



FACULTAD DE
FILOSOFÍA Y
HUMANIDADES
UNIVERSIDAD DE CHILE

Departamento de Filosofía

CEREBROS DE SILICONA

Un experimento mental para explorar la ontología de la conciencia

Tesina para optar al grado de Licenciado en Filosofía

Autor

Jonathan E. Lagos Pino

Profesor guía

Rodrigo González Fernández, Ph.D.

Seminario

Inteligencia, mente y sociedad:
de protones y máquinas a presidentes.

Santiago, Enero de 2016

*Para mi madre y mi padre,
cuyo proyecto de vida ha consistido en fabricar un
sinfín de días felices para mis hermanos y para mí.*

Agradecimientos

Quisiera agradecer a todas las personas que han sido un apoyo fundamental durante mi formación y vida. En primer lugar, a Montserrat Pino y Manuel Lagos, por ser mis padres; pero sobre todo, por encarnar toda la dulzura que he experimentado en mi vida.

A Damaris, Carlos y Rodrigo, por ser cariñosos hermanos y excelentes amigos.

A Paula, Verónica y Elena Lagos, por la alegría y amor que me entregan a pesar de la distancia y el poco tiempo que les dedico.

La vida me ha honrado con la amistad de pocas pero grandiosas personas. Todas y todos son dignos de mención. Le agradezco a Francesca Olate, Britt Benitez, Barbara Naeilson, Hernán Díaz, Gabriela Menanteau, Macarena Pagels, Enrique Costa, Mauricio Fuenzalida, Page Navarrete, Camila Petit, Susana Soto, Daniel Garcés, Rodrigo Alarcón, Tania Pinto y Ariel Calderón. Cada uno ha sido, tal vez sin saberlo, una fuente constante de apoyo emocional e intelectual. Estoy en infinita deuda con ellos.

Esta modesta investigación no hubiese sido posible sin el apoyo constante del Prof. Rodrigo González, quien a pesar de su apretada agenda académica, aceptó amablemente dirigir esta tesina. Su consejo, tanto humano como intelectual, han dejado una huella muy profunda en mí y en mi forma de abordar la filosofía. Quisiera mencionar también al Prof. Alejandro Ramírez, quien me ha inculcado, al participar en sus clases, el rigor argumental necesario para un adecuado pensar filosófico. Seguramente, sin las lecciones que he aprendido de ambos, esta obra tendría muchas más deficiencias de las que posee. Las que aún permanecen, son de mi absoluta responsabilidad.

Finalmente, agradezco a la Universidad de Chile, mi *alma mater*. Las aulas de la Facultad de Filosofía y Humanidades han sido como un invernadero en el que han germinado y crecido los conocimientos que hoy humildemente cosecho.

Jonathan Lagos Pino
Santiago, diciembre de 2015

Contenidos

Agradecimientos	ii
Introducción	v

Capítulo 1: Herramientas de análisis

1.1. Introducción	8
1.2. James Brown y los Experimentos Mentales Platónicos	9
1.3. John Norton y los experimentos mentales como argumentos	11
1.4. Principios para el estudio de <i>Cerebros de Silicona</i>	12
1.4.1. Primer Principio	13
1.4.2. Segundo Principio	13
1.4.3. Tercer Principio	17
1.5. La ontología subjetiva	19
1.6. La Filosofía de la Mente	22
1.7. Conclusión	23

Capítulo 2: Cerebros de Silicona

2.1. Introducción	25
2.2. El planteamiento de <i>Cerebros de Silicona</i>	25
2.3. Aplicación de los principios	27
2.3.1. <i>Cerebros de Silicona</i> y P1	27
2.3.2. <i>Cerebros de Silicona</i> y P2	30
2.3.3. <i>Cerebros de Silicona</i> y P3	34
2.4. Conclusión	36

Capítulo 3: Implicaciones para la ontología de la conciencia y el problema de las otras mentes

3.1. Introducción	38
3.2. Conclusiones para la ontología de la conciencia	39
3.3. Conclusiones para la epistemología de la conciencia	41
3.4. Criterios de atribución de conciencia	43
3.5. Una objeción a los criterios propuestos	46

Conclusión	48
------------	----

Referencias bibliográficas	49
----------------------------	----

Y hubo un momento, hace cuatro o cinco mil millones de años, en que no había mente alguna, ni simple ni compleja...o, por lo menos, no en este planeta. ¿Cuáles fueron los cambios, en qué orden se dieron y por qué? Los pasos principales parecen claros, aunque los detalles de las fechas y lugares sólo puedan ser conjeturas. Una vez contada esa historia, por lo menos tendremos un marco general en el que situar nuestros dilemas.

Daniel Dennett, *Tipos de Mente*

Introducción

Uno de los problemas centrales en el estudio de los fenómenos mentales conscientes, es precisar la relación entre la adscripción de conciencia a un sistema y el comportamiento que dicho sistema exhibe. La pregunta es: ¿es suficiente el comportamiento para la existencia de conciencia? Desde el análisis empirista, la conducta es la única evidencia a mano para atribuir estados mentales a un sistema. Sin embargo, en lo que concierne a la metafísica de la conciencia, el comportamiento de un sistema es irrelevante en relación con la existencia de estados mentales conscientes. En la literatura existen varios esfuerzos por intentar salvar este importante tropiezo. En efecto, una incorrecta aproximación puede conducir a una reducción inadecuada y problemática de la conciencia, en tanto es posible el comportamiento inteligente pero la ausencia de ella.

Uno de los pensadores que ha intentado clarificar este problema es John Searle. Varios de sus escritos presentan experimentos mentales que proporcionan importantes intuiciones acerca de la metafísica de la conciencia. Entre ellos destaca el experimento llamado *Cerebros de Silicona*, por las implicaciones referidas a la relación entre cerebro, conciencia y conducta que proporciona. A partir de él, Searle hace interesantes observaciones acerca del rol de control del comportamiento que se atribuye evolutivamente a la conciencia.

El objetivo general de esta investigación consiste en establecer adecuadamente las conclusiones que dicho experimento mental ofrece acerca de la relación entre la conciencia y la conducta. Debido a que los experimentos mentales son herramientas controvertidas, lo anterior requerirá formular condiciones y criterios que permitan lidiar con estos dispositivos epistémicos, de tal modo que las conclusiones no se vean debilitadas por las objeciones que se puedan presentar en contra de los experimentos mentales.

El objetivo específico que persigue este estudio es proponer dos criterios de atribución de conciencia sobre la base de las intuiciones obtenidas de *Cerebros de Silicona*. Argumentaré que, a pesar de las dificultades filosóficas asociadas al problema mente-cuerpo y al problema de las otras mentes, es posible determinar si un sistema posee o no estados mentales conscientes. En otras palabras, si no satisfacen tales condiciones, esto es, si no es posible detectar directa o indirectamente evidencia de los factores asociados a la existencia de estados mentales como correlato de la conducta de un sistema, la atribución de conciencia a dicho sistema no resulta justificada.

Estos objetivos serán abordados a lo largo de tres capítulos. En el primer capítulo, presento dos de las posturas más importantes sobre los experimentos mentales. Debido a que son posiciones parcialmente opuestas, me propongo elaborar con base en ellas criterios objetivos para la evaluación de los experimentos mentales; por lo tanto, todo el primer capítulo puede ser entendido como un marco conceptual. En el segundo capítulo, trataré extensamente los pormenores del experimento mental antes mencionado y aplicaré sobre él los principios propuestos en el primer capítulo con el fin de validarlo. En el tercer capítulo, abordaré las conclusiones (ya validadas) de *Cerebros de Silicona* y lo que contribuyen

al debate en filosofía de la mente. Finalmente, argumentaré a favor de los dos criterios de atribución de conciencia.

Herramientas de análisis

1.1. Introducción

El uso de experimentos mentales es una práctica extendida en las ciencias. Particularmente, en física existe un conjunto de ejemplos históricamente célebres. *El gato de Schrödinger* y la *Caída de los libros*, son algunos de los más importantes en la historia del progreso de esta ciencia. También en biología evolutiva se encuentra una muestra de ellos: Darwin, por ejemplo, imaginaba jirafas con cuellos de distinta extensión en *El Origen de las Especies*.

Mientras no está en duda la validez y distintos roles que estos dispositivos epistémicos cumplen en las ciencias naturales, su papel en filosofía es objeto de algunas sospechas. En general, se cuestiona la validez de las intuiciones o conclusiones que se extraen de su aplicación. Adicionalmente, las condiciones iniciales estipuladas en muchos casos parecen problemáticas por tratarse de escenarios exóticos e intuitivamente desconocidos. Esta última condición es común sobre todo en algunos experimentos mentales utilizados en filosofía de la mente y epistemología.

Este capítulo constituye un ejercicio de análisis que pretende formular precauciones metodológicas mínimas a la hora de estudiar experimentos mentales en filosofía de la mente. Mi investigación se centra en el experimento mental *Cerebros de Silicona*, en concreto, en el uso que John Searle hace de él para respaldar un conjunto de conclusiones referidas a la metafísica de la conciencia. Los principios obtenidos aquí serán aplicados en el segundo capítulo de esta investigación.

Mediante el contraste de las observaciones realizadas por John Norton y James Brown, pretendo extraer, a modo de síntesis, tres principios metodológicos que regulen el análisis del experimento mental. Serán herramientas para testear las observaciones, con el objetivo de no dar pasos inválidos durante la investigación.

En el primer apartado presento un resumen de la posición de James Brown. Me detengo en lo que él llama Experimentos Mentales Platónicos. En el segundo apartado presento la posición de John Norton. Lo importante aquí será ilustrar lo que él ha llamado *elimination thesis* o tesis de la reducción. Posteriormente, formularé los tres principios a partir de la presentación de las posiciones de los autores arriba mencionados. Luego, me referiré a análisis de la noción de ontología subjetiva y su relación con perspectiva de la primera persona. Para concluir, expondré en breve el rol de los experimentos mentales en filosofía de la mente en el contexto de interés actual y generalizado por la conciencia.

1.2. James Brown y los Experimentos Mentales Platónicos

En primer lugar, me referiré al análisis que James Brown ofrece acerca de los experimentos mentales. Su punto de vista difiere de la posición de Norton, en que atribuye un papel epistémico singular a una clase de experimentos mentales específicos. Dicha atribución se debe a la capacidad que tienen de permitir un acceso *a priori* a las leyes fundamentales de la realidad.

Antes de presentar los detalles de la posición de Brown, es indispensable una descripción general de los experimentos mentales. Asumiré que el lector está familiarizado con más de algún experimento de esta clase, ya que en este punto no haré referencia a ningún autor en particular. En general, es posible mencionar que existen dos razones para realizar un experimento mental. La primera motivación, es la ausencia de datos experimentales reales requeridos por el tipo de investigación que realiza el presentador del experimento. La segunda motivación, radica en que los experimentos reales no pueden ser ejecutados a causa de problemas de financiamiento u obstáculos técnicos no superados, pero que son en principio ejecutables. Entre los objetivos de estos dispositivos se hallan los siguientes: ilustrar la o las correctas interpretaciones empíricas de una teoría, refutar una teoría por las implicaciones empíricas que es posible extraer de ella, y apoyar o descartar ciertas intuiciones *a priori* sobre un ámbito específico de la experiencia. Esto es conseguido mediante la presentación de escenarios hipotéticos o contra-fácticos en un contexto de datos relevantes para los fines del experimento. Tales escenarios deben ser consistentes con la comprensión que se tiene del mundo.

Con esta descripción general en mente, es posible presentar el planteamiento de Brown. Típicamente, para él existen dos grandes clases: los experimentos mentales *destructivos* y los *constructivos*. La primera clase reúne aquellos diseñados para refutar una teoría específica mediante una *reductio ad absurdum* (Brown, 1991, p.34); mientras que la segunda agrupa a los de tipo *mediativo*¹, *conjetural* o *directo*. El primer tipo define a aquellos experimentos² que pretenden facilitar el paso de las premisas a las conclusiones, debido a que estas últimas pueden ser muy complejas o contra intuitivas. El segundo tipo define a aquellos experimentos en los que el presentador no parte de una teoría dada, sino que pretende formularla para explicar un fenómeno problemático (Brown, 1991, p.40). El tercer tipo es mixto. En él se comienza con un fenómeno explicado parcialmente para luego finalizar con una teoría bien articulada sobre tal fenómeno (Brown, 1991, p.41).

La particularidad del análisis de Brown, es que postula un cuarto tipo de experimento mental que posee rasgos de las dos clases primarias. Este tipo define a un número reducido de experimentos mentales agrupados bajo el nombre de *Experimentos Mentales Platónicos* (Brown, 1991, p.43). El rasgo común a este tipo de experimento, es que no sólo destruye una teoría previa, como los experimentos destructivos, sino que también formula una nueva teoría que expresa un orden fundamental en la realidad, como los experimentos constructivos. Lo interesante en esta clase, es que no se llega a la nueva teoría del mismo modo en que se llega a una conclusión, sino que dicha teoría es percibida (en analogía a la percepción sensorial) como un objeto teórico existente en el mundo platónico de las ideas

¹Utilizo el término «mediativo» para traducir la palabra inglesa «mediative».

²En adelante, cuando use el término «experimento» me estaré refiriendo a los «experimentos mentales». Cuando quiera referirme a experimentos que se efectúan realmente, haré las precisiones necesarias para evitar la confusión, aunque en general me referiré a ellos como «experimentos reales».

o leyes.

Para ilustrar el tipo anterior, consideraré uno de los experimentos mentales más notables en filosofía. Al parecer, no se ha advertido que Descartes utilizó esta clase de experimento para apoyar algunas de sus distinciones teóricas fundamentales. En un clásico pasaje de sus *Meditaciones Metafísicas*, el filósofo examina un cuerpo específico, un trozo de cera. En él se refiere a las distintas cualidades de este cuerpo. Pero luego, sostiene Descartes, se le acerca el fuego y:

[...] lo que en él quedaba de sabor se exhala, el olor se evapora, el color se cambia, su figura se pierde, su tamaño aumenta, se hace líquido, se abrasa, apenas se le puede manejar, y, aunque se le hiera encima, no dará sonido alguno (Descartes, 2011, p.95).

En este pasaje, Descartes describe el comportamiento del fenómeno experimental imaginario que ha propuesto como si se tratara del reporte directo de una genuina observación. Luego de esto, formula algunas preguntas que lo conducen hacia conclusiones metafísicas fundamentales. El filósofo francés continúa señalando lo siguiente:

¿Sigue siendo después de este cambio la misma cera? Preciso es confesar que sigue siendo la misma; nadie podrá dudarlo. ¿Qué es, pues, lo que se conocía en ese pedazo de cera con tanta distinción? Ciertamente nada puede ser de cuanto se observa por mediación de los sentidos, puesto que todas las cosas que caían bajo el dominio del gusto, del olfato, de la vista, del tacto y del oído, se encuentran cambiadas; y, sin embargo, la misma cera permanece (Descartes, 2011, p.95).

Lo que Descartes intenta establecer, es que este ejercicio produce un tipo de intuición no sensible de un rasgo del trozo de cera que permanece constante ante todas las transformaciones mencionadas. El filósofo finalmente concluye lo siguiente: “Considerémoslo atentamente, y, descartando todas las cosas que no pertenecen a la cera, veamos lo que resta. Ciertamente no permanece sino algo extenso, flexible y mutable” (Descartes, 2011, p.95). De este modo, Descartes determina la ontología común a todos los entes materiales, la *res extensa*. Este experimento mental es un caso de libro de experimento platónico, su conclusión aprehende la cualidad fundamental de la realidad material, la extensión. Esta «percepción» no es sensible, sino que captada por el entendimiento y mediante intuiciones modales³. Lo interesante de este caso, es que la «extensión» es un concepto geométrico, intuitivamente obtenido, pero independiente del tipo de experiencia que lo origina, por lo tanto, a partir de él es posible deducir matemáticamente el comportamiento de objetos físicos, ya que para Descartes la física es pura matemática. Este ejemplo ilustra claramente la cercanía entre la concepción de Brown y el racionalismo cartesiano.

Es posible establecer otro punto de conexión entre Descartes y Brown. El filósofo francés señala en el *Discurso del Método* que los silogismos, el tipo de argumento más usado en la época, “[...] sirven mejor para explicar a los otros las cosas que uno sabe [...] que para aprenderlas” (Descartes,

³Las intuiciones modales son una herramienta de pensamiento que permite indagar la separabilidad lógica de entidades metafísicamente independientes, mediante los conceptos lógicos de *necesidad* y *posibilidad*, y los conceptos metafísicos de *substancia* y *accidente*.

1982, p.50). Dicho de otro modo, no son herramientas útiles para el descubrimiento. Es posible conjeturar que esta opinión es lo que motiva el uso de experimentos como el del trozo de cera, y no intrincados argumentos analíticos a la hora de exponer sus conceptos fundamentales. Brown por su parte, reconoce un importante rol heurístico en los experimentos mentales, pues mediante ellos es posible descubrir los principios que regulan la realidad a un nivel fundamental. Su taxonomía hace evidente el rol propedéutico que algunos experimentos tienen. En efecto, algunos ejemplos como el que Newton utiliza en los *Principia* para explicar porqué la luna permanece en su órbita, poseen la capacidad de introducir y facilitar la enseñanza de las ciencias.

Sobre la base de la exposición anterior, es posible caracterizar el planteamiento de Brown como platónico. Epistémicamente concibe que los experimentos mentales, o al menos cierta clase específica y restringida, constituyen una puerta de acceso a un mundo de entidades teóricas fundamentales que describen la realidad física. Finalmente, los experimentos mentales, en tanto dispositivos epistémicos útiles, poseen un carácter extraordinario que contrasta con los experimentos reales.

1.3. John Norton y los experimentos mentales como argumentos

El término “experimento mental” ha sido claramente acuñado en analogía a los experimentos que efectivamente se llevan a cabo en los laboratorios o en otros ambientes controlados de investigación. Estos últimos, se realizan debido a objetivos diversos, entre los cuales destaca la contrastación de teorías o hipótesis sobre un ámbito restringido de la experiencia. En efecto, tras la ejecución de un experimento es posible refutar o confirmar un modelo teórico. Si el resultado es favorable a la teoría, se ha adquirido un nuevo conocimiento sobre la realidad. El sentido de esta afirmación consiste en que un conjunto de proposiciones referidas al comportamiento de ciertos objetos, describen de manera correcta dicho comportamiento, al punto de posibilitar la predicción de la conducta futura.

De la exposición anterior, se sigue que el propósito de la experimentación real es llegar a forjar conocimiento acerca de la realidad. Ahora bien, cuando se trata de experimentos mentales –y con el fin de sostener la analogía propuesta– cabe preguntar: ¿de dónde proviene la información sobre el mundo físico que proporcionan los experimentos mentales? Esta es la pregunta central que John Norton formula para cuestionar la validez de estos dispositivos epistémicos (Norton, 1996, p.333).

Su planteamiento es enunciado sobre la base de una filosofía empirista de la ciencia, esto es, que el conocimiento científico no es más que un conjunto de experiencias que han sido generalizadas y ordenadas de un modo consistente (Norton, 1996, p.335). El método constitutivo de esta posición es la experimentación empírica, por lo tanto, parece natural que desde su perspectiva los experimentos mentales sean objeto de algunas sospechas. En este contexto, Norton señala que: “[...] los filósofos empiristas de la ciencia como yo deben resistir toda sugerencia de que uno de los principales fundamentos experimentales de la ciencia, los experimentos reales, puedan ser reemplazados por las fantasías de la imaginación”⁴ (Norton, 1996, p.335). No obstante, el filósofo no pretende desestimar los experimentos mentales; reconoce que estos dispositivos han sido de gran utilidad en la ciencia y que aún queda

⁴ “[...] empiricist philosophers of science like me must resist all suggestion that one of the principal experimental foundation of science, real experiments, can be replaced by the fantasies of the imagination”

mucho que aprender de su correcta aplicación (Norton 1996, p.335).

Con el propósito de compatibilizar el uso de experimentos mentales con su empirismo filosófico en las ciencias, Norton propone una interpretación opuesta a la presentada por Brown. Debido a que estos dispositivos no pueden aportar nuevos datos experimentales al no ser efectivamente realizados, su única función debe consistir en reorganizar o establecer nuevas generalizaciones de los datos que ya se poseen sobre el mundo físico. De este modo, los experimentos mentales quedan definidos como dispositivos que agrupan supuestos empíricamente fundamentados. Esto ocurre con el objetivo de generar una organización diferente que logre proporcionar nuevas conclusiones, dinámica que exhibe su naturaleza argumental: “Es decir, estos dispositivos son argumentos que nos conducen desde los supuestos a la conclusión, el resultado del experimento mental”⁵ (Norton, 1996, p.335).

De presentar a los experimentos mentales como nada más que argumentos pintorescos, se sigue que todas las restricciones lógicas de validez mediante las cuales se lidia con los argumentos son aplicables a los experimentos mentales. De hecho, es central para esta interpretación que todo experimento mental pueda ser reemplazado por un argumento, esto es lo que Norton ha llamado *elimination thesis* o *tesis de la reducción* (Norton, 1996, p.336). A partir de esta reducción, es posible estudiar el experimento mental en su forma de argumento y determinar su validez. Se verá, como consecuencia de esta tesis, que el grado efectividad observado en un experimento mental coincide con la validez que es posible detectar o no en su forma argumental.

De acuerdo con esta serie de precisiones, Norton concluye que no hay nada extraordinario en los experimentos mentales desde el punto de vista epistémico, pero que parecen ser dispositivos útiles a la hora de argumentar y explicar teorías. Así pues, tratamientos como el de Brown, introducen innecesariamente complicaciones teóricas que abultan la taxonomía de experimentos mentales, añadiendo, en lugar de claridad, nuevas complejidades a su estudio.

1.4. Principios para el estudio de *Cerebros de Silicona*

La presentación sucinta de las teorías previamente expuestas, tiene el objetivo de fundamentar el diseño de un marco teórico para el adecuado tratamiento del experimento mental *Cerebros de Silicona* de John Searle, en una disciplina (la Filosofía de la Mente) en la que el uso de estos dispositivos a veces ha sido tratado con desconfianza. Como Brown señala, algunos críticos sostienen que en muchos casos la información del trasfondo supuesto en los experimentos no es relevante, y que frecuentemente se proponen escenarios extravagantes acerca de los cuales no se tiene intuiciones claras (Brown, 1991, p.29). No es mi objetivo responder a estas objeciones ni proponer una interpretación nueva sobre los experimentos mentales. En esta sección se elaborarán los tres principios metodológico para trabajar adecuadamente con estos dispositivos epistémicos. En otras palabras, sobre la base de las interpretaciones de Brown y Norton, se propondrán tres principios guía que serán utilizado en el análisis del experimento mental de Searle.

⁵“That is, these devices are arguments that carry us from our assumptions to a conclusion, the outcome of the thought experiment”

1.4.1. Primer principio

Uno de los supuestos implícito en el tratamiento de Brown, consiste en la posibilidad de realizar afirmaciones *a priori* acerca de fenómenos empíricos, sobre la base de ciertas reglas de inferencia lógica. Dicho supuesto, está en los cimientos de los experimentos mentales destructivos. Estos consisten en estipular un escenario hipotético relevante para una teoría, y a continuación deducir qué hechos se siguen de ella. Tales hechos aparecen como imposibles o absurdos de modo *a priori*, para dicha teoría. Gracias a este mecanismo, es factible descartar la plausibilidad de ciertos hechos y de teorías completas. De esto se desprende, que uno puede seleccionar como viables aquellas explicaciones que resistan la reducción al absurdo y explorar especulativamente sus implicaciones lógicas.

Teniendo en cuenta la observación de Norton (1996, p.335) referida al carácter imaginario de estos dispositivos de experimentación mental, resulta tremendamente útil poder contar con un mecanismo de testeo que excluya o elimine los experimentos mentales que no sean plausibles, es decir, aquellos que no superen una reducción al absurdo.

De acuerdo con este orden de razones, están dadas las condiciones para formular el primer principio metodológico:

P1. Descartar mediante reducción al absurdo tanto el escenario hipotético o contra-fáctico estipulado en el experimento mental como el comportamiento que se atribuye al fenómeno.

Tal como lo indica el principio, su objetivo consiste en ser un primer criterio de revisión del experimento mental, para precisar si es lógicamente aceptable. Resultaría «psicológicamente» más convincente lograr establecer si es o no empíricamente posible. Sin embargo, tal estipulación es sólo deseable. Lo que en última instancia es imprescindible, es determinar su posibilidad lógica. Esta distinción radica en, como afirma Schlick: “Lo que es empíricamente imposible sigue siendo concebible, pero lo que es lógicamente imposible es contradictorio y, por ende, no puede ser pensado” (Schlick, 1981, p.95). De este modo se elimina la dificultad que podría inducir implicaciones contradictorias a la hora de interpretar el experimento. Las condiciones de dicha interpretación serán discutidas en el siguiente apartado.

1.4.2. Segundo principio

En este apartado se enunciará el segundo principio metodológico de esta investigación. Para su formulación, se tomará como punto de partida la taxonomía de los experimentos mentales realizada por Brown. El primer principio que se ha formulado, puede ser entendido como una adaptación de la clase de experimentos mentales destructivos. Tales experimentos no son blanco habitual de las objeciones comunes contra el uso de experimentos mentales. Es posible que esto se deba a que dicha clase no tiene en general un carácter propositivo, es decir, su fin no es respaldar una teoría sobre la realidad sino más bien descartarla *a priori*. De modo contrario, los experimentos constructivos producen algunas serias sospechas debido a que su resultado consiste en un conjunto proposiciones sobre la realidad física que pretenden validez *a priori*. Entre ellos, los experimentos mediativos muestran ser los menos cuestionables.

Pero, ¿por qué razón los experimentos mentales mediativos son menos problemáticos? Para responder adecuadamente esta pregunta, es preciso volver a la caracterización realizada por Brown, quien señala que a pesar de que los tres subtipos constructivos comparten el objetivo de establecer resultados experimentales positivos, poseen estructuras diferentes (Brown, 1991, p.36). Es necesario aclarar que el filósofo afirma esto sólo a propósito de los tres tipos mencionados. Los experimentos platónicos, como señalé más arriba, proporcionan intuiciones fundamentales pero de un modo bastante enigmático, y son el blanco preferido de las críticas de Norton. Por lo tanto, para los fines de esta investigación, lo conveniente será no considerarlos como una categoría disponible. Retomando el punto, esta distinción estructural entre los tres subtipos, permitiría valorar los experimentos mediativos al margen de sus compañeros de clase.

Lo crucial en esta distinción, es que mientras los experimentos conjeturales y directos buscan formular o completar teorías a partir de un escenario estipulado, los mediativos sólo pretenden «ilustrar» o facilitar la comprensión de conclusiones específicas. Tales conclusiones, son generalmente contra-intuitivas o no parecen seguirse de las premisas que conforman la teoría ya existente. Para ilustrar este punto, será útil considerar un experimento mental que Brown clasifica como mediativo (Brown, 1991, p.36). El destacado físico austriaco Erwin Schrödinger, propuso en 1935 un experimento mental que pretendía ilustrar una de las interpretaciones de la física cuántica que va en contra hasta del sentido común más atrevido, «la superposición». El científico imaginó un sistema compuesto por un gato, una botella de gas venenoso, una caja de paredes opacas y un dispositivo radiactivo. Tanto el gato, la botella con veneno y el dispositivo radiactivo se encuentran dentro de la caja sellada. El dispositivo radiactivo al activarse emitiría un fotón que pasaría por un detector y al ser detectado, provocaría la liberación del gas venenoso, causando la muerte del gato. El físico señaló que, debido a que en este nivel de la física cuántica una partícula (el fotón) no está en una posición específica, podría ser o no ser detectada por el sensor. En consecuencia, el gas podría haber sido liberado o no, y finalmente el gato podría estar muerto o vivo. La única forma de saberlo es abriendo la caja, pero la observación fijaría al fenómeno en un estado específico. Es por eso que, desde el punto de vista teórico de la física cuántica, el sistema tiene estados superpuestos. En otras palabras, el gato está vivo y muerto simultáneamente.

La singularidad de este experimento reside en que intenta explicar en términos macroscópicos un fenómeno de nivel cuántico, el comportamiento de un fotón en términos de los estados de un gato. Schrödinger no pretende demostrar la superposición cuántica a partir del resultado de este experimento, y tampoco el experimento tiene un rol en el desarrollo histórico de la teoría, cuyo origen puede situarse en el experimento de la doble rendija. Su clara intención es mostrar que los medios ordinarios de intuición de la realidad son absolutamente inútiles a la hora estudiarla en un nivel fundamental, y cómo la observación o medición afecta al fenómeno. Por lo tanto, la validez de la teoría en ningún caso depende de la plausibilidad del experimento. No obstante, es un mecanismo bastante eficaz al momento de interpretar y explicar la teoría cuántica. Es bastante sencillo notar que este uso de la experimentación mental no supone una falta a los mecanismos epistémicos convencionales utilizados en la verificación experimental, como sugiere Norton (1996, p.333), porque de hecho, no hay verificación en lo absoluto.

Se ha establecido porqué los experimentos mediativos no representan ningún problema epistémico. Queda referirnos a la dificultad asociada al uso de experimentos conjeturales y directos. Como se verá, la dificultad radica en la ausencia de una teoría previa, o bien de una teoría adecuadamente articulada que sirva de base para realizar la experimentación mental. Tal como Brown señala: “Hay una importante clase en la que no empezamos desde una teoría dada. El punto de tal experimento mental es establecer algún (mental-experimental) fenómeno; entonces suponemos una teoría para explicar ese fenómeno”⁶ (Brown, 1991, p.40). Así es como el filósofo define los experimentos conjeturales, debido a que el procedimiento consiste en conjeturar una explicación para un fenómeno experimentado mentalmente. En otro pasaje, a propósito de los experimentos directos, señala que:

Ellos se parecen a los experimentos mentales mediativos en que empiezan con fenómenos (mental-experimentales) no problemáticos, mas que con fenómenos conjeturales. Y por otra parte, los experimentos mentales directos, como los conjeturales, no comienzan desde una teoría dada y bien articulada –ellos *terminan con una*⁷ (Brown, 1991, p.41, énfasis del autor).

Un ejemplo de este último tipo, de acuerdo con Brown, se encuentra en el experimento «caída de los libres» que Galileo presenta en *Dialogues Concerning Two New Sciences* (1974, p.65). En él se toma como punto de partida la teoría aristotélica sobre la caída de los cuerpos, la que expone que su velocidad de caída en un mismo medio es una función de su *peso*. Galileo procede mostrando que esta teoría posee una inconsistencia interna, para luego proponer la suya propia, en la que señala que la velocidad de caída de dos cuerpos es independiente de sus pesos, y que por lo tanto, caen a la misma velocidad.

En vista de estas definiciones, Norton indicaría que, a pesar del cuestionable origen imaginario de los datos experimentales que pasan a formar parte del cuerpo teórico, los experimentos descritos no son más que argumentos de mayor sofisticación. En efecto, a propósito del experimento de Galileo, señala que: “No puede haber duda de que lo que Galileo nos ofrece aquí es simplemente un argumento”⁸ (Norton, 1996, p.341).

Sin perjuicio de lo anterior, existen otras razones que apoyan esta interpretación, por las que es conveniente prescindir de la utilidad de ambos tipos de experimento. Como se ha señalado anteriormente, el fin primordial de la experimentación real en ciencias es la contrastación de las hipótesis o teorías. En ocasiones, además de lo anterior, los experimentos son utilizados como métodos de descubrimiento (Hempel, 1993, p.41), es decir, se presupone una hipótesis general como guía, pero el fin del experimento es hallar relaciones que no han sido postuladas con anterioridad. Ya se mencionó más arriba que, tanto Brown como Descartes, reconocen el rol heurístico en el caso de los experimentos mentales.

⁶“there is an important class in which we do not start from a given theory. The point of such a thought experiment is to establish some (thought-experimental) phenomenon; we then hypothesize a theory to explain that phenomenon”.

⁷“They resemble mediative thought experiments in that they start with unproblematic (thought-experimental) phenomena, rather than conjectured phenomena. On the other hand, direct thought experiments, like conjectural ones, do not start from a given well-articulated theory—they *end with one*”.

⁸“There can be little question that what Galileo gives us here is simply an argument”

En ciencias, este procedimiento es confiable debido a que las relaciones que se observan surgen espontáneamente de la realidad, esto es, no son establecidas por el observador. ¿Ocurre lo mismo en la experimentación mental conjetural o directa? Claramente no. En la experimentación mental, las condiciones iniciales se estipulan y el comportamiento del fenómeno se imagina, es decir, se trata de un proceso dirigido. A la luz de esta observación, es razonable la desconfianza de Norton en relación con el origen de los datos en la experimentación mental.

A su vez, es un requerimiento lógico esencial en la experimentación real, la formulación previa de teorías o hipótesis que propongan una explicación del fenómeno. De lo contrario, se carece de un criterio de relevancia que posibilite una recolección de datos adecuada. A propósito de esto, Hempel señala lo siguiente:

En resumen: la máxima según la cual la obtención de datos debería realizarse sin la existencia de hipótesis antecedentes que sirvieran para orientarnos acerca de las conexiones entre los hechos que se están estudiando es una máxima que se autorrefuta, y a la que la investigación científica no se atiene. Al contrario: las hipótesis, en cuanto intentos de respuesta, son necesarias para servir de guía a la investigación científica. Esas hipótesis determinan, entre otras cosas, cuál es el tipo de datos que se han de reunir en un momento dado de una investigación científica (Hempel, 1993, pp.29-30).

Como se puede advertir, Brown pretende que los experimentos conjeturales y directos hagan exactamente lo contrario a lo apuntado por Hempel, dicho de otra manera, que la teoría sea formulada al final del experimento. Aunque en rigor para el primero no se trate de un ejercicio de observación empírica, sino en un mecanismo heurístico, es importante tener en cuenta la necesidad de hipótesis previas, sin importar lo generales que puedan ser. El sentido de toda esta observación, es que las exigencias lógicas y epistémicas que se aplican a la experimentación real, deberían ser igualmente empleadas con la experimentación mental, incluso con mayor firmeza al considerar su naturaleza imaginaria y, por lo tanto, susceptible a la manipulación.

Hasta aquí se han presentado las razones por las que, el diseño de esta metodología, descarta los tipos de experimentos mentales conjeturales y directos como categorías adecuadas para el análisis. De este modo, los antecedentes suficientes para formular el segundo principio metodológico están dados. En conformidad con la línea argumental seguida, el segundo principio estipula lo siguiente:

P2. Los experimentos mentales plausibles deben caer en la categoría de mediativos o ilustrativos, de lo contrario deben ser descartados como epistémicamente problemáticos.

Espero que hasta aquí el análisis haya sido entendido con claridad y que el lector esté de acuerdo en que las conclusiones están adecuadamente establecidas. El sentido de esta exposición, es que el principio metodológico propuesto al final de cada apartado, se muestre intuitivamente adecuado y necesario.

1.4.3. Tercer principio

Para formular la tercera restricción o principio metodológico, es obligatorio volver sobre el planteamiento de John Norton. Su posición frente a los experimentos mentales entendidos como dispositivos ordinarios, es una consecuencia del empirismo filosófico en ciencias al que él adscribe. Aunque efectivamente su posición es consistente con el empirismo, no depende lógicamente de él. Norton aclara lo anterior afirmando que: “[...] en principio, uno puede apoyar la perspectiva argumental sin ningún compromiso respecto al origen de las premisas usadas en los argumentos y su conexión con la experiencia”⁹ (Norton, 1996, pp.336-337).

Lo interesante de esta indicación, es que permite obviar el contexto o disciplina en la que se utiliza la experimentación mental, debido a que los criterios de aplicación son mayoritariamente formales. No obstante, es conveniente estipular la relación entre las premisas y las conclusiones. El filósofo insiste sobre esto al apuntar lo siguiente: “Por ejemplo, uno tendría que asumir que no puede generar conclusiones con contenido experiencial a menos hayan premisas con al menos tanto contenido experiencial”¹⁰ (Norton, 1996, p.336). De acuerdo con esta observación, simplemente es forzoso regirse a una restricción ontológica básica que consiste en no desplazarse desde premisas referidas a un ámbito de la realidad, a conclusiones en un ámbito distinto.

El anterior, es un error frecuente en un gran número de argumentos deístas que proponen, a modo de ejemplo, lo siguiente: «Dado que es esencial a la definición de un triángulo el que posea tres lados, asimismo, la existencia es esencial a la definición de Dios. Tal como es absurdo concebir un triángulo sin tres lados, es absurdo concebir a Dios como no existente». He seleccionado este ejemplo porque el cambio de dominio ontológico es evidente, se pretende que el significado de una entidad lingüística implique la existencia de un ente material.

De este modo, y en conexión con los dos principios metodológicos presentados más arriba, la propuesta de Norton será utilizada como tercera restricción para el estudio del experimento mental de Searle. Como se habrá hecho evidente, la metodología que se ha estado proponiendo es conservadora. El objetivo de este conservadurismo es no dar paso en falso a la hora del análisis, asegurar que la aproximación al caso de estudio sea adecuada y que las conclusiones sean válidas.

Volviendo al planteamiento de Norton, el elemento central de su posición es lo que él ha llamado *tesis de reducción*. Dado el estatus no excepcional de los dispositivos de experimentación mental, su único rol debe consistir, en palabras del autor, en: “[...] organizar o generalizar lo que ya sabemos del mundo físico y hacerlo explícito”¹¹ (Norton, 1996, p.335). Para ilustrar esta aclaración, sería útil recordar el experimento mental de Schrödinger. Efectivamente, en ese caso, el experimento pretende presentar de manera sencilla un conocimiento adquirido con antelación, para hacer explícita una de sus interpretaciones.

Así, el resultado de la experimentación mental será confiable en la medida en que los supuestos que

⁹ “[...] in principle, one may hold the argument view without any commitments concerning the origin of the premises used in the arguments and their connection with experience”

¹⁰ “For example, one would have to assume that one cannot generate conclusions with experiential content unless there are premises with as least as much experiential content”

¹¹ “[...] organize or generalize what we already know about the physical world and make it explicit”

la teoría presenta sean correctos y las inferencias a partir de ellos sean válidas (Norton, 1996, p.335).

En función de la *tesis de reducción*, la que establece que cualquier experimento mental puede ser reemplazado por un argumento que no posea dicho carácter, los criterios para su evaluación, en tanto dispositivo epistémico, deben ser los mismos que se utilizan en la evaluación de inferencias. Norton destaca este punto de la siguiente manera: “Un buen experimento mental es buen argumento; un mal experimento mental es un mal argumento”¹²(Norton, 1996, p.335).

Como resultado de este análisis, quedan expuestos los antecedentes para formular el tercer principio metodológico. A la luz, nuevamente, de la tesis de reducción:

P3. Se debe evaluar la validez epistémica del experimento mental en su forma argumental, obtenida mediante la reducción del experimento.

Es necesario, una vez propuesta y justificada la utilización de los tres principios, mostrar el modo en que se articulan como un sistema o aparato metodológico consistente. El principio P1: *Descartar mediante reducción al absurdo tanto el escenario hipotético o contra-fáctico estipulado en el experimento mental como el comportamiento que se atribuye al fenómeno*, tiene el objetivo de filtrar el experimento determinando la plausibilidad del escenario que propone.

El principio P2: *Los experimentos mentales plausibles deben caer en la categoría de mediativos o ilustrativos, de lo contrario deben ser descartados como epistémicamente problemáticos*, tiene el objetivo de proporcionar un marco de interpretación para el experimento.

El principio P3: *Se debe evaluar la validez epistémica del experimento mental en su forma argumental, obtenida mediante la reducción del experimento*, tiene como fin evaluar la validez de las implicaciones obtenidas por el experimento mediante criterios lógicos. Lo anterior queda resumido en la siguiente tabla:

PRINCIPIO	FUNCIÓN
P1	Excluye o elimina
P2	Interpreta
P3	Evalúa

Cabe finalmente, hacer explícito un criterio que subyace a toda la exposición anterior. Se habrá advertido que las fuentes utilizadas se refieren específicamente al uso de experimentos mentales en las ciencias. Además, es patente que la propuesta de este marco teórico tiene como fin el ser aplicado a un experimento mental en Filosofía de la Mente. De lo anterior surge la pregunta: ¿es válida la aplicación de criterios extraídos del análisis de los experimentos mentales en ciencia, a los experimentos mentales en filosofía? No existen, a mi juicio, razones evidentes que impidan el uso de los mismos criterios. Sin perjuicio de lo anterior, es necesario recordar que el punto central de distinción entre los experimentos mentales en ciencia y filosofía es el origen de los datos que el experimento recoge: para la primera es la realidad física; para la segunda es la intuición y la lógica. Es decir, la distinción no radica en la estructura interna de los experimentos mentales, sino que en elementos externos.

¹²“A good thought experiment is a good argument; a bad thought experiment is a bad argument”

Ahora bien, tradicionalmente se sostiene una distinción fundamental entre las ciencias y la filosofía, particularmente en relación con la metodología. Dicha distinción fue señalada en el párrafo anterior y no está en discusión. Sin embargo, existen razones para pensar que la pretendida y marcada distinción entre estas disciplinas responde a motivos ajenos al orden intelectual, y se aproxima más a una infundada rivalidad entre las humanidades y las ciencias en su búsqueda de conocimiento.

Precisando de una vez, en tanto instrumentos para la creación de conocimiento teórico sobre realidad, los experimentos mentales son dispositivos primordialmente formales. Se ha señalado que es posible reducir un experimento mental a su forma argumental. Los criterios de evaluación de tal argumento, consisten en leyes lógicas de inferencia, y por definición, dichas leyes son de naturaleza formal. Por consiguiente, debido a ese carácter formal, son indiferentes al contenido semántico o compromiso ontológico de las premisas. En otras palabras, lo que guía el uso de experimentos mentales es la lógica, y la aplicación de la lógica es transversal a todas las formas de conocimiento. Entonces, obviando los requerimientos propios de cada disciplina, los experimentos mentales son válidos en todo tipo de investigación, al menos en principio.

Finalmente, y en el mismo sentido, como Searle señala (2003, pp.138-139), el conocimiento filosófico sobre un fenómeno es una etapa previa, y tal vez incierta, que establece los fundamentos metafísicos y epistémicos necesarios para el desarrollo de una explicación científica. En cuanto la pregunta por el fenómeno ha adquirido forma precisa, de modo que es posible darle una respuesta que todos puedan ver como correcta, entonces es llamada ciencia. Por consiguiente, es perfectamente posible concebir el análisis filosófico en continuidad con el análisis científico.

1.5. La ontología subjetiva

Hasta aquí sólo han sido propuestos los tres principios metodológicos, justificando su necesidad para un correcto examen del experimento mental *Cerebros de Silicona*. En esta sección pretendo proporcionar otra noción clave para el desarrollo de esta investigación: la ontología subjetiva. Será una introducción para un importante comentario realizado por Searle en relación con el empirismo en filosofía de la mente que se tratará más abajo.

Como supuesto fundamental a toda investigación de los hechos en el mundo, ya sea científica o de sentido común, se encuentra la creencia justificada –lo que Searle llama *Presuposición de Trasfondo* (2001, p.23)– de que existe una realidad externa. El sentido de «externa» debe entenderse como existiendo independientemente de cualquier representación interna (léase mental) que se haga de ella. Los sistemas de percepción sensorial y de representación lingüística carecerían de todo sentido sin esta realidad independiente. Asociado a esto, otro supuesto central señala que tales representaciones sensoriales y lingüísticas son, para todos los efectos epistémicos y prácticos, esencialmente idénticas. En otros términos, dos sujetos enfrentados a un triángulo rojo, por ejemplo, realizarán reportes que coincidirán en la mayoría de los detalles esenciales de la escena, considerándose como aceptables pequeñas discrepancias tal vez referidas a la intensidad del color o al tamaño de la figura. Esta universalidad de las representaciones tiene un potencial tremendo. Gracias a ella es posible alcanzar, o al menos intentar alcanzar, la objetividad epistémica imprescindible en las ciencias.

De hecho, uno de los criterios epistémicos que da sentido al método científico es que, dada una investigación precisada paso a paso con resultados específicos, cualquier otro investigador puede repetir el mismo procedimiento y constatar que el reporte original se ajuste a los hechos. Esta posibilidad es la que justifica, en última instancia, la «verdad» e «ínter-subjetividad» de los hallazgos de la ciencia. En síntesis, cuando se afirma que un enunciado es objetivo y verdadero, se sostiene que dicho enunciado describe adecuadamente la existencia de tal estado de cosas. Esta concepción es llamada *Teoría de la Verdad por Correspondencia* (Searle, 2001, p.23).

Ahora bien, como fue señalado, este modelo supone un acceso universal al estado de cosas descrito, que funciona perfectamente para casi todo. Los problemas aparecen cuando se intenta realizar enunciados para describir hechos o estados de cosas que por su naturaleza ontológica no son universalmente accesibles. Este rasgo es lo que define la ontología de los fenómenos mentales y de la conciencia en general.

Tal distinción, que resulta casi evidente, no ha sido realizada sino hasta décadas recientes. Para ser precisos, lo que ha sido comprendido hasta hace poco es cómo lidiar epistémicamente con dicha ontología. Bajo el criterio de acceso universal, un fenómeno cuya existencia sólo puede ser constatada desde la perspectiva de la primera persona, no es un fenómeno «real» o susceptible de análisis científico. Los reportes subjetivos parecen incompatibles con una investigación objetiva y científica. Además, no es posible perforar el cráneo de un sujeto para ver si hay una conciencia dentro de su cabeza. Tampoco tiene sentido identificar la conciencia con el cerebro, ya que claramente al usar la expresión «perdió la conciencia», en ningún caso queremos significar que el sujeto ha «perdido el cerebro». Aquella parece ser un estado del cerebro más que el cerebro mismo.

Este hecho obliga a repensar el criterio de objetividad científica, específicamente el modelo de observación pictórica y los conceptos que lo constituyen, a saber, el de sujeto y objeto. Este modelo es el que ha entorpecido los intentos de análisis de la conciencia, y es responsable de que teorías como el «conductismo» fueran ampliamente aceptadas en el pasado.

Como siempre, la historia proporciona lecciones muy elocuentes. A principios del siglo XVII, el modelo cosmológico de Ptolomeo había entrado en crisis tras la teoría heliocéntrica de Copérnico, que fue expresamente apoyada por Galileo como resultado de su observación de Júpiter. Si bien desplazar a la Tierra del centro del sistema fue un gesto revolucionario, lo que finalmente consolidó el modelo heliocéntrico fue la modificación de la trayectoria orbital. Kepler, al notar que su teoría matemática tenía una leve imprecisión, se vio obligado a abandonar la perfección geométrica del círculo heredada por los griegos, y explicar las intrincadas órbitas planetarias como trazando elipses. El científico consideró, sin embargo, que esta se trataba de una hipótesis *ad hoc*, aunque por primera vez las predicciones se unían a las observaciones (Hawking, 1989, p.5). De modo similar, lo que se ha intentado realizar en los sucesivos análisis de la conciencia, es proponer concepciones inapropiadas que permiten su estudio en términos del modelo de objetividad clásico. Lo que un grupo de investigadores ha comprendido, es que dado que el modelo no permite describir los hechos, es necesario idear otro y no ignorar los hechos. Es probable que una modificación de la noción de acceso universal parezca también una hipótesis *ad hoc*, pero su modificación está justificada por el imperativo de ajustar la teoría a los hechos. Sólo tras tal modificación será posible considerar a los fenómenos mentales, con su

particular modo de existencia, como materia adecuada de una indagación científica.

John Searle presenta en varias de sus obras (1992, 1997, 2001 y 2010), una distinción conceptual que le ha permitido lidiar consistentemente con este tipo de ontología. Los términos «objetivo» y «subjetivo», señala el filósofo, tienen al menos dos sentidos diferentes. Considérense los siguientes enunciados:

- [a] El Sr. Maturana está experimentando un fuerte dolor. ¡Enfermera, administre un calmante!
- [b] El Lago Vostok está a 3.700 metros de profundidad bajo el hielo.
- [c] El jugo de naranja es más ácido que el jugo de pomelo.
- [d] El sentido de la moda de mi esposa es excelente.

Si se examinan con detención, es fácil detectar lo que distingue y vincula a estas cuatro proposiciones.

Hay dos tipos de enunciados: los que se refieren a entidades independientes, es decir, cosas que existen separadas de la mente de un sujeto; y los que se refieren a entidades que sólo existen vinculadas con la mentalidad de un sujeto consciente. Las proposiciones [b] y [c] son ejemplo del primer tipo; [a] y [d] son ejemplo del segundo tipo.

Además de lo anterior, hay enunciados que realizan afirmaciones cuya verdad no depende de las preferencias o deseos del sujeto que las emite o de los receptores; y enunciados cuya verdad se establece sólo en virtud de las preferencias y deseos del emisor y los receptores. Las proposiciones [a] y [b] son ejemplo de lo primero; [c] y [d] son ejemplo de lo segundo.

Hecha esta distinción, parece clara la dualidad de sentido de los términos «objetivo» y «subjetivo». Existe un sentido «ontológico», referido al modo de existencia de una entidad, y un sentido «epistémico», referido a la validez de un juicio. En los enunciados propuestos arriba, hay una mixtura de ambos sentidos. En [a], por ejemplo, se formula un juicio epistémicamente objetivo acerca de una entidad ontológicamente subjetiva. Por otro lado, en [c] se formula un juicio epistémicamente subjetivo acerca de una entidad ontológicamente objetiva. En [b] y [d] se dan las otras dos combinaciones posibles. La siguiente tabla las resume a todas:

–	Objetividad Epistémica	Subjetividad Epistémica
Objetividad Ontológica	[b]	[c]
Subjetividad Ontológica	[a]	[d]

El objetivo de esta distinción es mostrar que frecuentemente, en el uso cotidiano del lenguaje, se realizan enunciados objetivos sobre entidades subjetivas. Si Searle está en lo correcto, no existe ninguna dificultad real respecto a poder realizar juicios objetivos de carácter científico sobre una entidad ontológicamente subjetiva como la conciencia.

No obstante la aceptación de este tipo de ontología, queda pendiente el problema del acceso a los estados conscientes. Lo que se ha hecho en esta sección es mostrar que realmente «el genio está en la botella», hace falta descubrir cómo sacarlo de ella. Un punto de discordia en esta materia, es la confiabilidad de los reportes que los sujetos puedan ofrecer en el contexto de un estudio riguroso

de la conciencia. Sin embargo, no parece imposible idear mecanismos que permitan maximizar la transparencia de tales reportes, para así validar su uso en esta clase de investigación.

Lo anterior sugiere que este tipo de indagación está confinada a un acceso indirecto a su objeto de estudio. No hay razones, sin embargo, para pensar que esto constituye una dificultad insalvable. Es posible que mediante la sofisticación de métodos indirectos, se obtenga un conocimiento muy preciso acerca de la naturaleza de la conciencia. Hay muchos ejemplos de este tipo en la ciencia, en los que por razones empíricas no es posible acceder al objeto que se pretende estudiar. La astronomía, sin ir más lejos, ha evolucionado en la medida que ha aprendido a examinar una serie de hechos o fenómenos vinculados con los cuerpos que pretende conocer. Así es como mediante el estudio de la luz reflejada por un planeta lejano, es posible conocer de qué está compuesto sin tener que aterrizar en su superficie y tomar muestras.

Para finalizar esta sección, me parece necesario destacar un punto central. Las complejidades asociadas al estudio de la conciencia y los fenómenos mentales son de orden empírico. No existe ninguna barrera lógica que impida tener un acceso cada vez más directo a estos fenómenos. El límite de exploración, como se señaló, está marcado por su ontología. Pero en lo que respecta al contenido de los estados mentales, es posible que eventualmente se diseñen dispositivos que permitan transferir el contenido de los estados mentales de una persona a otra, o tener acceso «en directo» a la vida consciente de un sujeto visualizada en una pantalla, por ejemplo. Ignoro si en la actualidad existe algún proyecto de esta clase, por lo que aclaro que se trata sólo de elucubraciones.

1.6. La Filosofía de la Mente

Para tener una idea más clara de los alcances y límites de esta investigación, es necesario incluir algunas nociones más. El fenómeno que se intenta abordar, se inscribe en un ámbito conocido como Filosofía de la Mente. Parece, por consiguiente, adecuado decir unas cuantas cosas sobre esta rama de estudio, con el propósito de completar el cuadro que ha sido esbozado hasta aquí y justificar un ejercicio de esta clase.

La preocupación filosófica y sistemática por los fenómenos mentales y la conciencia, empezó a consolidarse a finales de la primera mitad del siglo XX. La impresión general afirmaba la necesidad de un tratamiento riguroso de los problemas asociados a la vida mental de los sujetos conscientes, cuestión que tradicionalmente había sido terreno del misticismo y la religión. Ni los problemas ni la preocupación por ellos eran nuevos, pero el creciente desarrollo de las distintas ramas de la ciencia obligaba a formular teorías que intentaran salvar lo que Searle ha llamado “una importante laguna en la vida intelectual del siglo veinte” (1989, p.413). Sin embargo, el origen de las preguntas por la mente puede ser rastreado hasta los tiempos de Sócrates. Ya en Platón, su discípulo, aparecen formuladas las dos interrogantes centrales acerca de lo mental. Para él, la naturaleza inmortal del alma era lo que explicaba el acceso al conocimiento o «recuerdo» de las ideas, que constituyen el núcleo de su teoría del conocimiento. De este modo, la metafísica de la conciencia y su papel en la cognición, pasan a ser los temas primordiales de esta disciplina. En los inicios de la modernidad, también para Descartes, estos ámbitos aparecen estrechamente vinculados. A este importante filósofo francés se deben los enigmas

más estimulantes de la historia de la filosofía que han preocupado a generaciones de pensadores hasta la actualidad. De hecho, el objetivo de esta investigación es abordar uno de esos enigmas, el problema de la relación entre las mentes y los cuerpos (específicamente los cerebros) en los que se ejecutan.

Hay dos aspectos del análisis filosófico que suelen ser desconocidos o mal entendidos entre la comunidad actual de investigadores que estudian la conciencia desde otras disciplinas. Según Bechtel, estas dudas son respecto a la metodología que usan los filósofos de la mente, y la relación de sus indagaciones con los proyectos o avances en ciencia cognitiva (Bechtel, 1991, p.15). El autor señala a propósito de esto que, si bien la filosofía no posee un soporte experimental particular, cuenta con otras herramientas teóricas de gran utilidad. Una de ellas es la distinción entre conocimiento *a priori* y conocimiento *a posteriori*. Algunos filósofos sostienen que es posible establecer verdades determinadas sobre la mente sin la participación de la experiencia. Esto sería especialmente cierto cuando se aborda la metafísica de la conciencia. Una muestra de aquello ha sido expuesta más arriba, al comentar la distinción que Descartes establece entre la mente y el cuerpo mediante intuiciones modales.

Pero, ¿cuál es la relevancia que el análisis filosófico de la mente posee? ¿Cuál es su función en esta ola de interés por la mente? Aunque es posible afirmar, con temor a ser objetado, que los problemas de la mente y la conciencia pertenecen al dominio de la ciencia, existen una serie de dificultades conceptuales que es necesario resolver desde el comienzo. Al resolver estos problemas teóricos, se pretende fundamentar adecuadamente los incipientes estudios científicos de la conciencia.

Ahora bien, ¿qué hace tan especiales a estos rompecabezas conceptuales, que no pueden ser resueltos por los científicos cognitivos? En principio, los expertos en ciencia cognitiva están en perfecta condición para intentar resolver estos problemas. Tal vez la falta de entrenamiento para abordar estos largos obstáculos conceptuales, los disuade del intento debido a que típicamente se trata de problemas para los cuales no es posible diseñar experimentos (Bechtel, 1991, p.17).

En suma: se trata de abordar importantes problemas conceptuales que pretenden fundamentar los intentos científicos de estudio de la mente. Los filósofos, por su particular idiosincrasia, parecen ser los más adecuados para intentar resolver tales problemas, presentando argumentos a favor o en contra de tal o cual propuesta. Para llevar a cabo su labor, precisan de instrumentos que les permitan pensar con claridad y precisión. Los experimentos mentales, entre otras, son poderosas herramientas que facilitan dicha tarea. Pero, como toda herramienta poderosa, su uso implica algunos riesgos, por lo que es necesario tomar las precauciones adecuadas. La totalidad de este capítulo, constituye un esfuerzo por presentar un mínimo de precauciones y nociones previas para utilizar esta herramienta con seguridad. Inevitablemente su extensión ha superado el plan inicial, pero, como suele ocurrir en las obras de construcción, la preparación del terreno es indispensable para lograr agilizar las tareas posteriores.

1.7. Conclusión

A modo de conclusión, expondré a continuación los principales puntos discutidos en este capítulo, para que la totalidad del cuadro esbozado hasta aquí pueda ser visto de una sola mirada.

A partir de los planteamientos de Brown y Norton, fueron expuestos tres principios metodológicos

que tienen como fin realizar tres tareas independientes. La primera tarea consistía en verificar el estatus lógico del planteamiento del experimento mental mediante una típica reducción al absurdo. La segunda tarea consistía en proporcionar un marco de interpretación del experimento poniendo énfasis en las restricciones ontológicas que estos dispositivos imaginarios poseen. Para atender a dichas restricciones fue propuesto un modelo mediativo de interpretación, esto es, los experimentos mentales son mecanismos que median el paso de las premisas a la conclusión, sobre todo cuando esta última es compleja o contra-intuitiva. La tercera tarea consistía en idear un mecanismo de validación del experimento mental, debido a que resulta complejo juzgar sus contribuciones en su forma de relato. A partir de la *tesis de reducción* ideada por Norton, fue propuesto que la validez del experimento mental se determine al darle al relato forma argumental, de tal modo que su eficacia como experimento mental sea proporcional a su validez argumental.

En la parte final de este capítulo se introdujo la noción de ontología subjetiva. Dicha noción es de central importancia, porque permite entender el modo de existencia de los fenómenos mentales, y a su vez, fundamentar epistémicamente su investigación. Para cerrar el capítulo, se expuso brevemente la relación de la Filosofía de la Mente con el uso de experimentos mentales, en tanto mecanismos de indagación *a priori*.

Finalmente, el capítulo introdujo todas las nociones centrales y necesarias para enfrentar adecuadamente el análisis de *Cerebros de Silicona* con un marco conceptual consistente.

Cerebros de Silicona

2.1. Introducción

En este capítulo se presentará detalladamente el experimento mental *Cerebros de Silicona*. Searle formula tres variantes, las que serán tratadas extensamente sin hacer referencia a las implicaciones de cada una. El objetivo general consiste en aplicar los principios que fueron elaborados anteriormente, para luego dar lugar al estudio de las conclusiones que se extraen en el tercer capítulo.

El texto está compuesto por tres secciones principales que abordan uno de los tres principios, aplicándolos sobre el experimento mental. Al final de cada una de estas secciones principales se indica si el experimento superó o no con éxito los requerimientos que cada principio enuncia.

2.2. El planteamiento de *Cerebros de Silicona*

El objetivo que Searle persigue con este experimento mental, es desafiar la hipótesis según la cual lo mental o la conciencia mantendría una conexión esencial con la conducta (Searle, 1992, p.65). Su estrategia de argumentación consiste en presentar un escenario hipotético en el cual surgen fuertes intuiciones acerca de la metafísica de la conciencia.

El filósofo invita al lector a que imagine que su cerebro comienza a deteriorarse a causa de alguna extraña enfermedad. La zona que se muestra más afectada es el córtex visual. Los efectos ya empiezan a notarse, usted experimenta una reducción en su capacidad para detectar colores, cambios de luz y la calibración de distancia, entre otras. Sus médicos están bastante angustiados, han administrado todos los tratamientos conocidos, pero ninguno ha dado resultado. Su visión sigue disminuyendo a paso acelerado. Como última opción, deciden emplear un tratamiento experimental que consiste en conectar chips de silicona en su córtex visual. Tras la intervención, todos quedan asombrados, el tratamiento ha dado frutos. Los chips de silicona restauran su visión de manera óptima. Desafortunadamente, no se trata de una cura. La enfermedad sigue atacando agresivamente las otras áreas de su cerebro. En la medida que el daño se expande, los médicos reemplazan las zonas dañadas por más chips de silicona. Eventualmente, todo su cerebro es reemplazado por chips, al punto que usted llega a oír el sonido que emiten al procesar los impulsos eléctricos.

La anterior es la escena descrita por John Searle. He intentado enfatizar los detalles para que se capte con claridad la sencillez del planteamiento. A primera vista, no existe ningún elemento que pueda ser considerado como problemático de modo *a priori*. A continuación, el filósofo considera tres resultados posibles que se seguirían de la aplicación del tratamiento, es decir, los efectos que podría

tener en los procesos biológicos vinculados a la actividad cerebral y la conciencia.

En primer lugar, bajo el supuesto de que los chips de silicona son capaces de duplicar tanto los procesos *input-output* como los fenómenos mentales conscientes y no-conscientes que son dependientes de y realizados por tales procesos, la vida mental del sujeto (en este caso usted, el lector) se mantiene intacta. Tal como cuando sólo es reemplazado el córtex visual, esta vez usted no detecta ninguna anomalía en su experiencia consciente. Todo va bien.

El segundo resultado posible consiste en que, tras la implantación progresiva de los chips de silicona en un área cada vez mayor de su cerebro, usted comienza a notar que la experiencia consciente asociada a las zonas sustituidas desaparece. Supongamos que aún no es reemplazada el área cerebral asociada a los procesos auditivos. Es sabido, que este tipo de operaciones son realizadas con el sujeto en estado de vigilia, de manera consciente. Esto permite a los médicos ir testeando los efectos que la manipulación del tejido cerebral produce en la experiencia subjetiva del individuo. Como decía, la corteza auditiva primaria no se ha reemplazado, por lo tanto logra escuchar que uno de los médicos le señala que han situado frente a usted un objeto rojo. A continuación le preguntan: «¿Qué es lo que ve?». Usted se desespera, no ve absolutamente nada y quiere responder: «¡No veo nada, estoy totalmente ciego!». Pero, para su desconcierto, usted escucha su inconfundible voz diciendo: «Veo un objeto rojo frente a mí», y por más que se esfuerza en rectificar su respuesta, no lo logra.

En el anterior resultado, se aprecia que la capacidad de duplicación total que fue asignada a los chips de silicona en el primer caso, no es lograda aquí. Claramente las funciones *input-output* son eficazmente duplicadas, pero los chips no son capaces de ejecutar los procesos conscientes. Esto tiene como resultado que la capacidad de producción de la conducta permanezca intacta, pero desvinculada de la experiencia consciente. En este escenario, tras el reemplazo total de su cerebro, perderá totalmente su vida mental consciente.

La tercera variación que Searle propone es la siguiente: como en los casos anteriores, su cerebro es progresivamente reemplazado por los chips. Los médicos observan que reacciona bien a las modificaciones y usted no experimenta alteración alguna en sus pensamientos, recuerdos, sentimientos e intenciones. Sin embargo, en la medida que la implantación continúa, su conducta externa va menguando hasta que eventualmente desaparece. A pesar de que su vida mental está intacta, no puede poner en acción ninguna de sus intenciones de hacer o decir tal o cual cosa. En efecto, usted puede escuchar que los médicos discuten sobre su situación, señalando que, a pesar de la mantención eficaz de los procesos vitales, se encuentra en un estado de muerte cerebral. Lo que el protocolo de bioética instruye en estos casos, es que usted debe ser desconectado debido a que carece absolutamente de vida mental. Usted estaría totalmente conmocionado, usted sabe que está totalmente consciente y que puede percibir claramente todo lo que ocurre a su alrededor, pero que, por más que intenta, no puede causar ningún movimiento corporal.

Han quedado así expuestos los tres resultados que Searle se propone como posibles. He desplegado de manera más detallada los escenarios presentados con el fin, como dije más arriba, de que la comprensión de los efectos sea óptima. En cualquier caso, ninguna de las modificaciones que he efectuado distorsiona el sentido que intentan transmitir.

Antes de someter el experimento a la maquinaria de análisis propuesta en el capítulo anterior,

resultaría conveniente detenerse en algunas de las observaciones que el filósofo norteamericano realiza en paralelo a su presentación. Primeramente, señala que, en su opinión, el hipotético procedimiento de reemplazo y duplicación de las capacidades causales del cerebro es remoto, experimentalmente hablando. Searle lo expresa del siguiente modo: “Creo que es empíricamente absurdo suponer que podríamos duplicar los poderes causales de las neuronas totalmente en silicona”¹³ (Searle, 1992, p.66). Sin embargo, la validez de su planteamiento, como él mismo reconoce, descansa en la posibilidad lógica o conceptual del experimento mental. En definitiva, su escepticismo sólo podría ser confirmado *a posteriori*. Del mismo modo, la posibilidad de los resultados de las tres variaciones del experimento, es lógicamente sostenible en la medida que ninguno puede ser descartado a la luz de objeciones *a priori*.

La segunda observación realizada por Searle, se refiere a la importancia de imaginar este experimento desde la perspectiva de la primera persona: “Pregúntese, «¿Cómo sería ser como para mí?»”¹⁴, insiste el filósofo (Searle, 1992, p.67). De este modo, es perfectamente concebible que su comportamiento permanezca normal, pero que usted se encuentre en estado de muerte cerebral. Asimismo, se debe tener claridad acerca de la base biológica de esta separación entre conducta y consciencia. El filósofo señala que, es muy probable que la base de realización de la consciencia se encuentre en alguna zona del cerebro. Una candidata es la formación reticular. Si esto es correcto, es fácil conjeturar que dicha región específica (sin importar cuál sea realmente) pueda deteriorarse a tal punto que la consciencia de un sujeto quede eliminada. Pero dado que el resto del sistema continúa intacto, la conducta externa puede seguir siendo mantenida.

2.3. Aplicación de los principios

A continuación se presentará el examen detallado del escenario planteado por John Searle, a la luz de los principios propuestos en el marco teórico desarrollado en el primer capítulo.

2.3.1. Cerebros de Silicona y P1

Lo que el Principio P1 indica, es que la primera prueba que un experimento mental debe superar apunta a su consistencia lógica. Mediante una típica reducción al absurdo, se examinará si el escenario posee o implica alguna contradicción.

El experimento mental *Cerebros de Silicona* posee tres capas, a las que me referiré con los nombres de «escenario», «maniobra» y «resultados». Para organizar adecuadamente el tratamiento del caso, serán inspeccionadas por separado, y sólo al final de este apartado las abordaré en bloque. Algunas de las observaciones que se harán aquí parecerán obvias y de sentido común, pero con el fin de iluminar todos los rincones del planteamiento de Searle, es importante mencionarlas.

La primera capa es el escenario. No hay nada complejo en su presentación. Se trata de un sujeto que padece una extraña enfermedad que aqueja a su cerebro, y para la cual no existe un tratamiento estandarizado. Lo particular de esta afección es que, al menos en principio, un trasplante no es un tratamiento adecuado, como lo sería, por ejemplo, si tratara de una enfermedad cardíaca. Bastaría

¹³ “I think it is empirically absurd to suppose that we could duplicate the causal powers of neurons entirely in silicon”

¹⁴ “Ask yourself, «What would it be like for me?»”

con disponer de un donante compatible de corazón, y la intervención se realizaría de inmediato. Pero ya que se trata de un cerebro, se entiende que la «identidad personal», en algún sentido del término, reside en el cerebro del sujeto. Luego, si se trasplantara un cerebro distinto a su cuerpo, se produciría un escenario bastante exótico, del cual resulta difícil hablar especulativamente. Dicho sea de paso, este escenario revela que la visión materialista-biológica de la mente acierta al vincular la mentalidad de un sujeto a su biología cerebral. En otras palabras, si se trasplantara otro cerebro a su cuerpo no se esperaría que la mente del propietario del cuerpo se «ejecute» intacta en su nuevo *wetware*, sino que parece claro que se trataría de la mente de otro.

Gran parte de lo anterior es, como se dijo, bastante evidente. Lo importante es que destaca la peculiaridad del caso médico. Si aplicamos el P1 es, hasta donde consigo ver, manifiesto que no existe nada contradictorio en suponer que alguna enfermedad pueda atacar de este modo destructivo al cerebro de un sujeto. Luego, al menos a nivel del escenario, el experimento mental es no contradictorio y, por lo tanto, lógicamente concebible.

¿Qué ocurre con la maniobra? Lo que los médicos de este paciente imaginario proponen es conectar a la zona dañada, reemplazando el tejido cerebral, un conjunto de chips de silicona. Lo que en esta etapa está supuesto, es que los chips cuentan con la capacidad para duplicar tanto el flujo de información electro-química de la zona como la experiencia consciente asociada a ese tipo de actividad, la experiencia visual. Searle aclara que se trata de un riesgo, los médicos no están seguros del éxito del tratamiento. El reemplazo empieza sólo por el córtex visual y luego abarca la totalidad del cerebro. Ya fue advertido el escepticismo del filósofo respecto al éxito de una intervención de este tipo, pero lo que aquí será destacado es la dificultad de establecer *a priori* el fracaso de la misma. Por lo tanto, es totalmente concebible que la maniobra pueda ser ejecutada con posible éxito.

Como nota de lo anterior, vale la pena mencionar que recientemente se han realizado experimentos reales similares al presentado por Searle. El sujeto ha sido un ratón al que han conectado una prótesis cognitiva en el hipocampo con resultados aparentemente exitosos¹⁵. Es previsible que el siguiente paso sea su testeo en humanos, lo que obliga a moderar el escepticismo del filósofo norteamericano hacia el éxito de este tipo de proyectos.

Lo siguiente es examinar los resultados estipulados por Searle. El primer resultado es totalmente exitoso. Los chips duplican efectivamente los estímulos *input-output* del córtex visual, y presuntamente ocurriría lo mismo una vez reemplazado todo el cerebro. Dado este efecto, no parece haber nada inconcebible o contradictorio, pero me permitiré mencionar una implicación, tal vez menor, de este resultado. Me parece que del éxito de este tratamiento se sigue que, si fuera posible simular la totalidad de la compleja información electro-química que el cerebro procesa, por ejemplo, en un computador y se hiciera circular por un cerebro de silicona, se obtendría un caso aislado de conciencia realizada o ejecutada en un complejo de chips de silicona. Es decir, se darían experiencias de color, sabor, etc., en esa masa de silicona aislada. Se presentaría una dificultad clara en este caso, no sería posible constatar la ocurrencia de conciencia, ya que la ontología de la conciencia es impermeable, y además no podría

¹⁵Para más detalles ver Berger, T. W. (2012), “A Hippocampal Cognitive Prosthesis: Multi-Input, Multi-output Nonlinear Modeling and VLSI Implementation”, IEEE Trans Neural Systems Rehabilitation Engineering. 20(02): 198-211.

ser deducida a partir del comportamiento porque no existe un cuerpo «conectado» a este cerebro. Por supuesto que el hecho de no poder constatarla no implica que no se dé realmente la conciencia. Llamo la atención sobre esta conjetura porque claramente supone una tesis compatible con el funcionalismo, posición que Searle no suscribe, lo que explica que no considere ampliamente este escenario.

En el segundo resultado estipulado por Searle, el énfasis es puesto en la disociación entre los procesos *input-output* y la experiencia consciente. Nuevamente se debe enfrentar una situación en la que es complejo negar *a priori* el resultado sugerido por el filósofo. Resulta, entonces, concebible que a causa de la incompleta duplicación de las funciones neuronales que los chips realizan, los procesos conscientes asociados al manejo de información sensorial (visual y auditiva, como lo sugiere el experimento), y la respuesta conductual hacia tales estímulos, se den de tal modo que los pensamientos conscientes pierdan su conexión causal con tales procesos.

Existe un tipo de «evidencia» que respalda la posibilidad real de esta disociación. En algunos manuales de neuroanatomía se define el estado de conciencia como el estado de vigilia en el que hay conocimiento de sí mismo y del entorno. Si suponemos que la formación reticular mantiene a un sujeto en estado de vigilia pero que la corteza cerebral está activa, es concebible que dicho sujeto reaccione al entorno y exhiba conducta consciente pero que no esté en estado de conciencia. Por ejemplo, un sonámbulo puede tener conductas bastante complejas mientras se encuentra dormido, y claramente nadie considera que un sonámbulo esté consciente. Por lo tanto, resulta además empíricamente posible que se dé conducta compleja (incluso la emisión de palabras) sin control consciente. No pretendo que este ejemplo sea demostrativo, pero me parece que apoya empíricamente la intuitiva separabilidad de la conducta y la percepción, de la conciencia.

El tercer resultado, nuevamente se refiere a la disociación entre la conducta y la conciencia. Sin embargo, esta vez la consecuencia de la separación es la parálisis total del individuo. Es decir, los chips son capaces de ejecutar los procesos conscientes y mantener las funciones vitales, pero el poder causal de la conciencia ha desaparecido. Al parecer, no es posible extraer ninguna contradicción de un supuesto de esta clase. De hecho, es un evento que es parte de la experiencia consciente de algunas personas. Existe un fenómeno llamado «parálisis del sueño» en el que un sujeto puede dejar el estado de vigilia, esto es, recuperar la conciencia, pero verse imposibilitado de realizar cualquier movimiento corporal. Si suponemos que alguien lo observa mientras experimenta el fenómeno, probablemente el reporte del observador exprese que el sujeto estaba totalmente inconsciente, cuando realmente era capaz de percibir su entorno. La presunta causa de este fenómeno radica en la falta de coordinación entre el control motor y la regulación del estado de vigilia por parte de la formación reticular. Con base en esto, el resultado de esta variante no sólo resulta concebible, sino también empíricamente posible.

Hasta aquí, cada una de las capas del experimento mental ha sido examinada a la luz de P1. El análisis ha mostrado que los elementos que componen tanto el escenario, la maniobra y los resultados son concebibles, no poseen ninguna contradicción interna y no dan lugar a conclusiones absurdas. Observados en bloque, parecen reflejar el comentario que Searle realiza respecto de su consistencia lógica, esto es, que no es posible desecharlos sobre la base de ninguna consideración *a priori*.

Al finalizar este apartado sería útil volver sobre una observación anterior. Como dije al inicio, es muy probable que gran parte de los comentarios hechos en esta sección resulten obvios a partir de

una buena comprensión del experimento. No obstante, uno de los objetivos transversales a toda esta investigación es hacer explícitos una serie de elementos relevantes a la hora de juzgar la validez de esta propuesta. De lo contrario, se corre el riesgo de comprometerse con una serie de supuestos en los que no se repararía en un sondeo superficial.

2.3.2. Cerebros de Silicona y P2

El segundo Principio indica que el experimento mental debe ajustarse a una interpretación «mediativa». Este criterio fue estipulado debido a los problemas que, según Norton, presentan las otras interpretaciones de los experimentos mentales, sobre todo la interpretación platónica de James Brown. Para poder aplicar el P2, será necesario dividir artificialmente el planteamiento total. Para estos efectos, se considerarán tres niveles: el nivel de los supuestos, el nivel de los medios, y el nivel de los objetivos.

Para hacer explícitos el contenido de los tres niveles mencionados, es necesario responder las siguientes tres preguntas: ¿Cuál es el objetivo del experimento, o qué tesis pretende defender? ¿Cómo o mediante qué estrategia intenta conseguirlo? Y, ¿cuáles son los supuestos iniciales de la exposición? Expondré el desarrollo del experimento en sentido inverso, desde los objetivos hasta los supuestos, en el mismo orden que he presentado las preguntas. Esto permitirá apreciar cómo funciona el experimento, a modo de *ingeniería inversa* o en similitud al trabajo de un detective frente a la escena del crimen, que retrotrae los acontecimientos hacia las causas que en el pasado lo desencadenaron.

La primera pregunta es: ¿Cuál es el objetivo del experimento, o qué tesis pretende ilustrar? Como ya fue señalado, el propósito de John Searle es argumentar que la conciencia o la mente es independiente de comportamiento. En un contexto de análisis filosófico, una tesis de este tipo puede ser fácilmente comprendida. La claridad que presenta no permite ambigüedades, o se acepta o se rechaza. Por supuesto que esta tesis no pretende señalar que el comportamiento es empíricamente irrelevante para la mente, solo que, tanto la mente como la conducta, pueden darse separadamente. Para el filósofo norteamericano no existe duda respecto a la realidad de esta separación. Sin embargo, el escepticismo que puede provocar parece estar justificado: lo que está detrás de esta afirmación es una tesis metafísica bastante fuerte.

Pero, ¿cuál es el problema con esta proposición metafísica, qué la hace tan fuerte? En pocas palabras, es posible apuntar a que la aceptación de esta tesis produce o podría producir obstáculos que entorpecen el estudio científico de la conciencia. Si la conciencia y los estados mentales no tienen un correlato de *outputs* conductuales, la tarea de estudiarlos se vuelve prácticamente imposible. Sería como tratar de examinar desde el exterior la estructura interna de una casa sin ventanas. Esta fue la conclusión a la que llegaron los conductistas de la primera mitad del siglo XX.

El problema es complejo, por una parte existe la evidencia subjetiva directa de estados mentales conscientes, la que es tan inmediata, que parece inadecuado tratar de demostrarla sin suponerla en el intento; y por otra parte, está la necesidad de contar con un modelo o mecanismo de acceso a los estados mentales que faciliten su estudio. Gran parte de los detalles de esta dificultad fueron discutidos más arriba, en la sección dedicada a la ontología subjetiva.

En este contexto, es fácil justificar el intento de argumentación que Searle realiza. Propone una

tesis metafísica fuerte y polémica, y como la mayoría de las tesis metafísicas, es formulada en términos muy abstractos, de ahí que el filósofo intente derivarla de manera intuitiva a partir de un experimento mental, donde propone un escenario empírico favorable a su tesis. En síntesis, pretende defender la tesis metafísica que señala que la *conciencia* (los estados mentales conscientes) *es ontológicamente independiente del comportamiento* (o estados conductuales).

La primera pregunta ha sido respondida, corresponde abordar la segunda: ¿cómo o mediante qué estrategia intenta conseguirlo? Hay un modo fácil de responder a esta pregunta, y un modo difícil. El modo fácil implica una respuesta obvia: su estrategia es usar un experimento mental. Ciertamente una respuesta de este tipo no contribuye mucho a la comprensión real de la estrategia de Searle. Además, toda esta investigación tiene como tarea transversal dar contenido al concepto de experimento mental y definir parámetros para los mismos. El punto no es si usa o no un experimento mental, sino más bien qué posee este experimento que lo hace tan sugestivo. La respuesta difícil tiene que ver con esto último. Una parte de ella fue mencionada en el análisis de la sección anterior, y está relacionada con la posibilidad lógica de concebir esta distinción a nivel conceptual. Esto es suficiente para aceptarla, y no desecharla por absurda o descabellada.

Pero, que la consistencia lógica de esta tesis haya sido establecida, no es suficiente para que valga la pena apoyarla. También es concebible que usted y yo no seamos más que cerebros flotantes en cubetas, química y electrónicamente estimulados, pero no veo como alguien podría sostener una tesis de esta clase más que como un estimulante desafío epistémico. Por esto me parece, que el punto fuerte del experimento *searleano* consiste en explotar un conjunto indefinido, y probablemente infinito, de intuiciones de lo que significa tener una mente consciente. Una de ellas, es que algunos estados mentales conscientes no son expresados conductualmente; y que incluso los sujetos conscientes pueden experimentar que controlan cuáles estados mentales tienen una realización conductual y cuáles no. Ahora mismo yo puedo experimentar el deseo de recostarme sobre la cama para descansar, pero determino que este estado mental no tenga expresión conductual, y que por otra parte mi intención de seguir escribiendo sí la tenga. Si alguna persona que me observa afirmara que yo nunca experimenté en este lapsus el deseo de recostarme sobre mi cama porque mi cuerpo no se movió de tal y tal forma, estaría claramente equivocado, debido a que la existencia de un estado mental, no está determinada por la realización de sus condiciones de satisfacción, tal como se infiere de las distinciones *searleanas* al respecto (Searle, 2001 y 2010)¹⁶. Como fue mencionado más arriba, existen ejemplos más exóticos (la parálisis del sueño y el sonambulismo) de esta separabilidad, pero la mayoría de estas intuiciones son más comunes que los casos exóticos.

Para resumir lo dicho a propósito de la segunda pregunta, Searle utiliza el experimento mental para aislar estas intuiciones, y son estas intuiciones las que refuerzan la distinción ontológica entre conciencia y conducta, más allá de la posibilidad lógica de dicha distinción.

Queda abordar la última pregunta: ¿cuáles son los supuestos iniciales de la exposición? Hasta aquí se ha sostenido que Searle apoya su tesis metafísica aislando un conjunto de intuiciones mediante el

¹⁶Es interesante notar que la base de esta distinción se remonta a Descartes, tanto en la independencia ontológica como en la formulación del juicio. Searle es, por lo tanto, marcadamente cartesiano en relación con la fenomenología de la conciencia y el modo de formular juicios sobre ella.

experimento mental que se ha descrito más arriba. Sin embargo, estas intuiciones no son primitivas, es decir, dependen lógicamente de al menos tres nociones o supuestos fundamentales.

La primera noción consiste en lo que Tim Crane ha llamado «imagen causal del pensamiento». Con esta noción Crane está afirmando tres cosas: [1] que la relación pensamiento-conducta implica la veracidad de algunos enunciados contrafácticos (si bebí una copa de vino es porque deseaba beber vino, de tal modo que si no hubiese experimentado ese deseo no hubiese bebido la copa de vino); [2] que citar a los estados mentales como causa de los estados conductuales es un modo de explicarlos, “[...] ya que citar causas es una manera de explicar efectos” (Crane, 2008, p.105); y [3] que las relaciones causales implican regularidades o leyes. Lo interesante de la noción de causalidad de Crane, es que en ningún momento es definida como una relación física, sino que una relación es causal si satisface estos tres criterios. Afirmer una relación de este tipo entre pensamientos o estados mentales y la conducta, implica que son conceptualmente distintos y empíricamente separables. En efecto, cuando se invoca la noción de causalidad es siempre entre dos entidades distintas: $A \longrightarrow B$ (A causa B). Si, por ejemplo, la entidad A pasa de estar en el estado A_1 a estar en el estado A_2 , no se describe dicha situación afirmando que $A_1 \longrightarrow A_2$ (A_1 causa A_2). Si no interviene otra entidad, se afirma simplemente que A cambió¹⁷. Señalo esto porque es importante aclarar que el vocabulario causal supone estas distinciones. Es debatible si esta separación conceptual implica categóricamente que se trate de una distinción real, pero insisto en señalar que el uso de esta noción es un indicio importante.

Debido a que suponemos la imagen causal de los pensamientos, es posible postular un conjunto de principios sobre el comportamiento de terceros y el propio. Estos principios se agrupan en lo que se ha llamado tradicionalmente «psicología popular»¹⁸ [*folk psychology*]. Aquí usaré la expresión propuesta por Crane, «psicología de sentido común», ya que el término «popular» posee una connotación negativa en algunos contextos. Para evitar ese mal entendido, he optado por la expresión alternativa. La psicología de sentido común, como iba diciendo, agrupa una serie de principios que pueden ser expuestos en términos más o menos explícitos. El principio que interesa aquí puede ser expresado del siguiente modo: «en general, las personas intentan llenar las condiciones de satisfacción de sus estados mentales o pensamientos, aunque ningún pensamiento implica necesariamente sus condiciones de satisfacción».

Considérese el siguiente ejemplo: un estudiante de medicina está estudiando en su escritorio, y suponga que alguien le pregunta: «¿Dónde te gustaría estar ahora mismo?» El estudiante de medicina responde: «Me gustaría estar en el jacuzzi de un hotel cinco estrellas junto a Mélanie Laurent». Asuma que el estudiante realmente tiene la posibilidad de satisfacer ese deseo, que tiene el dinero para hotel y que conoce a Mélanie. Ahora bien, no sólo tiene este deseo, también desea titularse como médico en cuatro años más, y para poder lograrlo necesita completar con éxito todos sus exámenes, uno de los cuales tiene mañana y esa es la causa por la que se encuentra estudiando.

¹⁷Claramente, el cambio de A puede ser descrito también como producido por una relación entre elementos internos de A, pero dicha relación causal se daría en un nivel explicativo más básico. En el nivel explicativo en que A es una entidad cerrada, A_1 no es causa de A_2 .

¹⁸No me interesa en este texto la discusión sobre la validez o relevancia de la psicología popular. Toda investigación parte de ciertos supuestos, uno de los supuestos de esta investigación es la aceptación de la psicología popular como una teoría psicológica válida.

El caso es que el estudiante tiene dos pensamientos distintos en relación con lo que gustaría estar haciendo ahora, pero sólo uno de estos pensamientos tiene expresión conductual. ¿Cómo es posible que ocurra esto si, como ya fue dicho, los pensamientos causan la conducta? Otro rasgo de la relación causal descrita por Crane es que tampoco dice nada acerca de la necesidad, es decir, que A cause B, no implica que lo cause con necesidad, sólo que es una conjunción regular. Del mismo modo, los pensamientos no implican con necesidad la conducta que representan. Un agente tiene la experiencia de modular su conducta, de elegir qué hacer y qué no.

Alguien podría objetar que existe una jerarquía de pensamientos, y que sólo el pensamiento que en un momento específico se encuentra en el primer lugar de la jerarquía se ejecuta. Como ocurre con los aviones en un aeropuerto: existe un protocolo para el despegue, y sólo puede despegar un avión por pista, pero en algún momento todos los aviones en la pista habrán despegado. Asimismo, eventualmente todos los pensamientos causarían el comportamiento que representan, en la medida que su posición en la jerarquía de ejecución cambie. Ahora bien, parece ser cierto que los individuos crean jerarquías de sus pensamientos o estados mentales, pero en cualquier caso, esa jerarquía emergería de la relación entre los distintos estados mentales de los individuos, es decir, no se trataría de un rasgo interno de los estados mentales. Por lo tanto, resulta adecuado concluir que, en lo que se refiere a su naturaleza interna, todos los estados mentales o pensamientos son causalmente equivalentes.

Lo anterior conduce al tercer supuesto. Si los pensamientos o estados mentales no causan con necesidad el comportamiento, y los agentes conscientes pueden tener la experiencia de que modulan a voluntad su conducta, se debe a que, a pesar de la imagen causal de los pensamientos, existe un vacío entre los estados mentales y la conducta. Esto es lo que Searle expresa con el término brecha [*gap*] (Searle, 2010, p.133), la separación que provoca la discontinuidad causal entre estados mentales y estados conductuales. La idea que el filósofo intenta expresar, es que un estado mental X no está conectado privativamente con un curso de acción Y , sino que X podría causar tanto el curso de acción Y como el curso de acción¹⁹ Z , o ninguno en absoluto.

Los tres supuestos han sido ya formulados por separado. Sería conveniente tratar de expresar el modo en que se articulan. Los agentes conscientes suponen que sus pensamientos causan estados conductuales (imagen causal de los pensamientos), y que existen fuertes razones para pensar que todos los agentes experimentan este tipo de causalidad, lo que constituye una poderosa herramienta a la hora de que los agentes entiendan y expliquen el comportamiento de terceros, y lleguen a saber todo lo que saben de sus mentes y las mentes de terceros (psicología de sentido común). Una de las cosas que los agentes saben de otras mentes es que toda acción no refleja es *causada* por un estado mental, pero que los estados mentales no siempre causan el comportamiento que representan, o lo causan del mismo modo (la brecha). De este modo se articula el esquema de supuestos que permite al experimento mental funcionar, es decir, aislar el conjunto de intuiciones respecto de lo que significa ser consciente.

Finalmente es posible hacer un plano general de la escena, mostrando en acción a los supuestos, los medios y el objetivo de *Cerebros de Silicona*. Sobre la base de los supuestos de la imagen causal de los pensamientos, la psicología de sentido común y la brecha, el experimento muestra como un conjunto de

¹⁹Debe entenderse como «curso de acción» al contenido representado por estado mental, y que constituye sus condiciones de satisfacción.

intuiciones, consistentes con estos supuestos, apoyan la tesis de que la conciencia es ontológicamente independiente de la conducta. Este experimento no demuestra que esta separación sea real, sino más bien organiza las intuiciones que conducen a sostener dicha tesis, por más problemática que pueda parecer.

Como efecto secundario de la interpretación mediativa del experimento, éste queda protegido de cierto modo. Esto ocurre por dos razones: en primer lugar, debido a que es complejo definir claramente qué se querría decir cuando se afirma que un experimento mental *demuestra* una tesis; en segundo lugar, porque permite sostener que, dentro de las posibilidades epistémicas de este experimento mental, la conclusión que produce es válida y ciertamente contribuye a los intentos de dilucidar con mayor claridad los aspectos metafísicos de la conciencia. En términos coloquiales, que una escalera no sea el medio adecuado para alcanzar la cima del monte Everest, no quiere decir que las escaleras sean inútiles para objetivos más modestos. Asimismo, que los resultados de los experimentos mentales no sean demostrativos, no quiere decir que no aporten importantes distinciones en el contexto de investigaciones filosóficas de este tipo.

En síntesis, el experimento es consistente con la interpretación mediativa que ha sido tratada en este apartado. Espero que haya quedado claro que cualquier interpretación más arriesgada, corre el peligro de ser objetada fácilmente al tratar de exceder los alcances ontológicos y epistémicos de estos dispositivos epistémicos.

2.3.3. Cerebros de Silicona y P3

El Tercer Principio Metodológico indica que la validez del experimento mental debe ser determinada en función de la validez de su forma argumental. Esta sección consiste en darle forma de argumento al contenido del experimento. Esta tarea será fácil debido a que ya se han formulado gran parte de los elementos que serán considerados en esta sección. En primer lugar, presentaré las premisas que conforman el argumento junto a la conclusión, para luego examinarlas detenidamente.

Propongo el siguiente planteamiento argumental:

Premisa 1. La conciencia, tanto en sus rasgos fenoménicos como causales, está estrechamente ligada a la biología del cerebro.

Premisa 2. El cerebro de un sujeto puede ser parcial o totalmente modificado por elementos sintéticos, como chips de silicona, que reemplazan el tejido cerebral de dicho sujeto.

Premisa 3. Desde la perspectiva de la primera persona, debido al reemplazo del soporte que ejecuta la conciencia, el sujeto constatará una disociación entre su experiencia consciente y su capacidad de conducir o causar su conducta.

Conclusión. En lo que respecta a la ontología de lo mental, la conciencia es independiente de la conducta, tanto en su control como producción.

Me parece que el relato del experimento ha quedado adecuadamente expresado en las premisas anteriores. A continuación, consideraré las premisas por separado, con el fin de justificar su uso en este argumento. Luego de esto, intentaré determinar la validez del argumento.

Lo primero será examinar la primera premisa:

Premisa 1. La conciencia, tanto en sus rasgos fenoménicos como causales, está estrechamente ligada a la biología del cerebro.

Hasta aquí se ha dicho que existe un conjunto de supuestos que respaldan las intuiciones que el experimento ocasiona. Es difícil trazar una línea clara entre lo que son supuestos, y las intuiciones que estos apoyan. Aquí he usado como criterio el nivel que poseen estas intuiciones, es decir, lo que he catalogado como supuestos son intuiciones más básicas, aunque no pretendo que esta distinción sea estricta. Dichos supuestos expresan en general un conjunto de rasgos fenomenológicos y causales. Con ellos se posee información de cómo es tener una mente consciente en términos cualitativos, y qué cosas son capaces de hacer las mentes conscientes. Ahora bien, la idea de que la conciencia está ligada estrechamente a la biología del cerebro, si bien es un poco más elaborada, descansa igualmente en observaciones concretas. En términos muy generales, hasta donde ha sido constatado por la biología, todos los individuos humanos a quienes se atribuye conciencia tienen cerebro. Gran parte de las atribuciones de conciencia que hacemos a otros organismos, se fundamentan en la presencia de las mismas estructuras, o estructuras de funciones análogas. Si se descubriera mañana una especie submarina muy extraña, por ejemplo, y se determinara que posee un cerebro análogo al humano y un sistema nervioso también análogo, con seguridad se concluiría que es consciente, que posee una mente. En el mismo sentido, mientras más complejos sean los cerebros que se descubran en otros seres vivos, las criaturas que los poseen serán consideradas poseedoras de conciencias más complejas. Es decir, se concibe como posible que tales criaturas tengan experiencias cualitativas similares a las experiencias conscientes humanas²⁰, o que incluso tengan lenguaje. Ejemplo de esto es lo que ocurre con los delfines, muchas personas expresan, en sentido metafórico, que son «los humanos del mar». El punto que intento establecer, es que cuando se habla de la presencia de conciencia, siempre se constata la existencia de cerebro. Si uno deja al margen la creencia de que las mentes o almas pueden subsistir separadas de los cuerpos, establecer que la relación entre la conciencia y el cerebro es estrecha, no es para nada gratuito; de hecho, a estas alturas del debate, se da por supuesto. Por lo tanto, existen las condiciones para concluir que la primera premisa ha sido adecuadamente establecida.

Lo segunda premisa expresa lo siguiente:

Premisa 2. El cerebro de un sujeto puede ser parcial o totalmente modificado por elementos sintéticos, como chips de silicona, que reemplazan el tejido cerebral de dicho sujeto.

Existen dos razones para justificar esta premisa. En primer lugar, como ha sido mencionado en reiteradas ocasiones, no existe ninguna razón *a priori* por la cual este procedimiento no pueda ser realizado. En segundo lugar, ya se ha implementado en ratones la implantación de una prótesis de memoria que, al parecer, ha tenido buenos resultados. Posiblemente, sólo es cuestión de tiempo para que sea testeado en humanos. Es complejo tratar de aventurarse en relación con los resultados de un experimento de este tipo, sin embargo, en principio debería ser igualmente exitoso. Destaco nuevamente que, a pesar de la factibilidad empírica mencionada, lo esencial de esta premisa descansa en su posibilidad lógica, rasgo suficiente para considerarla como premisa válida.

Premisa 3. Desde la perspectiva de la primera persona, debido al reemplazo del soporte que ejecuta la conciencia, el sujeto constatará una disociación entre su experiencia consciente y su capacidad de

²⁰Hago esta comparación porque hasta donde es sabido la conciencia humana es la forma más compleja de conciencia.

conducir o causar su conducta.

Estas es la premisa más fuerte que es posible extraer del experimento mental. En estricto rigor, se trata de una conjetura con base en el supuesto que la primera premisa expresa, una especie de corolario, esto es, que la naturaleza biológica del cerebro está estrechamente relacionada con la existencia de conciencia. Por esto, es previsible que la experiencia consciente se vea trastocada si el soporte biológico la sustenta es alterado. Desde esta perspectiva, me parece que queda justificada la consideración de esta premisa. Nuevamente, cualquier refutación de la misma sólo puede ser presentada *a posteriori*.

Conclusión. En lo que respecta a la ontología de lo mental, la conciencia es independiente de la conducta, tanto en su control como producción.

La fuerza de esta conclusión, es también el conjunto de intuiciones que apoyan la separación postulada por la tercera premisa. Si bien la cadena de razonamiento ha conducido con rigor a esta conclusión –como ya fue dicho– este argumento no puede aspirar a implicaciones materiales. Es decir, la separación aquí defendida sólo se sostiene a modo conceptual, y se presentan las razones que apoyan con validez dicha distinción en ese nivel. Corresponde al análisis empírico de la conciencia verificar si la separación es también real. Una tarea como esta va más allá de los alcances de esta investigación y, me parece también, de la filosofía.

He intentado hacer la presentación argumental del experimento de Searle del modo más conciso posible. Esto no implica que no sea posible un análisis o revisión ulterior de lo que ha sido propuesto en este apartado, sino más bien que se trata de un análisis incompleto pero consistente. En este sentido, el argumento presentado parece válido, las premisas usadas se justifican en el contexto de esta investigación, y la conclusión se sigue de tales premisas. Siguiendo a Norton, es posible afirmar que, dado que se trata de un buen argumento, *Cerebros de Silicona* es un buen experimento mental.

2.4. Conclusión

El objetivo de este capítulo ha sido, junto con presentar exhaustivamente el experimento, aplicar los conceptos y principios obtenidos a partir de la investigación realizada en el capítulo uno. Hasta aquí, el total de esta investigación ha consistido en la preparación y empleo de las herramientas que se han considerado como mínimas a la hora de enfrentarse a experimentos mentales en filosofía de la mente. El siguiente capítulo consistirá en el examen de las conclusiones que Searle extrae de este experimento, y de cómo desarrolla una postura metafísica particular de la conciencia. Pero sería útil recapitular lo dicho en este capítulo, antes de dar el siguiente paso.

Tras presentar el experimento, fue establecido que tanto el planteamiento como la propuesta general de la conclusión, resistían una reducción al absurdo. La conclusión extraída de dicha exposición, fue que sólo era posible presentar objeciones empíricas a la propuesta del experimento. En otras palabras, que no existían razones *a priori* que permitieran descartarlo.

En la segunda sección, el objetivo consistía en mostrar cómo una interpretación «mediativa» era no sólo posible, sino también necesaria. A partir de lo presentado en el capítulo uno, se estipuló que los experimentos mentales sólo pueden ser interpretados en filosofía como medios de exposición de tesis, las que se caracterizan por ser audaces y complejas de extraer de las premisas e intuiciones que

las respaldan. Las implicaciones que proponen sólo pueden ser aceptadas *con seguridad* a un nivel conceptual. En síntesis, sólo constituyen medios de presentación y comprensión de tesis filosóficas, y en ningún caso representan un tipo de demostración, en el sentido fuerte del término.

Finalmente, en el tercer apartado, se dio forma argumental al experimento a partir de la *tesis de reducción* extraída del planteamiento de Norton. El uso de cada premisa fue justificado, y se determinó que la conclusión era consistente con ellas. Se destacó el carácter de este argumento, esto es, que la conclusión sólo puede ser defendida a nivel conceptual y no empírico. Sin perjuicio de lo anterior, la separabilidad conciencia/conducta postulada por esta tesis, posee toda la validez teórica posible.

Implicaciones para la ontología de la conciencia y el problema de las otras mentes

3.1. Introducción

Hasta aquí se han expuesto y aplicado el conjunto de herramientas diseñadas para el examen riguroso de las implicaciones que John Searle pretende extraer a partir del experimento mental Cerebros de Silicona. En lo que resta de esta investigación, se tratarán en detalle dichas conclusiones y se determinará cuán aceptables son.

Es notable la riqueza teórica que el filósofo norteamericano logra desarrollar con base en esta sencilla propuesta. Ciertamente, la totalidad de intuiciones adquiridas proporciona bastante apoyo a una cierta perspectiva sobre lo mental. El objetivo específico que ha motivado todo el despliegue metodológico de este escrito consiste en, de la mano de estas intuiciones, formular condiciones necesarias, aunque probablemente no suficientes, para la adscripción de conciencia a un sistema, ya sea orgánico o sintético. En efecto, es factible diseñar, con el conjunto de conclusiones de Cerebros de Silicona, una lista de comprobación [*checklist*] para justificar las atribuciones de conciencia que frecuentemente uno efectúa. Esta propuesta puede parecer extraña a primera vista, después de todo, en la vida diaria nadie cumple el rol de «inspector de mentes» y anda por el mundo verificando las condiciones que le permitirían atribuir mente a otros; más bien, es algo que se da por supuesto. Este emerge, se vuelve visible, cuando se pretende determinar si un animal o una máquina sofisticada poseen mente. De algún modo, uno sabe quién o qué posee una mente, pero no está muy claro cómo se llega a saber. El objetivo de este capítulo consiste en, una vez tratadas las conclusiones, presentar dos condiciones de adscripción de conciencia que, en principio, determinarían su ocurrencia real en un sistema.

El plan de presentación de este capítulo contempla tratar las conclusiones presentadas por Searle divididas en dos conjuntos. El primer conjunto está compuesto por las implicaciones ontológicas que el experimento posee, y el segundo conjunto está compuesto por las implicaciones de orden epistémico. Aunque las ontológicas tienen mayor importancia en el contexto de esta investigación, existen algunas importantes consideraciones epistémicas estrechamente vinculadas a los criterios de atribución que vale la pena exponer, y de las cuales se han hecho algunos comentarios que serán recogidos aquí. A continuación, formularé la lista de las condiciones de atribución de conciencia. Como ya fue dicho, no se pretende que el conjunto establezca condiciones suficientes para dicha atribución. Tras presentar y validar estas condiciones, expondré una objeción que es posible realizar en su contra al considerar algunas observaciones de Dennett.

3.2. Conclusiones para la ontología de la conciencia

Como ha sido recordado en varias ocasiones, esta investigación puede ser caracterizada como una de tipo ontológico. Es decir, su objetivo general es indagar y descubrir los rasgos fundamentales del modo de existencia de los fenómenos mentales conscientes y su relación con el cerebro y la conducta. Arriba se ha dicho que Searle pretende defender una tesis metafísica fuerte, pero intuitivamente sostenible. Esta tesis es la conclusión central de su planteamiento. El filósofo lo expresa del siguiente modo: “*Ontológicamente hablando, el comportamiento, el rol funcional, y las relaciones causales son irrelevantes para la existencia de los fenómenos mentales conscientes*”²¹ (Searle, 1992, p.69, énfasis del autor). Es necesario destacar el punto clave de esta afirmación: cuando Searle habla de la irrelevancia del comportamiento, debe entenderse estrictamente que dicha irrelevancia sólo es real en relación con la ontología de la conciencia. El filósofo mismo precisa su declaración señalando que: “[...] los fenómenos en cuestión pueden existir completamente y tener la totalidad de sus propiedades esenciales independientemente de cualquier resultado [*output*] conductual”²² (Searle, 1992, p.69). Pero intenta desglosar esta proposición a modo de un argumento que, según él, sería aceptado por aquellos filósofos que sostienen tesis contrarias a las suyas.

Mediante una astuta estrategia de consenso, el filósofo norteamericano presenta las siguientes proposiciones (Searle, 1992, p.69):

1. Los cerebros causan los estados mentales conscientes.

2. Hay alguna clase de conexión conceptual o lógica entre los fenómenos mentales conscientes y la conducta externa.

Sin embargo, sobre la base de lo planteado por el experimento mental, la conjunción de estas dos proposiciones no puede ser mantenida si se quiere aceptar una tercera proposición:

3. La capacidad del cerebro para causar la conciencia es conceptualmente distinta de su capacidad para causar el comportamiento motor. Un sistema podría tener conciencia sin comportamiento y comportamiento sin conciencia.

Como puede notarse, [2] y [3] son consistentes con [1] pero inconsistentes entre sí. Lo que [2] expresa es una relación fuerte entre la conciencia y el comportamiento, a tal punto que podría ser presentada como identidad. Si la conciencia y el comportamiento comparten su origen causal, y existen y dejan de existir de manera simultánea, es posible que sean la misma entidad, siguiendo la ley de Leibniz. Pero más arriba se ha mostrado por qué sería un error identificar o reducir la conciencia al comportamiento. Si se acepta como válido el experimento, en función de toda la investigación anterior, [2] no puede ser aceptada. Por lo tanto, la aceptación de [3] está justificada por su coherencia con el experimento.

Si la conjunción [1] y [3] es válida (y Searle piensa que es evidentemente válida), están dadas las condiciones para formular un principio ontológico de la conciencia. Searle ha llamado a esta máxima “*el principio de independencia de la conciencia y el comportamiento*”²³ (Searle, 1992, p.69, énfasis

²¹ “*Ontologically speaking, behavior, functional role, and causal relations are irrelevant to the existence of conscious mental phenomena*”.

²² “[...] the phenomena in question can exist completely and have all of their essential properties independent of any behavioral output”.

²³ “the principle of the independence of consciousness and behavior”

mío). Este principio tiene dos corolarios inmediatos: el primero afirma que el comportamiento es insuficiente para determinar la presencia de conciencia; el segundo afirma que, ante la presencia *real* de conciencia, la existencia o inexistencia de *outputs* conductuales no se da con necesidad. Si el lector recuerda la presentación del experimento realizada en el segundo capítulo, notará que el escenario dos describe exactamente lo señalado por el primer corolario. Del mismo modo, el segundo corolario es descrito por el tercer escenario. En síntesis, es posible el comportamiento sin conciencia, y es posible la conciencia sin comportamiento.

Ahora bien, lo presentado hasta aquí parece ser bastante consistente. Sin embargo, como fue adelantado en el apartado dedicado a la ontología subjetiva, esta disociación ontológica puede presentar algunos problemas. De hecho, para ser mucho más preciso, no se trata de que la posición apoyada hasta aquí tenga alguna inconsistencia interna, sino que dicha posición aparentemente no encaja con la ontología aceptada por lo general en la ciencia.

La penúltima sección del primer capítulo, presenta el problema de la ontología subjetiva y algunas dificultades epistémicas asociadas a ella. Aquí es necesario completar esa discusión antes de dar paso a las conclusiones epistémicas del experimento. En aquella sección se sostuvo que existen dos clases de hechos en el mundo: hechos objetivos y hechos subjetivos. Estos se definen como hechos que existen con independencia de los observadores, el caso de los objetivos, y hechos que existen sólo vinculados a un observador, el caso de los subjetivos. Esta distinción no supone ningún problema en primera instancia, después de todo, no parece haber malos entendidos cuando en el día a día nos enteramos de que alguien está enfermo, o que los padres de Javiera están llenos de orgullo por los éxitos de su hija, o que morder un limón puede ser bastante desagradable. Todos estos hechos son subjetivos, si los agentes que los experimentan desaparecieran, los hechos mismos dejarían de existir para siempre. Además, la medicina, por ejemplo, es un tipo de conocimiento científico que en gran parte se dedica al estudio de las enfermedades (que sólo existen asociadas a un agente), y todo el mundo parece estar de acuerdo en que un ejercicio de ese tipo es válido. Pero, ¿puede existir una ciencia del orgullo o de la sensación de acidez?

Esta pregunta pone en contraste a estos dos ejemplos con el primero, cuando en un comienzo se presentaron como hechos análogos. La clave de esta distinción radica principalmente en la reducción. Aunque desde la perspectiva de un paciente su deseo para eliminar el dolor puede ser el motivo por el cual asista al médico –y ciertamente para el médico la escala de dicho dolor determina de algún modo la urgencia del tratamiento–, parece no haber problemas en tratar la enfermedad como nada más que cierta alteración de los procesos biológicos que le permiten al individuo mantenerse vivo y sin malestares. En otras palabras, si se le administra un fuerte calmante al paciente el dolor desaparecerá, pero todos los médicos coincidirán en que eso no implica que el paciente se haya recuperado; sigue enfermo pero no experimenta dolor. Por el contrario, si fuera posible determinar con total exactitud que un paciente no posee absolutamente ninguna patología, pero incluso así afirmara experimentar un fuerte dolor, bastaría con administrar un calmante o un placebo. ¿Qué quiere decir todo esto? Que hay un conjunto de hechos subjetivos que pueden ser reducidos a hechos objetivos sin dejar atrás nada relevante o crucial. A pesar de la doble vida de algunos conceptos mentales, como Chalmers señala (1999, pp.41-49), es posible inclinar la balanza hacia un aspecto del fenómeno en perjuicio del otro.

Espero que parezca clara la distinción que está siendo realizada con los hechos restantes. Cuando se trata del orgullo y la sensación de acidez, parecería que no existe la misma disposición hacia la reducción. Estas experiencias no serían adecuadamente tratadas si se afirmara que no son nada más que el efecto de cierta actividad cerebral. Lo que importa en estos casos es el cómo se siente estar orgulloso, o cómo se siente la acidez del jugo de limón, es decir, el *cómo* fenomenológico más bien que el *por qué* biológico. La importancia de esta experiencia subjetiva radica en que tiene consecuencias epistémicas. Thomas Nagel ha sido quien ha señalado enfáticamente este rasgo: “La razón es que cada fenómeno subjetivo está esencialmente conectado con un único punto de vista, y parece inevitable que una teoría física objetiva deje de lado ese punto de vista”²⁴ (Nagel, 1974, p.220). Este punto de vista es el punto de vista de la primera persona. Esto quiere decir, que existen ciertos hechos subjetivos que sólo pueden ser *observados* desde la experiencia misma del fenómeno, debido a que cualquier forma de reducción que lo haga accesible a la perspectiva de la tercera persona, lo elimina.

Antes de concluir esta sección es necesario hacer una aclaración. Cuando se ha dicho que ciertos hechos no son accesibles a la perspectiva de tercera persona, por ningún motivo se está señalando que son una especie de misterio insondable, ni tampoco que la perspectiva tradicional del estudio científico sea errada. Lo que he intentado sugerir, es que la subjetividad ontológica debe ser aceptada como parte de la imagen objetiva de la realidad, y que son necesarios intentos serios por adaptar la perspectiva de la primera persona a los criterios de objetividad científica. Después de todo, la realidad siempre ha resultado ser más rara de lo que usualmente se ha previsto.

3.3. Conclusiones para la epistemología de la conciencia

Cerebros de Silicona tiene también interesantes lecciones que ofrecer en relación con el modo en que se conoce la conciencia, la de otros y la propia. A propósito de esto, es conveniente establecer desde el principio que el comportamiento, a pesar de su irrelevancia ontológica, es el principal elemento que permite conocer los estados mentales de terceros. Cuando más arriba fue mencionada la psicología de sentido común, fue precisamente porque en ella se organiza una inabarcable cantidad de conocimientos sobre cómo se supone que la gente actúa y, por lo tanto, qué conducta se espera de ellos en contextos específicos. La idea de que el comportamiento hace visibles los pensamientos (hasta el momento totalmente invisibles) es la que da sentido a la psicología de sentido común. Este hecho es consistente con el modelo empirista de estudio de los estados mentales. El objetivo de este apartado será mostrar cómo es que, a partir del planteamiento *searleano*, este modelo es inexacto.

La clave del empirismo en filosofía de la mente es que una aproximación científica a la conciencia sólo es posible sobre una base empírica. Esto supone la existencia de una conexión entre los eventos mentales de un sistema y su conducta. No me parece necesario intentar precisar este punto, debido que, a lo largo de esta investigación, se han presentado varias razones para mostrar que este supuesto es correcto. Por lo tanto, si estuviera en discusión la atribución de estados mentales a un sistema, el medio empírico para cerrar el debate es el estudio del comportamiento de dicho sistema. Claramente,

²⁴ “The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view”

este estudio supone algunos criterios de racionalidad mínimos, sobre todo cuando se trata de animales. Todo esto parece bastante razonable, y está en sintonía con lo que podría llamarse «metodología general de la ciencia».

El problema aparece cuando consideramos la otra mitad de la evidencia. Esa otra mitad es el conjunto de estados mentales que claramente no tiene un correlato conductual. Existen varios ejemplos de estos estados mentales: mi creencia en que la velocidad de la luz es 300.000 k/s, la intención que tiene un testigo de ser honesto con el juez, entre otros. Esto se suma a lo que el tercer escenario del experimento mental ilustra con claridad.

En una sección previa, se dijo que este tipo de experiencias conscientes no conductuales eran tan inmediatas y evidentes que parece inadecuado intentar demostrarlas. Todo agente consciente posee un acceso privilegiado a una parte sus eventos mentales que, por la ontología de tales eventos, son inaccesibles a otros. Si un tercero afirmara que tales experiencias no existen, a uno le parecerá que está claramente equivocado, pero no tendría medios para demostrar que se equivoca. En última instancia, la única prueba es que ese tercero acepte que él también las experimenta.

Pero, ¿por qué sistemáticamente se ignora esta otra mitad de evidencia? Desde el análisis que John Searle presenta, la causa de esta confusión es una ambigüedad en el uso del término «empírico». Al igual como ocurre con los términos «objetivo» y «subjetivo», existen dos sentidos en los que aquél puede ser entendido. El primer sentido es ontológico, esto es, cuando se afirma que un hecho es empírico, lo que se quiere significar es que se trata de un estado de cosas real en el mundo, contingente, en oposición a los hechos morales, o los hechos matemáticos, que existen de modo abstracto. El segundo sentido es epistémico, es decir, afirmar que un hecho es empírico quiere decir en este caso que es universalmente accesible, que todos los observadores aptos pueden testarlo. Dicho esto, parece clara la dificultad que surge al usar el término sin aclarar la distinción. Searle lo expresa de la siguiente manera: “Ahora, esta ambigüedad sistemática en el uso de la palabra “empírico” sugiere algo que es desde luego falso: que todos los hechos empíricos, en el sentido ontológico de ser hechos mundanos, son epistémicamente igual de accesibles a todos los observadores competentes”²⁵ (Searle, 1992, p.72). Si el lector ha seguido con atención la línea argumental que se ha desarrollado hasta aquí, coincidirá en que hay hechos en el mundo que no son igualmente accesibles para todos los observadores, principalmente por la naturaleza misma de los hechos. El caso más notable, es el que fue ilustrado en la sección anterior: las experiencias subjetivas.

Ahora bien, esta oposición necesita ser resuelta para poder formular la lista de criterios para la atribución de conciencia que se pretende aquí. Si el empirismo tradicional en filosofía de la mente es verdadero, y sólo el comportamiento es evidencia confiable de los estados mentales, esta investigación queda en un callejón sin salida, y la solidez de los argumentos presentados debe ser cuestionada. Pero, si las razones aquí presentadas han sido adecuadamente establecidas, es necesario abandonar el empirismo tradicional. Esto último es lo que sostengo. Por lo tanto, resulta necesario adaptar el empirismo a la naturaleza del fenómeno que se desea conocer.

²⁵ “Now this systematic ambiguity in the use of the word “empirical” suggests something that is certainly false: that all empirical facts, in the ontological sense of being facts in the world, are equally accessible epistemically to all competent observers”

La razón que está en la raíz del error empirista, es que supone que la realidad es, por decirlo de alguna manera, «transparente» a los observadores. Pero si uno se detiene en esta idea, de inmediato surge la pregunta: ¿por qué la totalidad de los fenómenos deberían ser igualmente accesibles a los métodos empíricos? Tal como señala Searle, tal supuesto es obviamente falso (Searle, 1992, p.73). Para que sea posible un estudio científico de la conciencia, es necesario incluir los hechos subjetivos en sentido ontológico, dentro del marco de la objetividad epistémica y abandonar el mito de la realidad «transparente».

Los dos conjuntos de conclusiones expuestos hasta aquí, son los principales aportes que el experimento mental realiza al análisis metafísico de la conciencia. Probablemente no suponen ningún tipo de novedad para el debate en Filosofía de la Mente, pero sin duda constituyen un esfuerzo por fundamentar y sistematizar un conjunto de ideas aisladas sobre la naturaleza de lo mental. Este ha sido el objetivo general de esta investigación, proporcionar un marco de análisis para *Cerebros de Silicona* que permitiera juzgar con fundamento los supuestos, deducciones y conclusiones que proporciona. Lo que resta ahora es concretar el objetivo específico. En el siguiente apartado, presentaré los dos criterios de atribución de conciencia formulados a partir del experimento mental.

3.4. Criterios de atribución de conciencia

A continuación serán propuestos y fundamentados dos criterios de atribución de conciencia, que tienen como fin enunciar condiciones necesarias, aunque no suficientes, para la atribución de conciencia. La estrategia de exposición de esta sección intentará ser lo más detallada posible.

Para evitar rodeos innecesarios en la presentación de los criterios, es conveniente enunciarlos de una vez:

Criterio 1. Un sistema consciente debe exhibir conducta intencional e inteligente.

Criterio 2. Un sistema consciente debe poseer estructuras o sistemas físicos análogos a los humanos, en donde se ejecuten los procesos mencionados en el Criterio 1.

No conlleva mucho esfuerzo notar la oposición que puede existir entre lo expuesto por los criterios y el resto de esta investigación. Sin duda el principal problema se encuentra en el Criterio 2, ¿acaso no fue establecido que la conducta era ontológicamente irrelevante para la existencia de conciencia? En efecto, esta fue conclusión más fuerte de *Cerebros de Silicona*. En lo que sigue, quedará claro que no existe contradicción alguna, y que tanto el criterio como la conclusión ontológica son lógicamente compatibles.

Para empezar a aportar claridad, expondré el propósito de estos criterios. Como fue señalado, su objetivo es proporcionar condiciones de atribución de conciencia. Ahora bien, ¿de qué tipo es esta tarea? Claramente se trata un objetivo práctico. En otras palabras, lo que está siendo requerido es un método que permita determinar cuándo un sistema tiene conciencia y cuándo se comporta como si tuviera. Esto hace alusión al problema de las otras mentes, y un poco más abajo me referiré a él. Lo necesario ahora es precisar mi anterior afirmación, para determinar qué estoy diciendo y qué no.

Considérese el siguiente escenario. Usted es un reclutador de guardias de alta seguridad. Su tarea consiste en seleccionar los mejores candidatos para ocupar las vacantes disponibles. Pero resulta que la

cantidad de postulantes al puesto es muy masiva, y le resulta imposible determinar uno por uno si los candidatos cumplen con las condiciones físicas requeridas para el puesto. Para solucionar el problema, usted recurre a los conocimientos que ha adquirido a lo largo de su carrera, y llega a la conclusión de que los mejores guardias que ha seleccionado cumplen estas dos condiciones: miden sobre 1.6 metros y tienen espaldas anchas. Resulta entonces que usted revisa las fichas de postulación y solo convoca a los postulantes que miden sobre 1.6 metros. El día de la reunión hace que se formen y selecciona aquellos que tienen las espaldas más anchas. De este modo agilizó sorprendentemente su proceso de selección, y cuenta con una gran cantidad de candidatos posiblemente competentes para el puesto de guardia de alta seguridad.

El punto que intento establecer es que, no importa si sus criterios son precisos o no, porque es posible que un porcentaje de los seleccionados resulte ser pésimo en su puesto, o que uno de los descartados haya sido un eventual excelente guardia; lo importante es que sus criterios le permitieron solucionar un problema práctico que determina su actitud hacia cada postulante. El segundo punto es que sus criterios no son antojadizos, debido que responden a un conjunto de intuiciones adquiridas y respaldadas por su experiencia. El tercer punto, es que usted no pretende que estos criterios sean exactos o suficientes a la hora de determinar quién es mejor guardia y quién no. Usted piensa que estos son criterios mínimos para tomar a alguien en serio como postulante al puesto. Y por supuesto, usted no tendría ningún problema si eventualmente apareciera un científico con un método tan eficiente como el suyo pero más certero, y en su selección postulantes con una estatura menor a 1.6 metros fueran seleccionados.

Este relato ilustra claramente el fin de los criterios de atribución de conciencia. En la práctica, existen situaciones en las que se debe juzgar si un sistema posee conciencia, pero es necesario que sea de una manera eficiente, ya que se corre el riesgo de estar cometiendo un gran error de implicaciones serias. En general, determinar la presencia de conciencia tiene consecuencias, la actitud de uno hacia los sistemas conscientes está llena de matices éticos, matices que no existen en la relación con los sistemas inertes y no conscientes.

Pero, ¿qué ocurre con el problema de las otras mentes? Precisamente ese el punto. Este problema sólo existe en el análisis abstracto o filosófico del estudio de la conciencia, en la práctica, no hay tal cosa (Searle, 1992, p.77). Cada uno asume que las personas con las que se relaciona poseen una mente consciente y que su comportamiento es un resultado de sus procesos mentales. En gran parte, esa atribución espontánea descansa sobre los criterios que han sido expuestos arriba. Esto quiere decir, que no existe ninguna objeción empírica que nos prohíba realizar esta atribución tan radical, a pesar de los problemas filosóficos. Esta no es una actitud aislada. Hay otros temas importantes en filosofía, que son problemáticos, pero con los que diariamente uno lidia. Estoy pensando en la existencia de Dios, cierta noción de justicia, o que las palabras tengan significados, etc. Estos tres temas poseen interesantes y complejos problemas si nos aproximamos filosóficamente a ellos, pero la existencia de tales problemas no impide que la gente sea creyente o atea, que tenga opiniones fuertes sobre lo que es junto o no, y que se exprese con la seguridad de que está diciendo cosas que tienen sentido. Esta es la distinción que mencioné al iniciar este apartado: cuando uno se dedica a la filosofía debe tener claro que, ontológicamente hablando, el comportamiento es irrelevante para la conciencia; pero

cuando pensamos con sentido práctico el comportamiento y el soporte físico de ese comportamiento, son absolutamente relevantes para la atribución de conciencia. En filosofía, el objetivo es aislar las propiedades esenciales de este fenómeno tan intrigante. Cuando se piensa con sentido práctico, el objetivo es juzgar sobre hechos concretos en el mundo. Es por esto que no existe contradicción alguna entre lo dicho a propósito del comportamiento antes, y lo propuesto aquí. Simplemente se trata de ámbitos distintos y adyacentes.

Pero parece haber un vacío. Cuando se trata de establecer si los sistemas complejos no humanos tienen conciencia, uno se encuentra frente a un puzle casi imposible. Las preguntas que surgen son del tipo: ¿Sobre qué base se justifica nuestra atribución de conciencia a tales sistemas? ¿Existe sólo un tipo de conciencia? ¿Cómo reconocer tipos de conciencia distintos de la conciencia humana? Estas son sólo algunas de las tantas posibles preguntas que pueden ser enunciadas.

Lo que pretendo argumentar es que, a pesar de las desafiantes preguntas que una teoría sobre las mentes no humanas necesita responder, existen criterios que permiten realizar tal atribución a los sistemas complejos no humanos.

Estos criterios ya fueron presentados más arriba. ¿En qué se justifican? Hasta donde es sabido, las mentes conscientes se ejecutan principalmente en cerebros y otras partes del sistema nervioso. No existe, hasta donde estoy enterado, ninguna evidencia sobre seres humanos conscientes pero sin cerebros, o árboles consciente, o mesas conscientes, o computadores conscientes. Estos hechos son muy elocuentes, porque permiten establecer una relación estrecha entre cerebros y mentes. Algo de esto fue dicho más arriba, en el segundo capítulo. Además, cuando un sistema se comporta intencionalmente y de modo inteligente, se le atribuye conciencia en función de dicho comportamiento.

Sin embargo, persiste el dilema del autómatas introducido por Descartes, es decir, puede ser que un sistema tenga un comportamiento que satisfaga el criterio de intencionalidad e inteligencia, pero que sea un robot programado para tales fines.

Si los criterios antes presentados se consideran de modo separado, existen claras deficiencias en la función que cumplen. Sin embargo, esto sólo dice que son insuficientes por sí solos. Si, al contrario, son considerados en conjunto, proporcionan una base consistente de condiciones necesarias. A pesar de que intuyo que están cerca de ser suficientes, me parece que no poseo actualmente evidencia para apoyar una afirmación tan pretenciosa.

Existe un pasaje de John Searle en el que sugiere la idea anterior:

“Supóngase que los marcianos llegasen a la tierra y concluyésemos que tiene estados mentales. Pero supóngase que cuando abriésemos sus cabezas se descubriese que todo lo que había allí dentro era una mucosidad verde. Bien, con todo, la mucosidad verde, si funcionase de manera tal que produjese conciencia y el resto de su vida mental, tendría que tener poderes causales iguales a los del cerebro humano” (Searle, 1985, p.48).

Lo que ocurre con los marcianos, vale para los animales y para los robots. Si su comportamiento sugiere que existe una vida mental tras él, el siguiente paso es determinar que las estructuras que explican ese comportamiento potencialmente consciente, son estructuras que tienen poderes causales idénticos a los poderes causales que causan la conciencia en humanos.

Para finalizar este apartado, quiero insistir en el eje de esta afirmación. Es cierto que especulativamente uno puede argumentar cierta insuficiencia lógica en los criterios de atribución, esto quedó expuesto en las preguntas que fueron enunciadas más arriba. No obstante, la base de mi propuesta es que, las razones empíricas y los fines prácticos para dicha atribución, tienen la fuerza argumental suficiente para imponerse a los reparos filosóficos; y que por lo tanto, tales criterios representan una posición válida para ser sostenida.

3.5. Una objeción a los criterios propuestos

Existe un conjunto de objeciones filosóficas que pueden ser presentadas contra la propuesta aquí desarrollada, y creo haber expuesto algunas razones que permiten descartarlas sin miedo a pensar que se está dejando algo crucial de lado. Pero hay al menos una objeción adicional que parece no encajar en la categoría de “objeciones filosóficas”, que vale la pena considerar antes de cerrar esta investigación. Me refiero a algunas de las ideas expuestas por Dennett con relación a los sesgos implícitos a los criterios generales de atribución de conciencia. Con criterios generales me refiero no específicamente a los expuestos más arriba, sino a una lista que reúne las opiniones más extendidas de la gente en relación con las mentes.

La primera noción a considerar para dar forma a esta objeción, es separar el término «conciencia» del término «mente». En el desarrollo del debate en filosofía de la mente, se puede decir que se han manifestado dos tendencias respecto a la naturaleza de lo mental. Por una parte, están aquellos que consideran que la conciencia es un fenómeno tan central y complejo, que el objetivo de las investigaciones sobre la mente puede ser resumido en determinar la naturaleza de los fenómenos mentales conscientes. Esto es lo que Chalmers llama el problema difícil de la conciencia, el que enuncia en la siguiente pregunta: “¿Por qué todo este procesamiento está acompañado por una vida interna que experimentamos? A veces se ignora por completo esta pregunta, a veces se la posterga para algún momento futuro y a veces simplemente se la declara resuelta.” (1999, p.16). El filósofo australiano aclara que, si bien los problemas acerca de la cognición y el procesamiento e integración de la información son importantes, no existe ninguna complejidad metafísica en la base de dichos problemas. De hecho, en las últimas décadas se han obtenido interesantes resultados que han contribuido a dar forma a un modelo básico de la mente humana. A pesar de este desarrollo, las experiencias conscientes cualitativas de la vida mental humana se resisten a las reducciones electroquímicas o informáticas que estos modelos de explicación proponen.

Por otra parte, están aquellos que consideran que los fenómenos conscientes, si bien son evidentemente raros en el contexto de un sistema de procesamiento físico, son simplemente una interfaz de usuario llena de colores y sonidos que hacen más amistosamente manejable al conjunto de comandos que cumplen funciones específicas en el *wetware*. Sin embargo, los procesos realmente importantes se ejecutan en segundo plano y son los que deben ser estudiados con mayor detención. Más aún, esta interfaz es absolutamente irrelevante para el funcionamiento del sistema y, para todos los efectos importantes, puede ser ignorada. Lo interesante de esta postura, es que el procesamiento no es exclusivo del cerebro o el sistema nervioso, sino que estos son solo una parte del sistema total de

procesamiento mental, que incluye ciertamente la totalidad del cuerpo. En el mismo sentido Dennett señala lo siguiente cuando se refiere a este modelo alternativo:

“[...] la de ver al cerebro (y con él, la mente) como un órgano entre muchos, un usurpador del control relativamente recién llegado y cuyas funciones no pueden comprenderse adecuadamente hasta que no lo veamos como el patrón sino solamente como uno de los sirvientes, un tanto díscolo, que trabajan en el fomento de los intereses del cuerpo que lo alberga, lo alimenta y da significado a sus acciones” (Dennett, 2000, p.96).

Ahora bien, ¿tiene mente una planta según los criterios de atribución? No, no la tiene. No hay conducta inteligente ni estructura anatómica causalmente análoga a la humana. Pero, ¿tiene mente una planta tras considerar el sesgo temporal de estos criterios? Posiblemente. Si uno imagina que las reacciones de las plantas pueden ocurrir a una velocidad mil veces menor que las humanas, reacciones en las que se puede apreciar cómo interactúan con los estímulos y responden a ellos para adaptarse, uno pensaría que son organismos muy inteligentes. De pensar que son inteligentes a atribuirles mente, sólo hay un paso.

El chovinismo de escala espacial funciona de modo similar. Los organismos por los cuales se suele preguntar si tienen o no mente, son de cierto tamaño, de tal modo que, respecto de los organismos más pequeños que ese cierto tamaño, no surge la pregunta por si tienen o no vida mental. Uno puede pensar que un hámster tiene mente, pero probablemente descartaría la posibilidad si se tratara de una pulga, por ejemplo. Pero, ¿qué pasaría con esta atribución si la pulga fuera del tamaño de un perro y el hámster tuviera el tamaño de una bacteria? Pienso que claramente la atribución se vería afectada por la modificación de la escala, tal vez se descarte que el hámster posea una mente y se le dé una segunda vuelta a la pregunta respecto a la pulga.

Espero que el lector coincida en que esta objeción no es precisamente una de carácter filosófico o metafísico, sino más bien una objeción práctica. Y que aunque arriba fue sostenido que los criterios no son adecuados para la atribución de mente, con base en el supuesto de la diversidad de tipos de mentes, los criterios siguen siendo útiles si se aplican a los sistemas de la escala temporal y espacial en la que los humanos se encuentran.

Conclusión

Los problemas que he intentado abordar ciertamente son de una complejidad que resulta imposible resolver en una investigación tan austera como esta. Sin embargo, para no producir la impresión de que los puntos centrales de esta discusión han sido evadidos, resulta necesario ponderar el alcance de los resultados obtenidos.

La etiqueta de esta investigación ha sido recordada en varios pasajes: se trata de una indagación ontológica. El objetivo general ha sido recoger las intuiciones que *Cerebros de Silicona* ofrece al respecto, y argumentar a favor de algunas distinciones conceptuales cruciales para el estudio de la conciencia. El objetivo de no ha sido intentar establecer si existen o no los estados mentales conscientes, o simplemente la conciencia. Esto restringe el alcance de la investigación, ya que en el contexto de una ontología de la conciencia, solo me he limitado a intentar determinar el *modo* de existencia de la misma. Como resultado de esta tarea, se ha argumentado a favor de dos afirmaciones: la primera es que la conciencia tiene un modo de existencia subjetivo, inaccesible al testeado en tercera persona, pero que dicha subjetividad no impide que se puedan establecer verdades epistémicamente objetivas sobre ella. La segunda afirmación declara que la conciencia existe de modo independiente al comportamiento, aunque esté de hecho ligada al él en los agentes conscientes. Esto último es un hecho de la biología.

Es claro que determinar el modo de existencia de la conciencia no resuelve la cuestión de *cómo* es posible que exista la conciencia en un universo completamente constituido por materia inerte y sin vida mental alguna. Pero, insisto, nunca fue el objetivo de esta investigación tratar de resolver dicha cuestión. Sin embargo, es posible que mediante la determinación de cuán distribuida está la conciencia en el universo y bajo qué condiciones, se logre allanar el camino hacia la resolución de dicha pregunta.

En conexión con lo anterior, el objetivo específico de esta investigación ha sido intentar proponer dos criterios de atribución de conciencia. El sentido de esta tarea no tiene que ver con asumir una actitud explicativa intencional sobre el comportamiento de algunos sistemas, sino con establecer de hecho si un agente es consciente o no. Obviando las implicaciones éticas y prácticas que estos criterios tienen como resultado, la atribución de conciencia puede ser vista como una tarea de carácter científico. No pretendo afirmar que lo que ha sido realizado a lo largo de esta investigación es un tipo de «ciencia de la conciencia», por ningún motivo. Simplemente intento señalar que las conclusiones que esta investigación ha aportado, pueden considerarse como la base conceptual para una aproximación científica a ella. Por esto, al final de este texto, me he volcado sobre la cuestión *práctica* de la atribución, por ser menos compleja, aclarando siempre que las dificultades *metafísicas* persisten.

A pesar, nuevamente, de las complejidades del «problema difícil de la conciencia», he intentado aproximarme a él aunque sea de manera oblicua, buscando resolver cuestiones más modestas como las tratadas en este trabajo. Sin duda las preguntas que yo he planteado y abordado –junto con otras mucho más interesantes– seguirán abiertas, y me parece que por mucho tiempo más.

Referencias bibliográficas

Ayer, A. J. (Edi.) (1965), *Positivismo Lógico*. Traducción de L. Aldama, U. Frisch, C. N. Molina, F. M. Torner y R. Ruiz Harrel. México, D. F.: Fondo de Cultura Económica.

Bechtel, W. (1991), *Filosofía de la Mente: una panorámica para la ciencia cognitiva*. Versión castellana de Luis Valdés Villanueva. Madrid: Tecnos.

Brown, J. R. (1991), *The Laboratory of the Mind*. New York: Routledge.

Chalmers, D. (1999), *La Mente Consciente*. Traducción de José Álvarez. Barcelona: Editorial Gedisa.

Crane, T. (2008), *La Mente Mecánica*. Traducción de Juan Almela. México, D.F.: Fondo de Cultura Económica.

Dennett, D. (2000), *Tipos de Mente*. Traducción de Francisco Páez. Madrid: Editorial Debate.

Descartes, R. (1982), *Discurso del Método*. Traducción de Antonio Gual Mir. Madrid: Edaf.

Descartes, R. (2011), *Meditaciones Metafísicas*. Introducción y traducción de Guillermo Graño Ferrer. Madrid: Alianza Editorial.

Hawking, S. W. (1989), *A Brief History of Time*. London: Bantam Press.

Hempel, C. G. (1993), *Filosofía de la Ciencia Natural*. Versión al español de Alfredo Deaño. Madrid: Alianza Editorial.

Nagel, T. (1974), What is it like to be a bat?, originalmente publicado en *Philosophical Review*. Reimpreso en Chalmers, D. (comp.), (2002), *Philosophy of Mind*. New York: OUP, pp.219- 226.

Norton, J. (1996), Are thought experiments just what you thought? *Canadian Journal of Philosophy*. Volume 3, Number 26, pp.333-366.

Searle, J. (1985), *Mentes, cerebros y ciencia*. Traducción de Luis Valdés. Madrid: Ediciones Cátedra.

Searle, J. (1989), *Mentes y Cerebros sin Programas*, originalmente publicado como *Mind and Brains*

Without Programs, en C. Blakemore y S. Greenfield (comps.), *Mindwaves*. Oxford: Blackwell. Reimpreso en E. Rabossi (comp.), *Filosofía de la Mente y Ciencia Cognitiva*. Barcelona: Paidós, pp.413-443.

Searle, J. (1992), *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.

Searle, J. (1997), *La Construcción de la Realidad Social*. Traducción de Antonio Doménech. Barcelona: Paidós.

Searle, J. (2001), *Lenguaje, Mente y Sociedad*. Traducción de Jesús Alborés. Madrid: Alianza Editorial.

Searle, J. (2003), *La Universidad desafiada*. Traducción y presentación de Edison Otero. Santiago: Bravo y Allende Editores.

Searle, J. (2010), *Making the Social World*. Oxford: OUP.
