



Universidad de Chile
Facultad de Filosofía y Humanidades
Centro de Estudios Cognitivos

Subjetividad, experimentos mentales y Ciencia Cognitiva

Una aproximación kuhneana

Luis Olea Peralta
Santiago de Chile, 2011

Tesis para optar al grado de Magíster en Estudios Cognitivos
Profesor guía: Rodrigo González

A mi familia, en cuyos hombros estoy parado

Agradecimientos

Las personas que nombraré a continuación han contribuido en diferentes momentos y de distintas maneras al desarrollo de esta tesis. Sin su colaboración habría sido muchísimo más difícil la concreción de este proyecto. Por esta razón, y teniendo en cuenta que mi memoria pueda olvidar a muchos otros, quiero agradecer encarecidamente a:

Rodrigo González, mi profesor guía, por su rigor e interés activo y genuino en el trabajo que he realizado durante este tiempo. Sus críticas, comentarios y correcciones han sido muy importantes en el desarrollo de la tesis.

Carolina Molina, bibliotecaria de la Facultad de Filosofía y Humanidades de la U. de Chile, por su generosidad, simpatía y disposición.

Aud Marit Fjerdingsren, por su indispensable compañía.

María y Hernán, mis padres, pilares de este proyecto.

Contenidos

Introducción.....	7
1 La conciencia	12
1.1 La ambigüedad del término conciencia: exploraciones preliminares	12
1.2 La conciencia y sus dos problemas	18
1.2.1 Conciencia visual: hacia la solución del problema fácil	24
1.2.2 Significado del estudio del problema fácil de la conciencia. Observaciones preliminares de los hallazgos neurocientíficos	25
1.3 La realidad no es solo realidad objetiva. El lugar de la subjetividad en la naturaleza y su carácter irreductible	27
2 La Ciencia Cognitiva como una concepción reduccionista de la mente.....	35
2.1 Psicología introspectiva, Conductismo y Ciencia Cognitiva	35
2.2 El paradigma computacional-representacional de la mente (PCR).....	37
2.3 El primer hito del paradigma computacional-representacional: la arquitectura clásica.....	45
2.4 Conexionismo: el turno de la metáfora del cerebro.....	49
3 Subjetividad, experimentos mentales y Ciencia Cognitiva. Una aproximación kuhneana.....	56
3.1 Concepción kuhneana del desarrollo científico	56
3.2 Definición y características de la ciencia normal	59
3.3 La anomalía en la ciencia normal. El preámbulo del descubrimiento científico.....	62
3.4 Características de las revoluciones científicas.....	65
3.5 La función de los experimentos mentales. Una visión de conjunto	68
3.6 El rol de la experimentación mental en la Filosofía de la Mente: el caso del murciélago de Nagel.....	78
3.6.1 ¿Es relevante la conciencia para el paradigma computacional-representacional de la mente?	84
3.6.2 ¿Por qué la conciencia es una anomalía si es un fenómeno familiar y cotidiano?	96
Conclusiones.....	102
Bibliografía	107

Introducción

La conciencia es un tema que, desde la segunda mitad del siglo XX hasta la actualidad, ha sido abordado por filósofos y científicos en el marco de la Ciencia Cognitiva, donde se incluyen disciplinas como la Filosofía de la Mente, la Inteligencia Artificial, la Neurociencia y la Psicología. Vital importancia en este debate han tenido los experimentos mentales (en adelante EMs) como herramientas analíticas que, desde la Filosofía, han contribuido a esta discusión. Considerando lo anterior, uno de los objetivos principales en la presente tesis es validar el rol de los EMs como un método relevante en el estudio y discusión de la conciencia.

La conciencia ha generado gran cantidad de bibliografía, por lo que su estudio se presenta problemático dadas las diferentes aproximaciones científicas y las distinciones conceptuales que dan cuenta de sus múltiples aspectos. Una de esas distinciones, hecha por David Chalmers (1996), ha tenido gran aceptación en la Filosofía de la Mente en las últimas décadas, estableciendo una división del estudio de la conciencia en dos partes, a saber, el *easy problem* y el *hard problem*. El primer problema puede descomponerse, a su vez, en tres interrogantes. En primer lugar, cómo el cerebro procesa la estimulación del medio ambiente; en segundo lugar, de qué manera el mismo cerebro integra la información sensorial, y, finalmente, cómo los individuos son capaces de entregar reportes de sus estados internos. La Neurociencia, por ejemplo, desde Broca y Wernicke hasta los estudios actuales de los correlatos neuronales de la conciencia, ha aportado información sobre las áreas de la corteza cerebral que actúan en la generación de este tipo de conciencia. La segunda parte del estudio de la conciencia, el problema difícil, trata de cómo el cerebro, aquella sustancia material, origina la experiencia subjetiva de un individuo con una mente y un punto de vista. El problema difícil es un elemento que forma parte constitutiva y esencial de la conciencia, pero que no puede ser abordado desde un enfoque objetivista tradicional dada su ontología de primera persona. Es decir, la existencia de la

conciencia subjetiva no es independiente del sujeto que la experimenta, por lo que la apariencia misma del fenómeno constituye su realidad (Searle, 1997: 456 y 2004: 68).

El presente trabajo aborda, desde un punto de vista histórico, el lugar que el estudio de la conciencia subjetiva ocupa y cómo puede ser descrito en términos de los problemas que presenta para el paradigma científico cognitivista. Para tal propósito, la teoría del desarrollo científico de Thomas Kuhn (1971) constituye un enfoque fundamental por su profunda comprensión de las características de la investigación científica.

Según Kuhn, el estado de una disciplina científica, en un momento dado, no se constituye sobre la base de la suma de los descubrimientos y teorías anteriores. Es decir, desde su punto de vista no hay un desarrollo acumulativo de los hallazgos y teorías, como desde una concepción tradicional podría entenderse. Por el contrario, la ciencia se desarrolla a través de la sucesión de modos incompatibles de investigación, fenómeno denominado *revolución científica*. Además, cada disciplina ejerce su actividad en el marco de un paradigma, término que se refiere a la “constelación de creencias, valores, técnicas, etc., que comparten los miembros de una comunidad dada” (Kuhn, 1971: 269). El paradigma determina los problemas a estudiarse así como las teorías y reglas para resolverlos. En este sentido, se pretende aplicar estas nociones en el análisis histórico de la Ciencia Cognitiva como disciplina científica conformada bajo un paradigma.

Concretamente, la hipótesis sostenida aquí plantea que el carácter subjetivo de la conciencia corresponde a lo que Thomas Kuhn llama *anomalía*, es decir, un fenómeno para el cual el paradigma computacional-representacional de la mente (PCR), en el marco de la Ciencia Cognitiva, no está preparado para estudiar. No lo está porque sus teorías, técnicas, metodología, etc. se sitúan desde la tercera persona objetiva, buscando describir las funciones implicadas en la cognición, lo que imposibilita cualquier aproximación a la conciencia en tanto fenómeno subjetivo. De esta forma, este paradigma es incapaz de dar cuenta de un hecho

que tiene lugar en la naturaleza y que usualmente es descrito como puro fenómeno subjetivo, sin consecuencia alguna.

Respecto del punto anterior, se intenta validar el EM de Nagel (1974) *What is it like to be a bat?* como una herramienta analítica importante en la discusión sobre la conciencia. Nagel sostiene que es posible tener una descripción neurofisiológica acabada de un murciélago, alcanzando un conocimiento completo de sus mecanismos funcionales desde la tercera persona. No obstante, este análisis ignora un elemento relevante: la *sensación de ser* ese determinado organismo. En efecto, no es posible imaginarse qué es ser como un murciélago. Esta dimensión fenoménica, el aspecto *what it is like*, es precisamente la que ignora un enfoque científico tradicional y que este EM identifica en el seno del paradigma computacional-representacional de la Ciencia Cognitiva. Por lo tanto, la función del EM *What is it like to be a bat?* es evidenciar la anomalía en el paradigma computacional-representacional de la mente, que corresponde a una concepción reduccionista de la realidad, a través de la identificación y problematización de la información inaccesible para el científico tradicional. El aspecto inaccesible es la experiencia subjetiva, y esto se explica porque el paradigma computacional-representacional estudia la conciencia en términos de *lo que hace*, es decir, considerando sus funciones y poderes causales, y no en cuanto a *lo que se siente* ser un determinado organismo. Este paradigma hegemónico en Ciencia Cognitiva corresponde a un estudio de lo mental enfocado en la descripción de las habilidades y las funciones. No obstante, cuando se trata de dilucidar el tema de la experiencia subjetiva, una explicación basada en estas consideraciones fracasa. El problema difícil persiste debido a que no es una cuestión acerca del desempeño de funciones, necesariamente. Va más allá de estas, porque aun al conseguirse la descripción de los procesos cognitivos, tales como la discriminación perceptual, la categorización, el acceso interno, el reporte verbal, etc., queda una pregunta sin responder, a saber, por qué estas funciones se acompañan de la experiencia subjetiva (Chalmers, 1995: 621).

En cuanto a la organización de los temas, la tesis se divide en tres capítulos. En el primero, se desarrollan los aspectos más recurrentes en la

discusión de la conciencia, abordando la ambigüedad de su significado y las distinciones conceptuales más importantes, en el contexto de una discusión filosófica, pero considerando los aportes de la Neurociencia. Se continúa con una descripción de la objetividad física, base de una concepción reduccionista de la realidad, y, finalmente, se considera el tema de la irreductibilidad del aspecto subjetivo.

En el segundo capítulo, se lleva a cabo una descripción del paradigma computacional-representacional de la mente, caracterizando sus fundamentos teóricos, su metodología y sus hitos más importantes. Por esta razón, se analizan las dos arquitecturas que configuran este paradigma, la clásica y la conexionista. El objetivo es caracterizar la Ciencia Cognitiva como una unidad que engloba varias disciplinas (Lingüística, Psicología, Inteligencia Artificial, etc.) que comparten rasgos en común, comprendidas en una concepción de la mente en términos computacionales y representacionales.

Finalmente, en el tercer capítulo, se desarrolla la hipótesis del presente trabajo, la cual justifica el rol de los EMs en Filosofía de la Mente como una herramienta de identificación de la anomalía en el paradigma computacional-representacional. En este sentido, se expone en detalle el punto de vista de Thomas Kuhn respecto del desarrollo de la ciencia en general, dando cuenta de sus conceptos clásicos, como los de *ciencia normal*, *paradigma*, *anomalía* y *revolución científica*, entre otros. Además, se expone un estado del arte de los EMs como género en ciencia y Filosofía, vinculándolo con las ideas kuhnianas y presentando a este tipo de experimentos como instrumentos importantes en el debate científico sobre la conciencia. En este marco, se explica en detalle el EM de Thomas Nagel (1974) y sus alcances. Adicionalmente, se abordarán otros EMs importantes en Filosofía de la Mente, como los de *Fred* y *Mary* (Jackson, 1982 y 1986), para sumarlos a la discusión del aspecto subjetivo y su relevancia.

Es importante aclarar que la presente tesis no solo intenta dar cuenta exhaustiva del problema de la conciencia. Aunque se desarrollan los temas centrales al respecto, el objetivo principal es abordar la conciencia subjetiva como un problema científico en el seno de la Ciencia Cognitiva. Por esta razón, la teoría

del desarrollo científico de Thomas Kuhn es central para este propósito. Con ello, se entrega un nuevo enfoque de la discusión sobre la conciencia desde una perspectiva histórica, que explique el lugar de la experiencia subjetiva y el significado que tiene dentro de las preocupaciones científicas del paradigma computacional-representacional de la mente.

1 La conciencia

1.1 La ambigüedad del término conciencia: exploraciones preliminares

Sin duda que al hablar de *conciencia* hay muchas cosas que se implican dependiendo del contexto. Es una palabra que se puede emplear tanto en situaciones comunes del diario vivir como en la ciencia o la Filosofía, cuyos ámbitos son más bien restringidos y especializados. Sin embargo, incluso allí su comprensión adolece de cierta ambigüedad. Por esta razón, se hace necesaria una exploración de las distintas acepciones del término *conciencia*.

Etimológicamente, la palabra *conciencia* deriva del latín *conscientia*. Su raíz está formada por *cum*, que es *con*, y *scio*, *conocer*. *Conscientia* era un saber compartido con otro o con uno mismo. También podía significar simplemente *conocimiento*. Zeman (2001) distingue, laxamente y de forma preliminar, tres significados para esta palabra. La primera acepción se refiere a la vigilia, es decir, al simple y cotidiano hecho de estar despierto como también a la capacidad de percibir e interactuar con el medio ambiente y las otras personas. Es posible medir este estado objetivamente, esto es, desde un punto de vista de tercera persona, obedeciendo a un criterio conductual. Este tipo de conciencia es gradual si se consideran los distintos estados de un individuo: la vigilia propiamente tal, el sueño y el estado de coma.

En segundo lugar, hay un sentido de conciencia relacionado con el fenómeno de la experiencia. Es una conciencia entendida como una percepción *de* algo o dirigida *hacia* algo, por tanto tiene una base intencional, aunque esto es discutido por Block (2002, véase más adelante). Esta segunda acepción se refiere a la subjetividad de lo experimentado. Los *qualia* son su esencia, término usado para denominar a aquellas características tradicionalmente descritas como inefables e indescriptibles de la experiencia individual, por ejemplo, el olor del vino, la intensidad del color de un objeto, esto es, propiedades que se perciben desde un punto de vista subjetivo difícil de medir y comparar.

Zeman presenta un tercer significado de la palabra *conciencia* que es equivalente al concepto de *mente*. En concreto, se alude al hecho de poseer un estado mental con un determinado contenido proposicional, es decir, a cualquier cosa que un individuo desee, espere, crea, tema, etc. Los deseos y las creencias respecto de algo o alguien son su fundamento. En comparación con las acepciones anteriores, que se relacionan o con la conducta o la experiencia, *ser consciente* en este tercer sentido puede utilizarse para informar el conocimiento que se tenga sobre cualquier asunto (Zeman, 2002: 20), por ejemplo, que un individuo esté consciente de que sus padres lo van a visitar el próximo fin de semana, de que la popularidad del presente gobierno ha bajado, de que su vecino tiene estrés, etc. Aquí “la esfera de la conciencia es simplemente la esfera de la mente: podemos describirnos a nosotros mismos como seres conscientes de cualquier cosa que pueda pasar a través de nuestras mentes” (Zeman, 2002: 21, traducción mía)

Al ejemplificar estos tres significados de conciencia, sería de la siguiente manera:

- 1.- Conciencia-vigilia: *Después del choque, el conductor perdió la conciencia.*
- 2.- Conciencia-subjetividad: *Soy consciente del olor de tus calcetines.*
- 3.- Conciencia-mente: *Soy consciente de que estoy colmando tu paciencia.*

Una disciplina como la neurociencia tradicional, la cual ha intentado encontrar los correlatos neuronales de la conciencia-vigilia, se puede relacionar con el primer significado expuesto por Zeman. Por su parte, la Filosofía de la Mente, en mayor medida, se ha ocupado del debate sobre la segunda acepción. En cambio, La tercera no ha logrado la misma atención por parte de filósofos y de científicos, quizá por abarcar un ámbito demasiado amplio para ser estudiado científicamente (Zeman, 2002: 21).

Block (2002), por su parte, también reconoce que el concepto de conciencia es un híbrido que contiene varias acepciones y que describe diferentes fenómenos. Para tratar de clarificar esta ambigüedad, Block distingue esencialmente dos significados. En primer lugar, se refiere a la conciencia

fenoménica [*Phenomenal Consciousness*], (en adelante, conciencia F) definiéndola como experiencia. Un estado es fenoménicamente consciente si existe algo que es como estar en ese estado, aludiendo a la dimensión *what it is like* desarrollada por Nagel (1974). Un individuo tiene estados fenoménicos cuando huele, oye, degusta alguna sustancia o siente dolor. Las propiedades de la conciencia F incluyen las propiedades de experiencia de la sensación, los sentimientos y las percepciones, pero también los pensamientos, los deseos y las emociones. Además, este tipo de conciencia tiene la característica de ser a menudo (quizá siempre) representacional. Para Block este tipo de conciencia también es distinto de cualquier propiedad cognitiva, intencional o funcional.

En cuanto al segundo sentido de conciencia, es decir, la conciencia de acceso [*Access-Consciousness*] (en adelante conciencia A), Block sostiene que una representación es consciente desde el punto de vista del acceso si es emitida [*broadcast*] para el libre uso en el razonamiento y también para el control racional de las acciones. Su carácter informativo [*reportability*] es uno de los rasgos más identificables, aunque no es necesariamente el más importante. El factor racional de este tipo de conciencia sirve para descartar acciones que implican control automático, tales como en el caso del *blindsight*, donde el individuo con un daño cerebral que afecta su visión en el primer nivel de procesamiento puede suponer, con un alto nivel de certeza, características visuales de estímulos que no son accesibles a él desde la percepción visual. Típicamente, los estados de conciencia A “son ‘estados de actitudes proposicionales’ tales como pensamientos, creencias y deseos, estados con contenido representacional expresados por cláusulas del tipo ‘que’ (por ejemplo, la creencia de que el pasto es verde)” (Block, 2002: 209, traducción mía). Dada esta descripción, la conciencia A coincide al menos con algunas características de la tercera acepción de conciencia que Zeman (2001) establece, es decir, la conciencia-mente.

Lo que pretende Block es hacer más precisa la noción de conciencia A con el propósito de considerarla un correlato, en el nivel de procesamiento de la información, de la conciencia F. La conciencia A, desde este punto de vista, es una buena candidata para una identificación reduccionista de la conciencia F

desde una aproximación funcionalista, aunque, ciertamente, posean diferentes características.

La distinción entre conciencia A y conciencia F surge de dos concepciones diferentes de la mente, a saber, la computacional y la biológica, respectivamente. El enfoque computacional supone que la mente puede ser explicada si se adopta la estrategia funcionalista, que considera la mente en términos del procesamiento de la información. Según esta aproximación, el nivel de abstracción para comprender la mente permitiría realizaciones múltiples, esto es, la mente podría implementarse en varios dispositivos físicos que funcionen, por ejemplo, sobre la base de un mecanismo hidráulico o eléctrico. En tanto, el enfoque biológico sostiene que el tipo de realización sí importa en la implementación de la mente y la conciencia. De esta manera, el sustrato biológico es irremplazable. Si la conciencia F es equivalente a la conciencia A, que puede implementarse en más de un dispositivo físico, la aproximación computacional de la mente es correcta. Sin embargo, si la naturaleza de la experiencia biológica es relevante, el tipo de realización de la mente sí importa, y, por esta razón, tanto las conciencias F y A serán, obviamente, distintas.

Block (2002) señala tres grandes diferencias entre ambos tipos de conciencia. Primero, y de modo general, el contenido consciente de la conciencia F está basado en la experiencia subjetiva, mientras el contenido de la conciencia A es representacional. Es una parte esencial de este último cumplir un rol fundamental en el razonamiento, porque en este solo el contenido representacional tiene lugar. Sin embargo, esta generalización puede ser un tanto injusta, pues muchos contenidos fenoménicos son también representacionales, y algunos de esos contenidos pueden, incluso, ser intrínsecos en ciertos contenidos fenoménicos. Adicionalmente, los estados de la conciencia A son necesariamente transitivos, en otras palabras, un estado de conciencia A debe ser siempre un estado de conciencia *de* algo. En cambio, los estados de conciencia F no siempre son transitivos. En este punto, existe una divergencia con lo planteado por Zeman (2001, véase el comienzo de este capítulo) respecto de la conciencia-subjetividad,

concepto equivalente al de conciencia F, puesto que Zeman la caracteriza como transitiva, esencialmente.

La segunda diferencia es que la conciencia A es una noción funcional, por tanto, relativa a un sistema. En este sentido, lo que un contenido de conciencia A hace es aquello que una representación de su contenido hace en un sistema. En contraste, la conciencia F no es funcional. Un contenido llega a ser consciente F [*Phenomenal conscious content*] en razón de lo que ocurre dentro del módulo de la conciencia F. No obstante, lo que convierte a un contenido en consciente A [*Access conscious content*] no corresponde a algo que pudiera estar dentro de un módulo, sino más bien a relaciones informacionales entre módulos. Cuando un contenido es consciente A, lo es en la medida en que alcanza el sistema ejecutivo, que está a cargo del control racional de la acción y del habla.

La tercera y final diferencia se relaciona con la posibilidad de clasificar un estado consciente F en una clase o tipo [*type*]. Es decir, distintas realizaciones [*tokens*] del dolor, por ejemplo, comparten un rasgo común presente en todas las manifestaciones del dolor. Por su parte, un estado consciente A en un momento dado podría ser inaccesible en un momento posterior, lo que lo haría imposible de clasificar en un tipo.

Adicionalmente a esta famosa dicotomía conciencia F/conciencia A, Block introduce las acepciones de conciencia reflexiva (en adelante conciencia R) [*Self consciousness*] y conciencia de monitoreo (en adelante conciencia M) [*Monitoring Consciousness*]. El alcance del primer término alude a la posesión del concepto de *self*, entendiéndolo como un concepto que denota la identidad de un individuo, y a la habilidad, por parte de este individuo, de usar este concepto para pensar acerca de sí mismo. Sin embargo, aunque en un primer momento las conciencias F y R parecen tener mucho en común, no se puede reducir la primera a la segunda. Son conceptos relacionados, pero distintos. La evidencia empírica muestra que los niños, por ejemplo, no adquieren una conciencia R sino hasta la segunda mitad de su segundo año de vida. Naturalmente, nadie podría decir que niños bajo esa edad no tienen estados fenoménicos frente al dolor o en otras situaciones. Si bien a menudo hay en los estados fenoménicos una referencia a la propia identidad,

una *me-ishness* en palabras de Block, puesto que el contenido fenoménico generalmente representa ese estado como un estado de *mí mismo*, el contenido fenoménico puede variar y no así el estado de conciencia R. A modo de ejemplo, al experimentar un individuo los colores rojo y verde no hay una diferencia en la orientación de la conciencia R, mientras que esos dos estados de percepción respecto de los colores son diferentes en la experiencia fenoménica. Se deduce, entonces, que una misma conciencia R experimenta dos estados fenoménicos diferentes y por ello no se puede reducir la conciencia fenoménica a la conciencia reflexiva.

Por otro lado, la conciencia M adquiere distintos significados. Uno de estos alude a un tipo de percepción interior, que podría corresponder a una forma de conciencia F de los propios estados o del *self*. La segunda noción está planteada en términos del procesamiento de la información, concretamente, este tipo de conciencia actuaría como un escáner interno. En tercer lugar, hay una noción metacognitiva que hace referencia a un estado consciente que, a su vez, es acompañado por un pensamiento que afirma que uno está en ese estado. Según Block (2002: 214), el concepto de conciencia M en sí mismo no es necesariamente erróneo, pero no debe confundirse con la conciencia F, porque al identificar ambas como lo mismo se propicia una postura eliminativista acerca de la conciencia F. De hecho, computadores del tipo *laptop* tienen integrado un mecanismo de monitoreo interno pero esto no los hace conscientes desde un punto de vista fenoménico. Block sostiene que la teoría defendida por Rosenthal (1997) sobre la conciencia, denominada HOT (*higher order thought*), da cuenta de la conciencia M. Tal teoría, precisamente, comete el error de confundir esta conciencia con la fenoménica. Lo que se plantea en la teoría HOT es que la conciencia es una materia de pensamiento acerca de estados mentales, es decir, la conciencia se analiza como conciencia *de* algo (Davies, 1999). En este sentido, ser consciente *de* algo se entiende en términos de tener un pensamiento acerca de ese algo. Sin embargo, siendo esto lo que enfatiza Block (2002), es obvio que la conciencia F sin HOT es posible y viceversa, al menos conceptualmente. Por ejemplo,

posiblemente perros y niños muy pequeños al sufrir de dolores no tienen pensamientos sobre ello.

El aporte teórico de Block sobre las diferentes acepciones del concepto de conciencia es muy significativo, ya que ayuda a clarificar qué se entiende por conciencia en determinados casos y, para efectos del presente trabajo, a evitar confusiones respecto de teorías y conceptos de conciencia que han intentado incluir su aspecto fenoménico de forma inapropiada. Precisamente, el concepto de conciencia F, descrito tal como lo hace Block, será una de las piedras angulares del presente trabajo y, en los capítulos siguientes, será referido, desarrollado y complementado con el aporte de otros autores.

1.2 La conciencia y sus dos problemas

En el estudio de la conciencia las distinciones conceptuales han sido bastante populares, no solo entre filósofos, como podría esperarse, sino también entre los científicos, como una estrategia para abordar su problemática. Dentro del primer grupo, particularmente fecunda ha sido la contribución hecha por David Chalmers (1996), quien divide el problema de la conciencia en dos, a saber, el *easy problem* y el *hard problem*. En primer lugar, el problema fácil (Chalmers, 1996: xi-xii) puede descomponerse en tres preguntas, a saber, cómo el cerebro procesa la estimulación del medio ambiente, de qué manera el mismo cerebro integra la información sensorial y, tercero, cómo los humanos son capaces de entregar reportes de sus estados internos. Tradicionalmente, tanto la Psicología como la Neurociencia se han ocupado de estas tareas.

Para ubicar el *easy problem* en un contexto más amplio, puede identificarse con el concepto de *enigma* científico, tal como lo entiende Thomas Kuhn (1971: 70). En este sentido, *enigma* alude a aquella categoría especial de problemas que sirven para poner a prueba la habilidad de un científico, y cuya solución está asegurada, o al menos puede generarse esa expectativa, porque el paradigma científico respectivo posee las herramientas apropiadas para resolverlo. Aquí entiéndase *paradigma* como el conjunto de creencias, métodos, técnicas, etc., que

comparten los miembros de una comunidad científica. Es decir, al catalogar este tipo de problemas como fáciles, Chalmers no quiere decir que son triviales, por el contrario, muchos de estos presentan grandes retos en el desarrollo de su investigación. Son fáciles en la medida en que el paradigma de una disciplina determinada, en este caso la Psicología, la Neurociencia o la Ciencia Cognitiva en su conjunto, posee las herramientas y los métodos necesarios para abordarlos y, de esta manera, generar una expectativa de solución. Propiamente tal, son aquellos problemas catalogados como científicos y que se enmarcan en lo que Kuhn llama ciencia normal.

En cuanto al *hard problem*, Chalmers (1996: xi-xii y 25) se pregunta por cómo y por qué la experiencia subjetiva acompaña a los procesos físicos del cerebro, es decir, como un sistema físico, en este caso el cerebro humano, origina la conciencia, entendiendo esta última como conciencia fenoménica. Este tipo de pregunta, a diferencia del *easy problem*, no ha sido seriamente abordada por la ciencia ni existe hoy en día una teoría aceptada unánimemente que la explique en detalle. Se podría decir que este problema no tiene paradigma que la sustente, en el sentido kuhneano del término. En otras palabras, no existen los métodos ni las herramientas adecuadas para abordarlo, y lo que se ha hecho tradicionalmente es ignorarlo, reducirlo o declararlo sin respuesta como si fuera un misterio o una ilusión. Sin embargo, paradójicamente, la subjetividad es uno de los aspectos centrales de la conciencia que la define en cuanto tal y que ha sido en las últimas décadas puesto en un lugar central por filósofos de la mente como Nagel (1974), Jackson (1986), Chalmers (1996), Searle (2004), entre muchos otros.

Teniendo en cuenta la concepción kuhneana del desarrollo científico, se puede considerar el *hard problem* de la conciencia como un término equivalente al concepto de *anomalía* en el contexto de la práctica científica. El concepto de *anomalía* es definido como un hecho que ha violado las expectativas de la práctica científica establecida y reconocida como tal, y para el cual el investigador, por consiguiente, no estaba preparado (Kuhn, 1971: 100). En un determinado momento, el conjunto de herramientas, métodos, teorías y creencias disponibles en una disciplina no sirve para abordar un determinado problema. Solucionarlo

parece demandar un reordenamiento tanto metodológico como teórico, debido a que no entra en el marco de expectativas de la comunidad científica. Es en este punto que la Filosofía juega un rol crucial, puesto que señala precisamente este resquebrajamiento de los métodos tradicionales y la aparición de un problema que merece ser considerado dentro de las preocupaciones de la ciencia, tal como lo hace Chalmers con esta identificación del *hard problem*. La función que ha tenido la Filosofía en el problema de la conciencia y su comprensión a partir del enfoque de Thomas Kuhn (1971) será tratado en profundidad en el capítulo 3.

Desde una perspectiva neurocientífica, Damasio (2002) divide el estudio de la conciencia también en dos partes, siendo esta distinción bastante similar a lo hecho por Chalmers (1996), aunque con algunos matices. Primero, es necesario establecer cómo se genera lo que él llama *la-película-en-el-cerebro*, metáfora para describir el compuesto unificado que es, a su vez, el resultado de la integración de diferentes imágenes sensoriales, tales como las imágenes visuales, auditivas, táctiles, olfativas, etc.; esto es, describir el proceso mediante el cual los inputs de los diversos sistemas sensoriales son integrados de forma coherente en un todo. El trabajo en neurociencia ha seguido esta dirección. Desde Broca y Wernicke, en el siglo XIX, hasta los actuales estudios de los correlatos neuronales de la conciencia han aportado información sobre qué partes de la corteza cerebral actúan en la generación de conciencia. Esto equivale al *easy problem* identificado por Chalmers (1996). En la mayoría de los casos, sin embargo, se aportan los datos de un sistema perceptivo en particular, no se considera el todo. Quizá lo que más se ha desarrollado en esta dirección ha sido el estudio de la conciencia visual. En las últimas décadas, el estudio del *binding* estableció un segundo nivel en este primer problema. El *binding*, definido ampliamente, es la “capacidad para integrar información a través del tiempo, el espacio, los atributos y la ideas” (Triesman, 1999: 105, traducción mía). En el sistema visual, específicamente, este tema se relaciona con el mecanismo mediante el cual las diferentes características de un objeto (forma, color, movimiento, ubicación) se asocian para entregar una imagen unificada de este. Pese a que el sistema visual ha acaparado una mayor atención, el *binding* puede entenderse también como una propiedad más general

en el procesamiento cognitivo. No basta con establecer cuáles neuronas se activan en un determinado proceso perceptivo específico, sino que debe plantearse cómo las informaciones visual, olfativa, auditiva, etc. convergen. Por lo tanto, con la metáfora de *la-película-en-el-cerebro* Damasio no hace otra cosa que plantear el problema en términos de un *suprabinding*, que asocia la información proveniente de los distintos *bindings* sensoriales. Roskies (1999), por su parte, esboza la posibilidad de que el *binding* sea la respuesta al carácter unitario de la conciencia que deberá ser confirmado por futuras investigaciones: “¿Son los mecanismos que nos permiten atribuir el color y la forma correctos a un objeto los mismos que nos conducen a la unidad de la experiencia fenoménica?” (Roskies, 1999: 9, traducción mía).

Según Damasio, el segundo problema de la conciencia es el *self*, esto es, la pregunta sobre cómo el cerebro, aquella sustancia material, origina el sentido de posesión para *la-película-en-el-cerebro*, es decir, la sensación de que un individuo tiene pensamientos, percepciones y emociones que sólo le pertenecen a este. Para Damasio (2002), esto ha constituido el problema más difícil de resolver hasta el momento. Su estudio debe complementarse con el primer problema, pues ambos están relacionados. Se trata, a fin de cuentas, de describir cómo se genera la subjetividad en *la-película-en-el-cerebro*.

Al tratar de expandir su definición de *self*, Damasio (2003) recurre a la introspección, en primer lugar. Así el *self* es descrito como un proceso que va desde lo más simple, esto es, la sensación de la existencia separada de otras entidades, hasta lo más complejo, a saber, una identidad configurada en razón de una variedad de detalles biográficos. Adicionalmente, Damasio sostiene que, combinando las perspectivas de la introspección y de la biología, se puede concebir el *self* como la representación de la individualidad y continuidad de un organismo viviente. El valor del concepto de *self* estriba en que funciona como una referencia para otros contenidos de la mente, tales como las representaciones de los objetos con los cuales los sujetos interactúan y las representaciones de los eventos en que estos mismos sujetos participan. Por tanto, el *self* se constituye en un verdadero protagonista para los objetos y eventos que tienen lugar en el

universo mental humano. De esta manera, es posible concebir las ideas de que tanto los objetos como los eventos se perciben desde una perspectiva específica, de que los procesos mentales pertenecen a alguien y, finalmente, de que se puede armar una autobiografía y reconstruir la identidad de un individuo incesantemente, con la ayuda de su memoria de los objetos y eventos pasados.

Este segundo problema planteado por Damasio, cuya solución es vista con escepticismo por la comunidad neurocientífica (Damasio, 2002: 8), puede ser relacionado con el *hard problem* de Chalmers (1996), así como el problema de *la película-en-el-cerebro* lo fue con el *easy problem*. Sin embargo, existen algunas diferencias. Exactamente, Chalmers se refiere a una forma más básica de subjetividad, aludiendo al aspecto *what it is like* planteado por Nagel (1974), el cual Block (2002), por su parte, define como conciencia fenoménica. En otras palabras, el problema difícil es cómo y por qué se origina la conciencia fenoménica de los procesos cerebrales. En cambio, Damasio es más específico porque se refiere al *self*, concepto equivalente al de conciencia reflexiva introducido por Block (2002). Como se señaló más arriba (véase 1.1), la conciencia reflexiva entraña el concepto de *self*, que denota la identidad de un individuo, y al uso de este concepto para pensar acerca de sí mismo. Pero hay que remarcar que conciencia reflexiva o *self* no son lo mismo que conciencia fenoménica, a pesar de que estén muy relacionados. No se puede reducir la conciencia fenoménica a la conciencia reflexiva (Block, 2002: 214), puesto que el sentido del *self* puede ser el mismo en estados cuyo contenido fenoménico es diferente. Un individuo puede experimentar dos estados fenoménicos distintos (por ejemplo, al experimentar dos colores como el rojo y el azul) bajo la misma experiencia de *self*. Naturalmente, tanto la conciencia reflexiva como la fenoménica aparecen a menudo juntas. Quizá una forma de conciliar ambos conceptos sea integrarlos en una idea más general de experiencia subjetiva, en donde se incluyan el *self* y los estados fenoménicos, conservando, ciertamente, las distinciones internas.

Respecto del estudio del *self* específicamente, las teorías en diferentes áreas que lo han intentado abordar se ubican en dos grandes grupos. Por un lado, las teorías del tipo *ego* y, por otro, las del tipo *bundle* (Blackmore, 2005: 66-69).

Las primeras consideran el *self* como un hecho natural que responde a nuestra intuición de ser individuos que tienen experiencias únicas. Este es el enfoque predominante de la Ciencia Cognitiva actual. En contraste, los partidarios de las teorías del tipo *bundle* lo consideran como una ilusión que se produce por el efecto de la memoria sobre un conjunto de impresiones obtenidas del ambiente. Según este enfoque, el *self* no es una entidad sino un haz de sensaciones que tiene un propietario ilusorio creado por la acción de la memoria. Aunque Blackmore piensa que este enfoque es válido, no hay un desarrollo mayor en Neurociencia ni en la Filosofía de la Mente. Como excepciones se puede nombrar al filósofo escocés David Hume y al budismo, pero este último desde una perspectiva religiosa.

Lo interesante del aporte de Damasio (2002) es que postula una hipótesis para el *self* desde la Neurociencia. Este autor piensa que es posible encontrar en el cerebro las estructuras que participan en la generación de la subjetividad. Su hipótesis plantea que el tálamo y la corteza cingulada mapean tanto el propio organismo como los objetos externos a éste. Así, se crea una representación de segundo orden, produciendo una interacción entre el cuerpo y el ambiente en la cual surge el sentido de propiedad sobre las percepciones. En este sentido, Damasio cree que la mente como tal se define en términos biológicos, aunque no niega otros niveles de descripción:

“No niego la existencia de la mente ni digo que una vez que sepamos lo que necesitamos saber sobre la Biología la mente deje de existir. Simplemente creo que la mente privada, personal, preciosa y única, efectivamente es biológica, y un día será descrita en términos tanto biológicos como mentales” (Damasio, 2002: 7, traducción mía).

Desde este punto de vista, lo biológico prima sobre cualquier otro enfoque de la mente. Es interesante el esfuerzo de Damasio por postular un anclaje cerebral del *self* y no plantearlo como un problema imposible de resolver para la ciencia, pues es un aspecto que siempre ha sido esquivo para la investigación al

no poder enfrentarse desde la tercera persona. Sin embargo, es necesario un desarrollo más profundo de esta hipótesis para determinar de qué forma se lleva a cabo ese “mapeo” de los ámbitos corporal y ambiental por parte de las estructuras cerebrales implicadas.

1.2.1 Conciencia visual: hacia la solución del problema fácil

Los estudios sobre conciencia tienen en el sistema visual un área en la que se ha avanzado bastante y se enmarcan dentro de lo que se ha denominado el correlato neuronal de la conciencia, esto es, la identificación de la actividad neuronal en los procesos conscientes. La pregunta pertinente en el tema de la conciencia visual es: ¿Qué regiones del cerebro se activan en el fenómeno? Y otras dos preguntas complementarias a la primera: ¿Hay zonas que tienen una mayor participación que otras? ¿Es una actividad que compromete a toda la corteza cerebral? Hay algunos experimentos que han permitido esbozar algunas respuestas.

Logothetis (2002) afirma que en la corteza cerebral solo una pequeña fracción de neuronas participa en la conciencia visual. La evidencia se obtuvo del experimento *binocular rivalry* aplicado en monos. Este experimento se basa en la percepción de ambigüedad visual provocada por la presentación simultánea de dos objetos a un observador. Ambos objetos compiten por ser percibidos conscientemente y, por tanto, cada uno de estos registraría una actividad neuronal diferente cuando la atención se centra en uno o en el otro. Cuando un sujeto se enfrenta a una situación como esta, su cerebro se vuelve consciente, primero, de la percepción de un objeto y luego del segundo, produciéndose una alternancia. Un cambio de percepción reportado por el sujeto significaría actividad consciente. Lo que intenta este experimento es trazar la actividad neuronal que corresponde a cada cambio de reporte. A partir de los resultados del *binocular rivalry*, Logothetis concluye que gran parte de la actividad neuronal no tiene una directa relación con la conciencia, pues solo una pequeña fracción de neuronas se activa en el proceso. Solo la corteza temporal inferior (CTI) muestra una fuerte actividad que bordea el 90%. Sin embargo, esto no significa que una región específica sea la

causante de la conciencia visual. El tipo de neuronas encontradas en CTI podrían estar en otras regiones como la corteza prefrontal y el lóbulo parietal, como lo señalan algunas pruebas en humanos. Por lo tanto, el correlato neuronal que da lugar a la conciencia estaría distribuido por toda la vía visual, aunque no de forma homogénea. Para Tononi y Edelman (1998), en el área CTI hay una concentración de esas neuronas que funcionaría junto con otros grupos neuronales ubicados en otras partes, formando un conjunto mayor¹. Estos autores realizaron sus propios experimentos de *binocular rivalry* en humanos y encontraron correlaciones entre activación de neuronas y estados perceptuales conscientes solo en algunos grupos en las cortezas frontal, occipital y temporal.

Por su parte, Krick y Koch (2002: 16) coinciden con Logothetis en señalar que no hay evidencia para afirmar la existencia de un lugar específico para la conciencia visual en el cerebro. Al parecer, la actividad neuronal que la provoca está distribuida por toda la neocorteza pero en distinta concentración según el área.

1.2.2 Significado del estudio del problema fácil de la conciencia. Observaciones preliminares de los hallazgos neurocientíficos

Según lo expuesto en las secciones 1.2 y 1.2.1 de este capítulo, se puede concluir que el estudio de la conciencia puede dividirse en dos problemas. Primero, determinar cómo la información sensorial de los distintos sistemas (olfativo, visual, auditivo, etc.) es integrada para producir una percepción unitaria. En segundo lugar, determinar cómo se genera la subjetividad, entendiendo esta en un sentido amplio, es decir, incluyendo tanto la conciencia fenoménica (Block, 2002) como el *self* (Damasio, 2002 y Block, 2002), y de qué manera se vincula con el primer problema.

¹ Tononi y Edelman (1998) usan estos datos de los estudios sobre el sistema visual para respaldar una teoría global de la conciencia, cuya hipótesis central es el *dynamic core*, la cual consiste en la activación de subgrupos neuronales en la generación de conciencia. Estas neuronas variarían su actividad dependiendo del tipo de acción ejecutada. Por ejemplo, algunas neuronas implicadas en la conciencia visual podrían no activarse en otro tipo de tareas inconscientes o que requieren una ejecución automática.

El estudio del primer problema ha tenido un mayor desarrollo que el segundo por cuestiones metodológicas. El estudio del *self* o la conciencia fenoménica se presenta como una cuestión compleja, pues es difícil acceder a este fenómeno desde la tercera persona, fundamental en la investigación científica tradicional. Lo que se ha hecho más bien es un intento por trazar los correlatos neuronales de la conciencia, tal como Damasio lo ha hecho con su hipótesis sobre la base neuronal del *self* (2002). En cuanto al problema más fácil, es decir, la convergencia de la información sensorial y al sentido unitario de la conciencia, se han alcanzado interesantes avances en los estudios de la visión. La información disponible sobre esta área (Logothetis, 2002; Krick y Koch 2002) apoya la hipótesis de que en la conciencia visual participan menos neuronas de las que se pensaba, aunque probablemente distribuidas por toda la neocorteza. Este hecho respalda la teoría de la conciencia propuesta por Tononi y Edelman (1998), que plantea que en las tareas que implican conciencia se activan ciertos grupos de neuronas organizados en conjuntos y que están repartidos por todo el cerebro. Este sistema de subgrupos de neuronas de la conciencia tiene un carácter dinámico, es decir, carecen de ubicación fija; se trata más bien, de un funcionamiento que implica la activación de neuronas dependiendo de una determinada tarea. Así, un grupo de neuronas puede activarse para servir a un procedimiento pero para otros no. Sin embargo, es necesario investigar qué ocurre con otros tipos de conciencia perceptiva, como la auditiva, olfatoria, gustativa, etc. para tener una comprensión más completa. En este punto, el estudio del *binding*, definido como la integración de la información sensorial, podría explicar cómo se genera la unidad de la conciencia. El *binding* de cada sistema sensorial debe ayudar a una comprensión del fenómeno más general de la integración de información, planteado en términos de un *suprabinding*.

En cuanto a la hipótesis de Damasio (2002) sobre el *self*, esta necesita ser comprobada con experimentos y un desarrollo mayor. Representa un avance importante en el estudio de los fenómenos subjetivos desde un punto de vista neurocientífico, pues generalmente esta comunidad se ha mostrado escéptica sobre el éxito de tal empresa. Lo interesante de lo planteado por Damasio (2002)

es que trata de responder al problema del *self* desde una dimensión biológica, haciendo una descripción detallada de las estructuras corticales participantes en ese fenómeno, como el tálamo y la corteza cingulada. Sin embargo, esto está lejos de trascender el mero correlato neuronal de la experiencia subjetiva, pero constituye una base ineludible. El fenómeno de la conciencia es mucho más complejo, pues tiene una ontología de primera persona (Searle, 2004: 67-68), en otras palabras, es un fenómeno subjetivo, una experiencia, y una teoría que pretenda explicarla desde la tercera persona objetiva se sitúa lejos del fenómeno o simplemente lo pierde de vista. Es posible que un neurocientífico tenga acceso a la actividad neuronal de un organismo, pero carece de la experiencia de ser ese organismo, carece, en definitiva del fenómeno que ocurre en la conciencia de ese individuo (para un mayor desarrollo de este punto, véase 1.3 y capítulo 3).

Por otro lado, en las últimas décadas, ha habido un intento de usar los datos de la Neurociencia para integrarlos en teorías más globales de la mente, como la arquitectura conexionista, incluida dentro del paradigma computacional-representacional de la mente (véase capítulo 2), que se relacionan con el procesamiento de la información, pero en general estas perspectivas, aunque inspiradas en los hallazgos de la Neurociencia, también han ignorado, relegado o reducido el aspecto fenoménico. Uno de los principales principios de estas aproximaciones es considerar la mente como un sistema computacional, asemejando las capacidades cognitivas humanas a un sistema de capacidades computacionales. En este sentido, cuando se usa la metáfora del computador se entiende la mente simplemente como un mecanismo de procesamiento de la información, siendo reducido, también, el aspecto fenoménico.

1.3 La realidad no es solo realidad objetiva. El lugar de la subjetividad en la naturaleza y su carácter irreductible

Aunque el *hard problem*, término que refiere a la conciencia fenoménica, fue acuñado por David Chalmers (1996), ya había sido descubierto antes por Thomas Nagel (1986: 29), quien lo sitúa en una discusión más amplia sobre la objetividad y

el conocimiento de la realidad que es posible alcanzar bajo sus parámetros. Nagel se pregunta cómo, dada nuestra comprensión objetiva de la realidad física, la disposición de ciertos materiales también físicos y básicos dan lugar a las extraordinarias capacidades de un organismo viviente y, al mismo tiempo, originan un ser con una mente, un punto de vista y experiencias subjetivas que no tienen cabida en la concepción tradicional de la objetividad.

A propósito de la objetividad física y del lugar que los seres con sensibilidad tienen en un mundo dominado por este punto de vista, Nagel (1986: 13-27) se pregunta si la mente puede ser realmente comprendida de manera objetiva. Respecto de la concepción física de la objetividad en sí misma, es necesario establecer que se ha desarrollado como parte de un método de llegar a una comprensión más verdadera del mundo físico, el cual, a su vez, se presenta de forma inexacta en la percepción sensorial. Semejante método se desarrolla en estados sucesivos que van instalando un mayor nivel de objetividad cada vez. Como primer paso, se trata de establecer que la percepción humana es causada por la acción de las cosas sobre los sistemas sensoriales de las personas. Lo siguiente es darse cuenta de que las propiedades físicas que provocan las percepciones en las personas deben separarse de su apariencia perceptual, puesto que esas mismas propiedades también causan efectos en otras cosas del mundo físico diferentes de los humanos, por tanto, sin causar ningún tipo de percepción. El tercer paso es establecer una concepción de la verdadera naturaleza de las propiedades físicas independientemente de la apariencia provocada en un individuo. En el fondo, no se trata solo de eliminar cualquier punto de vista propio sobre la realidad, sino también desacreditar cualquier punto de vista humano general basado en las percepciones. En otras palabras, las cualidades secundarias, tales como las apariencias, las sensaciones, los sabores, etc., son relegadas del conjunto que constituye el mundo externo, mientras que las cualidades primarias, tales como la forma, el peso, el tamaño y el movimiento, adquieren un papel estructural en ese mundo.

Este método recién descrito ha sido muy exitoso en la ampliación que se tiene del mundo físico, recibiendo la ayuda de teorías y explicaciones que se valen

de conceptos apartados de cualquier punto de vista perceptual y humano. Esto último es muy importante, porque significa que es posible llegar a un conocimiento objetivo de la realidad sin tener la necesidad de poseer un sistema sensorial, con tal de que se tenga la racionalidad y una comprensión de la matemática y de las propiedades formales de la concepción objetiva del mundo físico. Como consecuencia, se puede llegar a las mismas conclusiones con seres que tuvieran un sistema perceptual muy distinto al humano y que, por tanto, percibieran las cosas de una forma bastante diferente. Por supuesto, esto ocurriría en el caso de que esos seres tuvieran las mismas capacidades, tanto racionales como matemáticas, de los humanos.

Pese a que la concepción física de la realidad ha sido una herramienta muy poderosa para la ciencia logrando grandes avances, encuentra dificultades si se la enfrenta al reto de llegar a una comprensión *completa* de la realidad. La tradicional concepción física de la realidad no da cuenta de las percepciones y de las experiencias subjetivas en general, y, sin embargo, parecen existir en compañía de las otras criaturas de la naturaleza. En este sentido, hay más cosas en la realidad de las que pueden tener cabida en la concepción física de la naturaleza. A pesar de su obviedad, históricamente esto se ha seguido ignorando, debido a la fuerte atracción del enfoque fisicalista que ha predominado sin contrapeso. Esta concepción ha negado la existencia de todo lo que no puede reducirse a parámetros físicos, incluida la mente.

De acuerdo con Nagel (1986: 15-16), este punto de vista en el campo de la Filosofía de la Mente ha sido defendido por varias formas de reduccionismo (causal, conductista, funcionalista, etc.), teorías motivadas por un criterio epistemológico de realidad que establece que solo lo que puede comprenderse desde una particular concepción de la realidad objetiva existe en el mundo. Un ejemplo de esta idea errónea lo da la analogía entre la mente y el computador que ha tenido lugar en la Ciencia Cognitiva (como se ve en el capítulo 2), particularmente en el paradigma computacional-representacional de la mente. Las máquinas, sin experiencias subjetivas, que han sido diseñadas para ejecutar impecablemente tareas a los ojos de un observador que también realizan los seres

conscientes constituirían, según Nagel, una pérdida de tiempo y los verdaderos principios que subyacen a la mente deberían ser descubiertos por una aproximación más directa, y no desde la tercera persona objetiva.

Al dar una definición general de reduccionismo, se puede describir como “una posición que sostiene que las teorías o cosas de una clase pueden dar cuenta exhaustiva de teorías o cosas de otra clase” (McCauley, 1999: 712, traducción mía). Respecto de la Ciencia Cognitiva, un punto de vista reduccionista afirma que las teorías y hallazgos de la Psicología pueden ser descritos por la Neurociencia, de manera que los estados y procesos psicológicos no son más que procesos y estados corporales (McCauley, *ibídem*). En este sentido, el reduccionismo conduce a una forma particular de relación de identidad que Searle (1997) denomina la *relación nada-sino* [*nothing but relation*]. Desde este punto de vista, A puede ser reducido a B, si y solo si A no es *nada sino* B.

Searle (*ibídem*) identifica cinco tipos de reducción dependiendo de los elementos que son relacionados. La primera acepción de reducción es la ontológica, que es la más importante, y consiste en la demostración de que objetos de cierto tipo no son *nada sino* objetos de otro tipo. Por ejemplo, cosas como una silla, un árbol, un vaso, etc. no son *nada sino* un conjunto de moléculas. En segundo lugar, existe la reducción ontológica de propiedades, similar a la primera forma pero atañendo, como su nombre lo indica, a la forma de ser de los objetos. Por ejemplo, el calor de un gas no es *nada sino* la mera energía cinética de los movimientos de las moléculas. En tercer lugar, está la reducción teórica, que consta de una relación entre teorías, en las que las leyes de la teoría reducida pueden deducirse de las leyes de la teoría por la que se establece la reducción. El ejemplo clásico de este tipo lo da la reducción de las leyes del gas por las leyes de la termodinámica estadística. En cuarto lugar está la reducción lógica, común entre filósofos. Es una relación entre palabras y oraciones. Aquí estas refieren a un tipo de entidad que puede traducirse en aquellas palabras y oraciones que refieren a otras clases de entidades. Por ejemplo, oraciones sobre el fontanero promedio en Santiago pueden reducirse a oraciones sobre fontaneros específicos en Santiago. En el final está la reducción causal, definida en términos de una

relación entre dos tipos de cosas que pueden tener poderes causales. En esta última clase de reducción se demuestra que los poderes causales de la entidad reducida se explican por los poderes causales del fenómeno que prima en la reducción. Como ejemplo está el hecho de que algunos objetos sean sólidos y que esto tenga consecuencias causales, a saber, los objetos sólidos son impenetrables por otros objetos, siendo a la vez resistentes a la presión y otros fenómenos. Sin embargo, estos poderes causales se explican por los de los movimientos vibratorios de las moléculas de esos objetos.

Para el problema de la conciencia, las acepciones de reducción importantes son la primera, la segunda y la última esbozadas en el párrafo anterior, es decir, los dos tipos de reducción ontológica y la causal. Searle (1997) defiende una posición sobre el problema mente/cerebro que implica una reducción causal. Específicamente, sostiene que las características mentales son causadas por procesos neurobiológicos. Desde su punto de vista, la conciencia es una propiedad emergente de la interacción de ciertos tipos de neuronas, de la misma manera que la solidez de un objeto es una propiedad emergente de un sistema de moléculas. Pero la conciencia no puede ser explicada solo de la estructura física de las neuronas, debe haber una descripción adicional de las relaciones causales entre estas.

En general, las reducciones causales han conducido a reducciones ontológicas, como en el ejemplo del calor que fue reducido a movimientos moleculares o el sonido a ondas de aire. En estos casos, se trata de distinguir la realidad física y objetiva por sobre la mera apariencia, esto es, hay un intento por separar la realidad de un fenómeno de la simple apariencia provocada en un observador. De esta manera, la reducción causal lleva a una reducción ontológica por medio de una redefinición de la expresión que identifica el fenómeno reducido. En el caso de la conciencia no es posible llegar a una reducción ontológica desde su reducción causal (Searle, 1997: 453). Lo particular de la conciencia es que no puede ser reducida de la forma en que se hizo con el calor, la solidez o el sonido. A diferencia del calor, por ejemplo, en la conciencia no es posible hacer una distinción entre realidad físico objetiva y apariencia subjetiva, esto es, a la manera

de energía cinética de movimientos moleculares, por un lado, y la sensación de calor, por el otro. Esto se debe a que la realidad de la conciencia consiste precisamente en su apariencia. Si se elimina la apariencia se elimina, por tanto, la conciencia. O en palabras de Searle (1997: 456, traducción mía), “la apariencia es la realidad”. Tal particularidad hace entender la conciencia subjetiva como un fenómeno irreductible, de forma que Searle en este punto concuerda con Nagel (1974 y 1986). Es por esta razón que la Neurociencia tradicional, entendida como una disciplina que ha hecho una reducción ontológica de la experiencia fenoménica a simples correlatos neurales, y que se basa en la concepción de objetividad física criticada por Nagel (1986), no posee las herramientas necesarias para abordar el *hard problem*. Pero esta perspectiva reduccionista no solo está en la neurociencia tradicional, es más bien un enfoque más amplio que incluye a la totalidad de la práctica científica. Concretamente en el caso de la Ciencia Cognitiva, que abarca varias disciplinas incluyendo la Neurociencia, se puede hablar de un paradigma computacional-representacional de la mente que se ha valido del reduccionismo como método de aproximación a los procesos cognitivos (véase capítulo 2 para un desarrollo extenso sobre este punto).

Volviendo a Nagel (1986: 17-19), a pesar de que este autor critica y descarta la particular concepción física de la objetividad, no renuncia al ideal de objetividad en sí mismo. Esta puede replantearse y entenderse en otros términos para dar cabida a los estados mentales. La nueva pregunta, entonces, es si es posible llegar a un concepto de objetividad mental. Hay muchos puntos de vista pertenecientes a distintos seres y es imposible acceder a ellos mediante la imaginación humana, pues esta tiene límites. Por tanto, se deben buscar otros métodos de acceso que conduzcan a una mayor comprensión. Hipotéticamente, si existiera un individuo capaz de tener una imaginación ilimitada no sería necesario una forma de pensamiento que permitiera el acceso a la subjetividad contenida en otros seres, pero dado que eso no es posible, una forma objetiva de acceso a lo mental sería beneficiosa. Por otro lado, para llegar a un concepto objetivo de mente, se debería incluir a los seres humanos junto con sus estados mentales y su cuerpo, esto es, su realidad tanto mental como física, en una concepción que

no esté anclada, paradójicamente, de forma exclusiva en el punto de vista humano. En otras palabras, sería necesario estudiar la subjetividad humana exteriormente, aunque por cierto, en términos mentales, no físicos. ¿Cómo hacerlo? Nagel parece no tenerlo claro, pero al menos plantea que es una idea concebible en principio. Sin embargo, la objetividad replanteada bajo estos nuevos lineamientos tampoco garantiza una descripción completa de la realidad, puesto que hay aspectos que solo se comprenden a través de la experiencia subjetiva. En este sentido, Nagel tiene una posición escéptica sobre la objetividad mental también, aunque la posiciona como un mejor método para alcanzar una comprensión más acabada de la naturaleza. Al respecto señala: “No hay razón para asumir que el mundo como es en sí mismo deba ser comprendido objetivamente, aun en un sentido amplio. Algunas cosas pueden ser solamente comprendidas desde dentro” (Nagel, 1986: 18, traducción mía).

La nueva comprensión de lo mental en términos objetivos defendida por Nagel debe traspasar la distinción tradicional entre apariencia y realidad, de manera de incluir la existencia de las apariencias en una realidad más elaborada, no reducirla. Además, no debe ser una concepción antropocéntrica. La subjetividad de los estados mentales humanos tiene que ser concebida como una manifestación de una propiedad mental general, dejando de ocupar el lugar más importante, aunque, paradójicamente, ocupen el lugar principal en la mente de un individuo humano. En este sentido, es necesario enfatizar que el mundo no es el mundo de la especie humana. Las experiencias subjetivas de una persona deberían ser entendidas como realizaciones de una característica general de la naturaleza. Semejante comprensión de lo mental en términos objetivos todavía está sin desarrollo, y Nagel es consciente de que tampoco pueda lograrse alguna vez, pero significa un intento de expandir y mejorar un concepto de objetividad física que ha ignorado aspectos esenciales de la realidad y, al mismo tiempo, escapar del solipsismo y antropocentrismo de un concepto de subjetividad mal entendido.

Searle (1997), por su parte, aunque reivindica la irreductibilidad de la conciencia, no plantea que esta sea un fenómeno misterioso incapaz de ser

reducido alguna vez, a diferencia de la posición escéptica de Nagel (1986), para quien una objetividad planteada incluso en términos mentales sería incompleta. Searle, más bien, sostiene que los patrones *actuales* de reducción son incapaces de considerar un fenómeno como el de la conciencia subjetiva, pero no descarta que en el futuro haya una revolución intelectual que pueda alcanzar un tipo de reducción que sí dé cuenta de la mente. Este punto es interesante, puesto que se infiere que Searle entiende de una u otra manera que la conciencia ha significado una anomalía en un paradigma científico determinado, en este caso la Ciencia Cognitiva, lo que corresponde a un momento histórico en su desarrollo y no un problema insalvable para la ciencia futura. Su irreductibilidad no es más que la consecuencia trivial de una forma de hacer y pensar la investigación científica en un momento de la historia. Nada más. En el futuro se podría concebir un tipo de reducción que no eliminara el fenómeno en sí. De este modo, para Searle parece absurdo pensar que el problema de la conciencia significa un abandono de la ciencia y la adopción de una posición dualista. La discusión que se desprende de lo anterior es plausible de ser analizada desde el enfoque que Thomas Kuhn (1971) tiene del desarrollo científico, cuestión que será tratada en el capítulo 3, particularmente cómo la conciencia puede considerarse una anomalía en el desarrollo histórico del paradigma de la Ciencia Cognitiva y el rol de los experimentos mentales en la problematización de esa anomalía.

2 La Ciencia Cognitiva como una concepción reduccionista de la mente

Entender la mente en términos de un mecanismo computacional ha sido la estrategia predominante de la Ciencia Cognitiva, que puede ser incluida en un enfoque más amplio de la naturaleza denominada la concepción física de la objetividad (véase 1.3). En esta concepción, la experiencia subjetiva ha sido ignorada como elemento constituyente de la realidad. En el siguiente capítulo, se abordarán las características del paradigma computacional-representacional de la mente (PCR) como representante de tal enfoque, identificando sus fundamentos teóricos y metodológicos más importantes. En este sentido, no se trata de describir una disciplina en particular dentro de los estudios de la cognición, sino establecer los principios que guían la investigación en la Ciencia Cognitiva en general. De esta manera, disciplinas como la Lingüística, la Neurociencia, la Psicología, la Inteligencia Artificial, la Antropología y la Filosofía, que tienen sus propios métodos y teorías, serán analizadas respecto de sus rasgos compartidos, los cuales apuntan a una interpretación de la mente en términos computacionales y representacionales (Thagard, 1996: 19).

2.1 Psicología introspectiva, Conductismo y Ciencia Cognitiva

Para dar cuenta del surgimiento de lo que se ha denominado Ciencia Cognitiva, considerando su etapa clásica o hegemónica, esto es, lo que constituye el modelo o paradigma computacional-representacional de la mente (Thagard, 1996: 10; Vallejos, 2008: 63)², es necesario remitirse a sus antecedentes más inmediatos, a saber, la Psicología introspectiva y su contraparte, el Conductismo.

² Tanto Thagard como Vallejos no hablan estrictamente de “paradigma”, sino de “comprensión” [*understanding*] y de “modelo” computacional-representacional de la mente, respectivamente. Sin embargo, dada la relación que en el presente trabajo se pretende hacer entre teoría kuhneana del desarrollo científico y el problema de la conciencia en Ciencia Cognitiva, se ha preferido el término “paradigma” para referirse a tal concepción hegemónica de la mente.

La Psicología introspectiva es una disciplina decimonónica cuyo postulado fundamental es que el acceso a la conciencia es posible a través del autoanálisis que hace de esta el propio individuo. Este principio es atribuido a Wundt, psicólogo alemán del siglo XIX. Según Wundt, la introspección es un elemento válido para acceder a la mente. Específicamente, el método de la introspección consistía en entrenar a algunos individuos para analizar sistemáticamente sus propias experiencias mentales de manera de identificar los elementos implicados en las mismas (Bechtel, 1998: 14). Los fenómenos observados desde la tercera persona adquieren una menor relevancia en el estudio de la mente en favor de la introspección.

El Conductismo, en su variedad metodológica³, planteará exactamente lo contrario. Toda evidencia que no se someta al ámbito de lo observable es descartada y no cuenta para la ciencia, de esta forma no es considerada parte de la Psicología. Desde este punto de vista, hay un énfasis en el estudio del comportamiento observable entendiéndolo como una respuesta al estímulo. El método de la introspección, por tanto, pierde toda validez para la descripción objetiva. La Psicología, de esta manera, es entendida como una teoría de la conducta y la mente es *explicativamente irrelevante*. Por lo mismo, el Conductismo considera la Psicología una disciplina rigurosamente experimental, y a la introspección como un método inadecuado para entregar datos fiables acerca de lo que clásicamente se denomina *mente*.

Desde el enfoque conductista, la Psicología Cognitiva parece inviable, puesto que el Conductismo niega su objeto de estudio, a saber, la mente. Por tanto, se presenta un reto metodológico para la Ciencia Cognitiva, es decir, ¿cómo obtener evidencia de lo mental desde la observación objetiva? A la luz de estas consideraciones, cabe preguntarse cómo la Ciencia Cognitiva responde a este reto conductista de obtener evidencia desde la tercera persona, que es la principal crítica que se le ha hecho a la Psicología introspectiva anterior. La manera de

³ Existen, a grandes rasgos, dos variedades de Conductismo. En primer lugar está el Conductismo metodológico que, aunque no niega lo mental, plantea que no se puede describir. La mente, desde este punto de vista, es explicativamente irrelevante. En segundo lugar, está el Conductismo radical o lógico, que niega lo mental.

responder a esto se ha logrado a partir del modelamiento de la mente como un computador, lo que Von Eckardt (1993) incluye en los supuestos sustantivos que describen a la Ciencia Cognitiva en cuanto tal. El computador sería un dispositivo físico a través del cual es posible obtener datos o reportes controlados. El dispositivo sería físico porque los procesos computacionales se implementan en el cerebro, que es un medio físico particular. Asimismo, a través del diseño de programas y su implementación es posible su descripción. Por lo tanto, si la mente (el computador) es un dispositivo físico, es accesible a la tercera persona y, de esta manera, se obtiene información objetiva. El reto conductista, finalmente, es superado.

2.2 El paradigma computacional-representacional de la mente (PCR)

De acuerdo con la propuesta de Von Eckardt (1993), el marco de investigación de la Ciencia Cognitiva en su versión hegemónica, esto es, el paradigma computacional-representacional de la mente (en adelante PCR), se basa en los siguientes supuestos sustantivos: el computacional y el representacional. Tales supuestos constituyen las directrices que guían la investigación en Ciencia Cognitiva.

El *supuesto computacional* se refiere al hecho de que un determinado sistema real comparte ciertas características o rasgos en relación con un sistema idealizado de un cierto tipo. A su vez, se divide en dos: el *supuesto computacional vinculante* y el *supuesto computacional de sistema*. El primero corresponde a la metáfora del computador y postula que la mente puede entenderse como un mecanismo computacional, y las capacidades de la cognición humana, por tanto, constan de un sistema de capacidades computacionales (Von Eckardt, 1996: 98).

Von Eckardt (1996: 103-4), basándose en Boyd (1979), desarrolla las funciones que la metáfora del computador desempeñaría en Ciencia Cognitiva. Desde su punto de vista, la metáfora del computador cumple varios requisitos básicos para ser una *metáfora constitutiva de teoría*, es decir, una metáfora cuya función es introducir terminología teórica nueva en la investigación científica de

una disciplina, lo que se logra a través de la estimulación de la exploración científica de similitudes y diferencias previamente desconocidas entre el objeto real y la metáfora. En Ciencia Cognitiva, el objeto real corresponde a la mente y el computador es la metáfora. Concretamente, hay cinco maneras en que el *supuesto computacional vinculante* funciona como una *metáfora constitutiva de teoría*. En primer lugar, las metáforas científicas introducen terminología teórica y especializada donde no la ha habido. A este respecto, los científicos interesados en la cognición piensan que la metáfora computacional es el único método fructífero para describir los procesos mentales en el nivel funcional. Segundo, las metáforas científicas son bastante abiertas, y la metáfora del computador permitiría un cierto espacio para el descubrimiento científico, pues todavía no se sabe a cabalidad en qué medida la mente o el cerebro opera de forma computacional. En otras palabras, la vaguedad de la relación entre los dos elementos implicados en la metáfora deja espacio para la exploración de nuevas relaciones entre la mente y el mecanismo. En tercer lugar, las similitudes entre metáfora científica y el objeto real al que alude deben ser establecidas a través del método científico. Es decir, la ciencia no se queda en la simple y aparente relación entre los elementos implicados en la relación metafórica. En cuanto a la Ciencia Cognitiva, esto se cumple, pues concebir la mente como un computador es solo el paso inicial para la posterior determinación de las similitudes entre el sistema ideal y el real. El cuarto requisito parece no presentar problemas tampoco, a saber, la metáfora en ciencia es propiedad de una comunidad científica, no de un solo individuo, y la comunidad científica de la Ciencia Cognitiva parece no comportarse de forma diferente a otras comunidades científicas. Finalmente, como quinto rasgo básico, la fuerza de la metáfora no se pierde con el uso repetido que se hace de esta. En el caso específico de la Ciencia Cognitiva, difícilmente podría agotarse esa fuerza porque continuamente se están desarrollando nuevos mecanismos computacionales, apareciendo nuevas propiedades que alimentan la relación entre el componente que sirve de vehículo de la metáfora, el computador, y el elemento objeto de la metáfora, la mente. Estas cinco características básicas hacen de la metáfora del computador una *metáfora-constitutiva-de-teoría*.

Sin embargo, ¿la comunidad científica piensa que la mente solo comparte propiedades con el computador o considera que la mente es en sí misma un sistema computacional? Si esta última apreciación fuera correcta, entonces tanto la mente como las máquinas construidas por el hombre serían mecanismos computacionales de igual forma. En este punto, es inapropiado considerar el computador como una metáfora de la mente a menos que sea claro que la metáfora es falsa, es decir, a menos que se considere que la mente *no es literalmente* un computador. Sin embargo, al parecer la comunidad científica se inclina por pensar literalmente que la mente sí lo es (Von Eckardt, 1993: 116), ya que los resultados de las investigaciones están enmarcados en términos del procesamiento de la información y, además, los científicos cognitivos piensan que no hay ningún elemento conocido sobre la estructura y el funcionamiento del cerebro que haga implausible considerar el funcionamiento de la mente/cerebro equivalente al del computador. Por lo tanto, de acuerdo con lo último, el estatus metafórico del supuesto computacional es bastante cuestionable, pues habría una relación más literal que metafórica entre la mente/cerebro y el computador (Von Eckardt, 1993: 116).

Por otro lado, el *supuesto computacional de sistema* señala las propiedades que tiene un mecanismo computacional. Específicamente, este supuesto establece que un mecanismo computacional es capaz de recibir información, almacenarla, manipularla y liberarla en virtud de recibir, almacenar, manipular y liberar representaciones de esa información (Von Eckardt, 1993: 114). Un conjunto finito de reglas tendría la función de procesar la información. Estas reglas residirían en la máquina misma, tal como sucede en los computadores convencionales del tipo von Neumann y de alguna manera también en los conexionistas⁴. Al postular esta definición del supuesto computacional de sistema, Von Eckardt tiene en mente la concepción del computador en cuanto *mecanismo de procesamiento de la información o procesamiento de datos*, concordante con

⁴ En los computadores convencionales las reglas se almacenan en la memoria del programa o en el *hardware*, mientras que en las máquinas conexionistas no se puede almacenar literalmente la información *en un lugar* porque no hay una memoria entendida como un componente separado; pero sí se puede decir que la información se da de manera implícita en la relaciones de las unidades (Von Eckardt, 1993: 137). La información en un computador conexionista es codificada por los nodos y las conexiones de la red global.

los desarrollos de Boden (1981) y el posterior de Stillings (1995). Esta aproximación sería más amplia y adecuada que la otra aproximación presente en la tradición de la Ciencia Cognitiva, la cual caracteriza al computador en términos de conceptos extraídos de la teoría autómeta y de la teoría matemática de la computabilidad, seguida por Newell (1980, 1990), Haugeland (1981, 1985) y Johnson-Laird (1988), entre otros. Esta última sería menos apropiada por ser demasiado específica, ya que caracteriza al computador de una forma puramente conductual, o como un mecanismo autómeta a la manera de la Máquina de Turing. Con ello se dejarían de lado los aspectos estructurales y el computador seguiría reglas de una forma mecánica, con lo cual se evidencia su comportamiento desde una perspectiva estrictamente conductual.

Para Von Eckardt (1993), el concepto de representación también forma parte esencial, así como el supuesto computacional, de la caracterización de la Ciencia Cognitiva. Particularmente, desarrolla el concepto de representación dentro de lo que denomina *supuesto representacional*, enmarcado, a su vez, dentro de los *supuestos sustantivos*.

El supuesto representacional, al igual que el computacional, se divide en dos: el *supuesto representacional vinculante* y el *supuesto representacional de sistema*. El primero dice relación con que la mente o cerebro cognitivos son un mecanismo representacional y, por lo tanto, las capacidades cognitivas de un individuo humano constan de un sistema de capacidades representacionales (Von Eckardt, 1993: 143). Este supuesto es vinculante porque se refiere al hecho de que conecta el sistema del mundo real de las capacidades cognitivas humanas con la idea más abstracta y teórica de un mecanismo de representaciones. Tal supuesto, sin embargo, debe acompañarse de otro que defina qué es un mecanismo representacional. Para ello es necesario postular un supuesto representacional de sistema. No obstante, tal objetivo se presenta dificultoso por no haber consenso sobre qué se entiende por representación. Stillings (1995: 27), por ejemplo, tiene en mente más bien un tipo de representación proposicional para la Ciencia Cognitiva. Esto es, se define la proposición como la unidad completa más simple de pensamiento, y una de sus principales características es que puede

ser analizada bajo un criterio de verdad-falsedad. Por ejemplo, *Mary* no es una proposición completa, ni lo es *likes*, pero sí *Mary likes John* porque expresa un pensamiento completo que puede ser verdadero o falso. Así, los hechos que tienen lugar en la mente están representados a través de proposiciones, una forma interna e inconsciente de representación que luego se traduce en el lenguaje natural. Por tanto, el pensamiento es un estado previo al lenguaje, entendiendo por este último las lenguas naturales como el español, sueco, chino, portugués, etc. Por otro lado, la proposición capta relaciones entre argumentos. Por ejemplo, en el enunciado *Mary likes John* los argumentos son *Mary* and *John*, los cuales no son intercambiables, pues *Mary likes John* and *John likes Mary* son proposiciones distintas. Asimismo, al compartir argumentos dos o más proposiciones pueden presentarse en una sola oración: *Mary likes the teacher who gave her an apple*.

Sin embargo, ni la noción de representación proposicional ni ninguna otra han generado total consenso en la Ciencia Cognitiva (Von Eckardt, 1993: 144), y ello ocurriría porque la fuente de la que se extrae el concepto de representación es amorfa e imprecisa, abarcando desde nuestra experiencia cotidiana hasta las teorizaciones de otras disciplinas. Por esta razón, no existiría acuerdo para aceptar un cuerpo sistemático de conocimiento que se refiera al significado y características de las representaciones. No obstante, Von Eckardt intenta solucionar tal dificultad recurriendo a Pierce y su teoría de la representación para llegar a algún consenso. De esta manera, el *supuesto representacional de sistema* define un mecanismo representacional como aquel que tiene estados o contiene entidades que son representaciones. Una representación poseerá cuatro aspectos esenciales, a saber, será realizado por un portador de representaciones, representará uno o más objetos representacionales, sus relaciones de representación serán de alguna manera establecidas [*grounded*] y, finalmente, deberá ser interpretable por algún intérprete existente actualmente (Von Eckardt, 1993: 159).

Con respecto a la metodología usada en Ciencia Cognitiva, Von Eckardt (1993) intenta unificar los diferentes criterios de las distintas subdisciplinas que la

constituyen estableciendo once supuestos metodológicos, los cuales se referirán a la forma adecuada de estudiar el campo en cuestión. Serán de dos tipos. Por una parte, los que son compartidos por todos o la gran mayoría de la comunidad científica, constituyéndose en verdaderos compromisos. Este primer tipo tiene un sentido más macro, aplicándose a los fundamentos, a los valores esenciales del quehacer científico. Se deduce que se aplica a la ciencia en su conjunto. Por otra parte, están los supuestos metodológicos que se relacionan con el marco de investigación específico en cuestión, o sea, el de la Ciencia Cognitiva. Estos supuestos, por corresponder a una disciplina que engloba a otras que tienen sus metodologías particulares (Psicología Cognitiva, Lingüística Cognitiva, Neurociencia Cognitiva, etc.), tienen un carácter general.

Del total de diez supuestos metodológicos (en adelante para referir estos supuestos se utilizará M1, M2, M3, etc.), los tres primeros tratan del dominio de la Ciencia Cognitiva. El M1 y el M2 plantean que es aceptable estudiar la cognición humana aislada del contexto social y de otros aspectos de la mente. Específicamente, se sostiene que la Ciencia Cognitiva se enfoca en la cognición individual como opuesta a lo social. La cognición puede ser estudiada de forma exitosa centrándose exclusivamente en el individuo donde tienen lugar los procesos cognitivos. La influencia de la cultura o la sociedad se explica a través del hecho de que tal influencia es mediatizada por la percepción individual y la representación. Adicionalmente, se postula que las capacidades cognitivas humanas son autónomas y pueden estudiarse aisladamente de otros aspectos de la mente, como el afecto y la personalidad. Asimismo, el M3 valida la división de la cognición humana general en capacidades cognitivas individuales, de manera tal que cada una de estas capacidades particulares pueda ser estudiada también de forma aislada.

Los siguientes dos supuestos, M4 y M5, se relacionan con el ANTCOG (*adult normal typical cognition*), esto es, con la cognición de un adulto normal típico. Para Von Eckardt (1993: 6), este es el componente central del programa de investigación de la Ciencia Cognitiva. En primer lugar, esto es así porque es incontrovertible, ya que no está puesto en duda por nadie. En cambio, no ocurre lo

mismo con otros tipos de cognición, como la de los computadores, los animales u otros fenómenos mentales humanos como las emociones. En segundo lugar, el ANTCOG funciona como un punto de referencia de la investigación. Los investigadores en cognición intentan saber cómo la cognición funciona típicamente en adultos, cómo se realiza en el cerebro, cómo varía a través de los individuos o a través de diferentes poblaciones de individuos y culturas, y es en este punto que el ANTOG articula las distintas aproximaciones que apuntan hacia un estado normal. Por ejemplo, muchos déficits cognitivos son analizados como desviaciones de la norma así como el estudio de la realización neural se enmarca en una realización típica de los procesos y estructuras. Entendido de esta manera, el parámetro dado por el ANTCOG es útil para resaltar la variación y el cambio en la investigación científica.

Concretamente, el M4 y el M5 establecen que es significativo distinguir entre cognición normal y anormal, aunque haya una variación cognitiva considerable en cómo los individuos realizan los procesos inteligentes. Y a pesar de que existan diferencias sobre cómo llevan a cabo la cognición diferentes individuos, los adultos son lo suficientemente semejantes al realizar procesos inteligentes. Por tanto, es significativo hablar de adulto típico y es posible llegar a generalizaciones acerca de la cognición que se apliquen a todos los adultos normales.

El M6 defiende la estrategia explicativa de la Ciencia Cognitiva, que consiste en la reconsideración de las capacidades cognitivas humanas como amplias capacidades de procesamiento de la información. En otras palabras, la cognición humana es entendida en el marco de un sistema de capacidades computacionales y representacionales. Esto es la consecuencia de los supuestos sustantivos vinculantes tanto computacionales como representacionales ya explicados más arriba. En el M6 se implica que las respuestas básicas de la Ciencia Cognitiva pueden resolverse a través de dos convicciones: primero, se debe identificar las capacidades cognitivas ordinarias con las amplias capacidades de procesamiento de la información, y, segundo, la explicación de aquellas se hará fundamentalmente en el nivel del procesamiento de la información. La

primera de estas convicciones se basa en la creencia de que por cada propiedad general básica de la mayoría de las capacidades cognitivas ordinarias hay una propiedad de procesamiento de la información que las explica y las realiza en el nivel de procesamiento de la información.

Los siguientes supuestos, del 7 al 11, tratan del vínculo de la Ciencia Cognitiva con la empresa científica en general y la relación de las subdisciplinas que la componen. Para lograr una adecuada comprensión de las capacidades cognitivas humanas se necesitarán las herramientas conceptuales y metodológicas de diferentes subdisciplinas al interior de lo que se ha denominado Ciencia Cognitiva.

El M7 señala que para dirimir entre dos hipótesis se debe usar una metodología científica, es decir, es necesario recurrir a la justificación empírica. En tanto el M8 y el M9 apelan a la participación de cada una de las áreas de investigación que componen la Ciencia Cognitiva, y particularmente al papel jugado por la Neurociencia, respectivamente. Sin una contribución de cada una de estas, no se tendrá una teoría completa de la cognición. En cuanto al papel de la Neurociencia propiamente tal, esta debe restringir con sus hallazgos las hipótesis que apelan a describir la cognición humana en términos del procesamiento de la información. El M10 apunta en esta misma dirección. Si se quiere desarrollar una teoría cognitiva de la mente/cerebro, la estrategia óptima es adoptar una aproximación co-evolucionista, es decir, establecer hipótesis que apelen al procesamiento de la información sobre la base de descubrimientos empíricos tanto de las ciencias cognitivas no neurales, como de las neurociencias. Esta retroalimentación de las disciplinas conllevaría el traspaso de conceptualizaciones de disciplinas no-neurales a la neurociencia y viceversa, lo que es más fructífero desde un punto de vista teórico. Por su parte, el M11 establece ciertos límites a la neurociencia, señalando que esta no es suficiente para dar cuenta por sí sola de los niveles más altos de la cognición. Tales niveles entran en el ámbito de las teorías del procesamiento de la información, las que no son por tanto, en principio, eliminables.

2.3 El primer hito del paradigma computacional-representacional: la arquitectura clásica

El término *arquitectura* en Ciencia Cognitiva se refiere, en una primera instancia, al diseño de una mente, es decir, la arquitectura trata de describir componentes y funciones dentro de ésta. Pero no debe entenderse como una hipótesis que necesita ser comprobada. Más que eso, debe comprenderse como un conjunto de principios que modelan la cognición. Dentro del modelo hegemónico de investigación de la cognición, es decir, el PCR de la mente, hay dos arquitecturas distinguibles: la clásica y la conexionista.

Si bien existen diferencias importantes entre estas dos arquitecturas, básicamente ambas corresponden a una concepción tanto computacional como representacional de la mente (Thagard, 1996: 10). Es decir, ambas conciben la cognición en términos de estructuras representacionales y de procedimientos computacionales que operan sobre precisamente esas estructuras en la mente. Ya sea a través de la metáfora del computador serial clásico o del cerebro realizado en un sistema computacional conexionista, de una u otra manera la estrategia del estudio de la cognición implica una reducción a un sistema computacional. En este sentido, la Ciencia Cognitiva ha trabajado a través de su historia con la triple analogía mente-cerebro-computación (Thagard, 1996: 11).

El primer hito dentro del PCR es, como aparece enunciado en el título de esta sección, la arquitectura clásica. Esta forma de entender la cognición está motivada principalmente por el computador von Neumann, denominado así por las contribuciones del matemático John von Neumann para su desarrollo⁵. La idea central es la técnica de almacenamiento de programas (Sloman 1999). En este tipo de computador, tanto los programas como los datos se almacenan juntos. En cuanto a su composición, un computador von Neumann consta de una unidad de procesamiento central, una unidad de memoria, unidades de input y outputs. Esto es, la información de entrada es almacenada y transformada algorítmicamente para derivar en un output.

⁵ Sin embargo, aunque tradicionalmente se le atribuye a von Neumann los fundamentos de esta arquitectura, la idea original corresponde a Mauchly y Eckert (véase Copeland, 1993: 4-10).

Aplicando estos conceptos específicamente a la cognición humana, Stillings (1995) señala que en el procesamiento de la información se implican *sistemas sensoriales*, que son subsistemas separados encargados de recibir los inputs, esto es, la información recibida a través de los sentidos. Como ejemplo de estos subsistemas están la visión, la audición, el gusto, el olfato y el tacto. La información de los diferentes subsistemas sensoriales llega al *procesador central*, donde también reside el pensamiento, la memoria, el aprendizaje y la atención. Sin embargo, el *procesador central* no interviene en la computación de los outputs sensoriales, porque la información que es ingresada a través de los *sistemas sensoriales* es encapsulada, por tanto, la información obtenida no es esencialmente modificable. Por otro lado, el *procesador central* envía instrucciones al sistema motor, tercer componente del sistema de procesamiento de la información, para ejecutar los movimientos, que corresponderían a los outputs.

Todas las arquitecturas que se derivan del computador von Neumann tienen en común la necesidad de una memoria de trabajo o, en algunos casos, de la memoria a largo plazo. Otro elemento compartido es la forma proposicional en el procesamiento de la información, aunque ésta se presentará de distinta manera en los distintos tipos de arquitectura.

Dentro del computador digital del tipo von Neumann hay esencialmente tres subtipos (Sloman 1999):

- *Arquitectura de sistemas de producción*: esta postula que la mente tiene una memoria de trabajo, un conjunto de reglas de producción de carácter condicional y otro conjunto de reglas para determinar el orden de descarga de esas reglas de producción. Las producciones tienen la forma proposicional condicional “si c, entonces a”. En “c”, reside la condición que debe satisfacerse para producir una determinada acción.
- *Arquitectura del procesamiento de la información*: es una modalidad más detallada y más apegada al diseño de un computador von Neumann. Propone una secuencia de estados de procesamiento (y no un lenguaje de

transformación de símbolos como la *arquitectura de sistemas de producción*), que va desde un input a un output, pasando por la decodificación, el almacenamiento de la memoria y la recuperación. Aquí el papel del *procesador central* es relevante porque regula y controla el flujo de la información. Requiere, asimismo, de varias memorias, como la de trabajo, la semántica y, eventualmente, la episódica.

- *Arquitecturas que enfatizan la representación verídica del conocimiento humano*: este tercer tipo de arquitectura obedece al intento de relacionar la Ciencia Cognitiva con las tareas cotidianas. Se distinguen, por una parte, los datos, y por otro, el programa. De este modo el computador tiene la facultad de colocar la mayoría de las estructuras que serán representadas en el programa computacional o en los datos que este programa opera, como en los modelos representacionales. Estos últimos usan estructuras de datos bastante sofisticadas para modelar el conocimiento organizado. La representación sigue siendo proposicional, pero se vuelve más compleja para la realización de acciones inteligentes. Además, se necesitan dos tipos de memoria: una de trabajo y otra para los datos estructurados. Según Stillings (1995), un ejemplo de estas formas de representar la información son los *frames*, estructuras para representar un concepto o una situación estereotipada como “salir a cenar”, agrupando toda la información relevante para estas situaciones, por ejemplo, las expectativas que se tienen sobre ellas y la información sobre qué hacer si esas expectativas no son cumplidas. De esta manera, los *frames* tienen como función capturar la esencia de los conceptos (para una descripción sintética de *frame*, véase Nebel, 1999 y Brewer, 1999). Los elementos dentro de estas estructuras estarían representados como proposiciones y conceptos separados.

Para entender los fundamentos de la arquitectura clásica, una noción central es la hipótesis del *sistema de símbolos físicos* (Newell 1980). Según esta teoría, la mente humana opera como si fuese un sistema computacional,

concibiendo la cognición como un conjunto de procesos de manipulación simbólica formal (Stillings 1995: 20). Este sistema simbólico es físico porque los procesos formales se implementan en el cerebro, un medio físico particular (Stillings, 1995: 21). Es necesaria la participación de una memoria en la cual los símbolos puedan ser almacenados y recuperados. Además, los símbolos son procesados sintácticamente. En este sentido, la arquitectura cognitiva humana es capaz de acumular estructuras complejas (además de los elementos simples), identificar las partes de los elementos complejos, así como su disposición, y llevar a cabo procesos dependientes de la estructura simbólica (Stillings, 1995: 24). Un buen ejemplo de este procesamiento sintáctico lo constituye la gramática de las lenguas naturales. Si se considera el enunciado *el hombre viejo* [*the old man*], puede afirmarse que es en sí mismo un símbolo complejo, el cual, a su vez, puede ser considerado parte de una estructura mayor: *el hombre viejo pateó al burro porfiado* [*The old man kicked the uncooperative donkey*]. Al transformar esta estructura a *el burro porfiado pateó al hombre viejo* [*the uncooperative donkey kicked the old man*], se concluye que el orden en cada oración es distinto y que esto afecta su significado. El agente de la oración va a cambiar si se intercambia el lugar de los dos sustantivos, en este caso el *hombre* y el *burro*. En la primera oración *hombre* es el que golpea y en la segunda lo es el *burro*. Esta capacidad de llevar a cabo computaciones a partir de la disposición de los símbolos se denomina *computación de estructura sensible* (Fodor y Pylyshyn, 1988, referido por Stillings, 1995: 24).

Adicionalmente, las computaciones dentro del modelo clásico necesitan de *sistematicidad y productividad*. Los símbolos deben procesarse de acuerdo con un conjunto de reglas restrictivas que constituyen las pautas de funcionamiento de la arquitectura, y tales reglas pueden aplicarse repetidamente para producir más estructuras simbólicas, lo que las convierte en reglas de carácter recursivo. Por otro lado, los procesos operan sobre representaciones, de tal manera de asegurar que las estructuras con similares características sean procesadas idénticamente (Stillings, 1995: 25).

2.4 Conexionismo: el turno de la metáfora del cerebro

El segundo hito en el estudio de la cognición se produce por el surgimiento de una nueva arquitectura. Si con la clásica prevaleció el enfoque computacional, con el conexionismo habrá un intento de hacer más real, desde un punto de vista psicológico, la concepción de una mente. Esto se debe a que la arquitectura conexionista pretende formalizar el tipo de procesamiento que tiene lugar en el sistema nervioso central (Smolensky, 1989: 237), por tanto, estará basado en la estructura neuronal del cerebro.

Uno de los principios fundamentales de la teoría conexionista es que el conocimiento residiría en las conexiones de las unidades, es decir, que el conocimiento estaría implícito en la estructura del mecanismo que realiza la tarea (Rumelhart, 1989: 208), más que estar explícitamente contenido en la forma proposicional como en la arquitectura clásica (Stillings, 1995). Además, el conexionismo, tal como lo presenta Rumelhart, no tiene como modelo la metáfora del computador serial. Más bien es un modelo neuralmente inspirado (Rumelhart, 1989: 206), lo que significa que es un sistema computacional basado en cómo funciona el cerebro humano. En otras palabras, se plantea una arquitectura para simular a través de algoritmos los procesamientos del cerebro en un computador. En este sentido, el modelo conexionista considerará como su principal unidad de procesamiento una especie de abstracción de una neurona –un nodo, se deduce (Rumelhart, 1989: 207). La metáfora del computador, punto de referencia de la arquitectura anterior, es complementada por la metáfora del cerebro.

Un aspecto crucial que confronta a la arquitectura conexionista con la clásica es el papel que representa el tiempo. En el computador serial clásico, por ejemplo, los componentes operan a una velocidad de procesamiento medida en *nanosegundos*, o sea, a una velocidad mucho mayor que la del cerebro humano, que opera en una escala de *milisegundos*. Esto implica que la velocidad de los componentes de cada arquitectura restrinja fuertemente su modo de procesar la información. El computador de la arquitectura clásica es más rápido, pero el procesamiento es serial, esto es, se ejecutan los pasos secuencialmente. En

cambio, el cerebro humano, más lento, tiene a la vez más componentes que un computador serial, y debe desplegarlos de modo que funcionen cooperativamente y en paralelo con el objetivo de ejecutar las tareas. Para el cerebro humano, el proceso de la percepción, recuperación de la memoria o la comprensión de oraciones, implica alrededor de un segundo de ejecución, requiriendo cien pasos programáticos dentro de ese segundo, es decir, en ese período de tiempo el cerebro humano es capaz de incluir alrededor de cien acciones secuenciales elementales (Feldman, 1985, citado en Rumelhart, 1989: 207).

Un tipo de restricción del modelo conexionista para la cognición que merece atención es el hecho de que el conocimiento se dé en las conexiones. En el paradigma anterior esta información estaba en la forma proposicional de las representaciones. Aunque puede existir una clase de almacenamiento a corto plazo que pueda darse en los estados de las unidades, el almacenamiento a largo plazo se da entre las conexiones de las unidades. Una importante consecuencia de esto es que el conocimiento está implícito en la estructura del mecanismo conexionista que realiza la tarea.

Existe otro tipo de limitaciones que se dan en un nivel de análisis más abstracto y computacional. Al considerar la conducta humana como producto de un procedimiento en el cual tiene lugar una gran cantidad de restricciones que se desempeñan simultáneamente, los algoritmos ejecutados por un computador de la arquitectura clásica no podrían desempeñar esta tarea fácilmente, pues este tipo de computador no tiene contemplado un funcionamiento no serial. Sin embargo, para los algoritmos de tipo conexionista dar cuenta de dicho funcionamiento sería mucho más simple, porque tienen la facultad de actuar paralelamente. Por tanto, el uso de sistemas computacionales neurales ofrece, por una parte, una caracterización de cómo nuestro cerebro efectúa ciertas tareas de procesamiento de la información y, al mismo tiempo, da soluciones a problemas computacionales que parecen difíciles de resolver desde un marco de investigación computacional más clásico (Rumelhart, 1989: 209).

Más arriba se dijo que el conexionismo era un intento de formalizar el procesamiento que tiene lugar en el sistema nervioso central. La razón estriba en

que se pretende hacer más psicológicamente real la concepción de la mente. Para tal propósito, un sistema conexionista debería tener siete componentes (Rumelhart, 1989: 210-216), a saber:

- *Un conjunto de unidades de procesamiento:* éstos pueden ser objetos conceptuales particulares tales como características, letras, palabras o conceptos, es decir, elementos abstractos sobre los cuales el patrón de significados puede ser definido. Estas unidades tienen la importante función de realizar todo el procesamiento del sistema conexionista. No existen unidades supervisoras, a la manera de un procesador central de la arquitectura clásica, hay solamente unidades simples, las cuales tienen por funciones también actividades simples, como por ejemplo recibir inputs de sus unidades vecinas para luego computar un valor de output que es enviado, a su vez, igualmente a las unidades vecinas. Rumelhart clasifica las unidades en: a) inputs, que reciben información desde fuentes externas como los sistemas sensoriales; b) outputs, los cuales envían información al exterior, generalmente afectando al sistema motor; y c) unidades escondidas, inputs y outputs que establecen conexiones en el interior del sistema solamente, siendo invisibles para los sistemas externos.
- *Estado de activación:* además de las unidades arriba descritas, se necesita una representación del estado del sistema. Esto se especifica por un vector que representa el patrón de activación sobre el conjunto de unidades de procesamiento. El patrón de activación capturará lo que el sistema está representando en un determinado momento.
- *La función de output:* las unidades interactúan a través de la transmisión de señales a sus vecinas. La fuerza de la señal estaría determinada por los niveles de activación. Una función output se asocia con cada unidad del sistema, y de esta forma mapea el estado actual de activación a una señal output.

- *El patrón de conectividad:* este componente determina la información del sistema así como la manera de responder a un input arbitrario. Al especificar el patrón de conectividad entre las unidades se especifica, al mismo tiempo, el sistema de procesamiento y el conocimiento cifrado en este. Por otro lado, y en términos generales, los inputs que una unidad recibe corresponden a la suma total de los inputs de las otras unidades con las que se establece una conexión. De este modo, el patrón puede ser representado si se considera el peso de cada una de las conexiones del sistema. Si el peso es positivo, entonces se trata de un input excitativo, y, por el contrario, si el peso es negativo, se está en presencia de un input inhibitorio. Asimismo, es importante responder a la interrogante de cuánta información es posible almacenar y, en segundo lugar, cuánto procesamiento serial es necesario. Rumelhart intenta responder a esta pregunta con lo que él denomina el *fan-in* y el *fan-out* de una unidad. El primer concepto está relacionado con el número de elementos que excita o inhibe una unidad dada. El *fan-out* corresponde al número de unidades afectadas directamente por una unidad.
- *La regla de activación:* es la regla a través de la cual los inputs de una determinada unidad se conectan con otra unidad y su estado actual para producir un nuevo estado de activación. Se necesita una función F_i , que se vincula con $a_i(t)$ y el input net, siendo éste $= \sum_j w_{ij} o_j(t)$, y de esta forma se produce un nuevo estado de activación. Con frecuencia el estado de activación depende tanto del input actual como del anterior. Por otro lado, puede ocurrir que las activaciones decaigan lentamente con el tiempo y, aunque no haya ningún input externo de activación, su valor no caerá directamente a cero.
- *Regla de aprendizaje:* se utilizan para modificar patrones de interconectividad, proceso que es necesario para producir un cambio en la

estructura de procesamiento o de conocimiento. Se pueden encontrar, en principio, tres clases de modificación: a) desarrollo de nuevas conexiones; b) pérdida de conexiones ya existentes; c) modificación de conexiones existentes. Sin embargo, las dos primeras pueden incluirse en la última: si la fuerza de conexión, desmarcándose del valor *zero*, adquiere valores positivos o negativos, se logrará el mismo efecto que al surgir una nueva conexión. Por el contrario, si la fuerza de conexión pierde sus valores positivos o negativos hasta llegar a *zero*, el resultado es el mismo que al perder una conexión.

- *El ambiente*: en el modelo conexionista el ambiente es representado como una función aleatoria con variación en el tiempo (*time-varying stochastic function*) que actúa sobre un espacio de posibles patrones de input, es decir, que por cada posible patrón de input existe alguna probabilidad de que, en algún momento dado, ese patrón esté afectando a las unidades inputs. Puede ser que la función de probabilidad dependa de la historia de los inputs hacia el sistema, así como también de los outputs del sistema mismo. Generalmente, en la mayoría de los modelos conexionistas el ambiente es caracterizado simplemente como una distribución de probabilidad estable que actúa sobre un conjunto de patrones de input, independientemente de los inputs o respuestas anteriores del sistema.

Si se toma como principio la realidad psicológica de las arquitecturas que pretenden modelar la cognición, el conexionismo, basándonos en Rumelhart, presentaría ventajas comparativas con su competidora, la arquitectura clásica. En primer lugar, habría que señalar que parece corresponder a un modo más fidedigno del procesamiento cognitivo humano al estar basado en la estructura neuronal del cerebro. Asimismo, con este modelo se daría cuenta del conocimiento que no se manifiesta de forma explícita, como aquel conocimiento procedural, que no es descrito por la arquitectura del computador de von Neumann, por lo menos en su versión más clásica, que necesita de una

explicitación del conocimiento a través de la forma proposicional. Sin embargo, aunque estas diferencias son importantes, tanto la arquitectura clásica como la conexionista corresponden al PCR, pues ambas intentan reducir la cognición humana a un sistema computacional. El conexionismo, en última instancia, lo que hace es simular a través de algoritmos los procesamientos del cerebro en un computador.

En relación con la conciencia, históricamente el PCR de la Ciencia Cognitiva ha explicado esta en términos funcionales (Chalmers, 1995: 620-622). Para elucidar una determinada función, se necesita identificar un mecanismo que la desempeñe. Los métodos de la Ciencia Cognitiva se complementan perfectamente con ese objetivo, que está en consonancia con el *easy problem*. Por ejemplo, si se quiere explicar la integración y control cognitivos, es necesario dar cuenta de cómo los procesos centrales de un sistema integran la información y su uso posterior en la facilitación de varias conductas. O si se pretende explicar el acceso interno, se hace necesario describir cómo el sistema central es afectado por sus estados internos y la manera en que se usan posteriormente en otros procesos. El desempeño de una función en el PCR se explica especificando un mecanismo que la ejecute. En este contexto, el modelamiento neurofisiológico y computacional son apropiados para tal tarea. Al situar el problema desde una explicación de nivel bajo, se especifica el mecanismo neural responsable para la función. En cambio, si la explicación se plantea en un nivel más abstracto, se especifica un mecanismo en términos computacionales.

En las ciencias de niveles más altos este tipo de explicación reductiva trabaja en la forma antes señalada. A modo de ejemplo, al dar cuenta del gen se necesita especificar el mecanismo, en este caso el ADN, que almacena y transmite la información hereditaria de generación en generación. Una vez explicada la forma en que la función es ejecutada, se explica el gen. Respecto de la Ciencia Cognitiva, cuando se explica el aprendizaje, se debe dar cuenta también de cómo las capacidades conductuales del sistema sufren una modificación al ser expuestas a la información del medioambiente, y en qué manera esa información puede ser recibida para apoyar la adaptación de las

acciones del sistema a su contexto. El aprendizaje se explica en la medida en que se identifica cómo el mecanismo computacional o neural realiza la tarea. La misma situación se aplica para otros procesos cognitivos, como la percepción, la memoria y el lenguaje, entre otros.

No obstante, cuando se trata de dilucidar el tema de la experiencia subjetiva de la conciencia, una explicación basada en las funciones fracasa. El *hard problem* persiste debido a que no es un problema acerca del desempeño de funciones. Va más allá, porque aun al conseguirse la descripción de la discriminación perceptual, la categorización, el acceso interno, el reporte verbal, etc., queda una pregunta sin responder, a saber, por qué estas funciones se acompañan de la experiencia subjetiva (Chalmers, 1995: 621). Empero, esto no significa que este tipo de experiencia carezca de algún rol en la cognición inteligente, sino que simplemente no se puede reducir a ello. No hay ninguna función cognitiva, según Chalmers, de la que pueda afirmarse automáticamente que su explicación vaya a entregar una respuesta sobre la subjetividad.

3 Subjetividad, experimentos mentales y Ciencia Cognitiva. Una aproximación kuhneana

3.1 Concepción kuhneana del desarrollo científico

No es poco común encontrar, ya sea en la vida cotidiana o en el ámbito científico mismo, la creencia de que el estado de la ciencia actual se ha alcanzado a través de la reunión de los datos, descubrimientos y teorías que se han producido a lo largo de todas las épocas. Esta manera de entender el desarrollo científico es criticada por Thomas Kuhn (1971) en su libro *La estructura de las revoluciones científicas*. Kuhn sostiene que la historia de la ciencia no responde a un desarrollo acumulativo y que la ciencia actual no es la suma de las tradiciones científicas anteriores. Por tanto, una teoría no es un incremento de lo que ya se conoce, sino que todo estado de una disciplina en un momento dado es el rechazo a un modo anterior de hacer y pensar la ciencia; es decir, el desarrollo científico se lleva a cabo a través de estados sucesivos de investigación que son, por naturaleza, incompatibles. Cada nuevo estado se presenta con la revolución científica, que es definida como aquellos episodios de desarrollo no acumulativo en que un paradigma imperante es reemplazado por otro nuevo e incompatible (Kuhn, 1971: 149).

En este contexto, *paradigma* tiene fundamentalmente dos acepciones. En primer lugar debe entenderse como la “constelación de creencias, valores, técnicas, etc., que comparten los miembros de una comunidad dada” (Kuhn, 1971: 269). En segundo lugar, puede considerarse el término paradigma como un elemento particular de esa constelación, a saber, los *ejemplares*, noción que será abordada más adelante. Respecto de la primera acepción, es necesario no confundir paradigma con teoría, pues esta tiene implicancias mucho más limitadas respecto de su aplicación. Para desarrollar este punto, Kuhn recurre al concepto de *matriz disciplinaria*. Tal *matriz* refiere a varios elementos de orígenes distintos

que actúan en conjunto y forman un todo. Por su parte, el término *disciplinaria* alude al hecho de que una comunidad científica de un área particular posee un grupo de rasgos caracterizadores.

Existen cuatro elementos principales que constituyen una matriz disciplinaria. En primer lugar, están las *generalizaciones simbólicas*. Algunas veces se les codifica con una forma, como por ejemplo $f = ma$; mientras otras veces se expresan a través de palabras: “Los elementos se combinan en proporción constante por el peso” o “acción igual reacción”. La aceptación de este tipo de afirmaciones posibilita acuerdos sobre los que se fundamentan las técnicas de la manipulación lógica y matemática a la hora de solucionar los problemas científicos. Sin embargo, no solo tiene esa función para una comunidad de investigadores. Si bien estas generalizaciones son en parte leyes de la naturaleza, funcionan asimismo como definiciones de algunos de los símbolos que muestran. Es decir, las generalizaciones simbólicas establecen qué significan los términos que la componen y cómo deben comprenderse. La aceptación de una ley científica nueva o de un paradigma revolucionario implica una redefinición de conceptos que antes fueron entendidos de otra manera.

Un segundo componente de esta matriz es el *metafísico*. Con ello, Kuhn hace referencia a aquellas creencias basadas en modelos particulares, tales como “el calor es la energía kinética [sic] de las partes constituyentes de los cuerpos” (Kuhn, 1971: 282) o “todos los fenómenos perceptibles se deben a la interacción de átomos cualitativamente neutrales en el vacío o bien, en cambio, a la materia y a la fuerza, o a los campos” (Kuhn, 1971: 282). Este tipo de creencias se constituyen en verdaderos compromisos teóricos dentro una comunidad científica. Los modelos entregan a los miembros de una comunidad analogías y metáforas. De esta manera, encauzan la investigación determinando cuáles son las posibles explicaciones y la forma en la que se solucionarán los problemas dentro de una disciplina, así como qué tipo de enigmas deben ser tomados en cuenta.

Un tercer componente corresponde a los *valores*, que son los criterios que una comunidad de científicos utiliza para conducirse en la empresa científica. Habitualmente se los comparte transversalmente en grupos científicos de

diferentes áreas, generando un sentido de comunidad que trasciende a una disciplina específica. Se hace notoria su presencia cuando es necesario identificar una crisis al interior de un paradigma, o al momento de escoger entre dos paradigmas incompatibles. Probablemente los valores más reconocibles se asocian a las predicciones. Estas deben poseer valores como la exactitud, la preferencia por el carácter cuantitativo por sobre el cualitativo, la promesa de solución de enigmas, la sencillez, la coherencia interna, la compatibilidad con teorías existentes, etc. En otras palabras, los valores se convierten en criterios por los cuales una teoría o predicción es evaluada.

El cuarto y último elemento de la matriz disciplinaria son los *ejemplares*, es decir, las soluciones concretas a problemas que los estudiantes de ciencia usan de guía para resolver problemas nuevos. Para Kuhn (1971: 269) este elemento de la matriz es la segunda acepción fundamental por la que puede entenderse el concepto de *paradigma*. Los ejemplares pueden aparecer en los libros de texto universitarios, en los exámenes o simplemente ser enseñados en el laboratorio. Tales elementos pueden llegar a reemplazar reglas explícitas en la metodología y dirección de una investigación. Así, el científico investiga inserto en un conjunto formado por creencias y técnicas que modelan su accionar, enfrentándose tanto a un marco de problemas como a sus posibles soluciones. Por otro lado, las diferencias de los conjuntos de ejemplares entregan a la comunidad una finísima estructura de la ciencia. A manera de ilustración, los estudiantes de física aprenden, en un comienzo, los mismos ejemplares, a saber, problemas como el plano inclinado, el péndulo cónico y las órbitas keplerianas. Sin embargo, en el posterior desarrollo de su actividad como científicos, las *generalizaciones simbólicas* en las que se basan se realizan cada vez más por diferentes ejemplares, lo que expande y agudiza su capacidad para resolver enigmas. Tener una concepción de la ciencia como esta significa trasladar el conocimiento científico desde las reglas y teorías a la práctica basada en un modelo anterior de resolución de problemas, es decir, como un problema que ya había sido identificado.

Al enfrentarse a distintos ejemplares, lo que aprende el estudiante es a percibir una similitud, un rasgo unificador entre una variedad de situaciones. Por ejemplo, dada una generalización simbólica del tipo $f = ma$, el estudiante debe identificar los conceptos de fuerza, masa y aceleración en una diversidad de situaciones físicas nunca antes encontradas, pero al mismo tiempo él diseña una versión adecuada de $f = ma$, de tal manera de relacionar todas las realizaciones de los diferentes contextos. Y lo logra precisamente porque es capaz de captar la analogía entre dos o más problemas. Por tanto, este proceso conduce a un conocimiento tácito que se adquiere practicando la ciencia y no aprendiendo leyes para practicarla.

3.2 Definición y características de la ciencia normal

El triunfo de un paradigma conduce a lo que Kuhn denomina *ciencia normal*, definida como la “investigación basada firmemente en una o más realizaciones científicas pasadas, realizaciones que alguna comunidad científica particular reconoce, durante cierto tiempo, como fundamento para su práctica posterior” (Kuhn, 1971: 33). Si se vuelve al significado de paradigma, especialmente a su segunda acepción, es decir, considerado como ejemplo o realización útil para la solución de problemas, es posible darse cuenta que son dos conceptos estrechamente relacionados. La ciencia normal corresponde a una investigación constituida a partir de un paradigma.

Sin embargo, estos no son los únicos elementos implicados cuando se trata de entender el proceso científico. Antes de que una disciplina comience a desarrollarse a través de paradigmas, ciencia normal y revoluciones, tiene lugar un periodo de ambigüedad e indefinición. En un determinado momento, diferentes escuelas llevan a cabo una competencia entre sí que da lugar a un enfoque heterogéneo. Si no hay un paradigma que guíe, todos los hechos que potencialmente se incluyen en un campo parecerán importantes, lo que de alguna manera derivará en la pérdida de tiempo y en la casi inexistencia del progreso científico. Pero este periodo llega a su fin cuando una de las escuelas logra

imponerse al resto, constituyéndose en paradigma. Este es un instrumento que permite una nueva y más rígida definición del campo. Cuando un científico desarrolla su actividad dentro de este, no necesita exponer en sus trabajos toda la gama de conceptos y principios que lo sustentan. Por otro lado, sus investigaciones no se conocerán a través de libros dirigidos a todo tipo de lectores, como en su momento le pasó a *El origen de las especies* de Darwin. En lugar de eso, lo que se escriba estará contenido en artículos de corta extensión incluidos en revistas especializadas, e irá directamente dirigido a profesionales de la disciplina pertinente. Estos son las únicas personas que serán capaces de leer lo que el texto dice, puesto que manejan los códigos del paradigma. En los campos en que el libro todavía es un medio de difusión de investigaciones, es posible que personas no especializadas puedan leer lo que contienen, siendo un signo inequívoco de que no hay una profesionalización muy desarrollada. Por consiguiente, el paradigma parece ser el elemento que constituye a un campo como ciencia.

Por otro lado, un paradigma parece ser una verdadera promesa de éxito para la solución de problemas. La ciencia normal se presenta como el cumplimiento de esa promesa. El proceso señalado es lo que ocupa a la gran mayoría de los científicos durante toda su vida profesional. No se intenta producir nuevas teorías ni descubrir fenómenos extraordinarios, y debido a que se impone un enfoque restringido, se logra investigar un segmento de la naturaleza en profundidad, inimaginable en otras condiciones.

En cuanto a las direcciones que adopta la investigación de este tipo de ciencia, se pueden establecer por lo menos tres si se considera el ámbito de la experimentación. Primero, se centra en hechos que el paradigma muestra como reveladores de la naturaleza de las cosas. Hay un intento de determinarlos con mayor precisión. En segundo lugar, se apunta a fenómenos que se pueden comparar directamente con predicciones de la teoría del paradigma. Tercero, se realizan experimentos para articular la teoría, resolviendo eventuales ambigüedades. El último de estos aspectos es el más importante de todos. Por otro lado, en el ámbito teórico se presentan a su vez tres tareas que coinciden con

las anteriores: a) uso de la teoría para predecir información fáctica de valor intrínseco; b) manipulaciones teóricas que muestran una nueva aplicación del paradigma o que precisan una anterior, lo que nace de la necesidad de confrontar naturaleza y teoría; c) problemas teóricos de articulación de paradigmas. Generalmente, se tiende a la aclaración de una teoría mediante la reformulación. En otras palabras, las tareas de la ciencia normal se reducen a tres: determinación del hecho significativo, acoplamiento del hecho con la teoría y la articulación de esta última.

Es pertinente remarcar que la existencia de un paradigma no implica necesariamente la de las reglas explícitas. Las investigaciones pueden dirigirse perfectamente sin éstas: “La existencia de un paradigma ni siquiera debe implicar la existencia de algún conjunto completo de reglas” (Kuhn, 1971: 82). En vez de aquello, los problemas y las técnicas de investigación pueden relacionarse con realizaciones ya establecidas de la comunidad. Con esto se puede explicar, por ejemplo, la dificultad de descubrir las reglas que han guiado a las tradiciones particulares y también el carácter preponderantemente práctico de la educación científica. Los estudiantes no aprenden las teorías en abstracto, sino a través de sus aplicaciones; esto es, el paradigma dirige la investigación por medio de modelos directos, las realizaciones ejemplares.

La resolución de enigmas es la principal función de lo se ha llamado *ciencia normal*. Con el término *enigma* se hace referencia a aquella categoría especial de problemas que sirven para poner a prueba el ingenio y la habilidad de un científico (Kuhn, 1971: 70). No se aspira a producir, por tanto, novedades importantes. Además, la solución de un enigma no tiene por qué ser relevante o interesante. Incluso, suele suceder lo contrario: los problemas que imperiosamente reclaman una solución por su importancia para la sociedad no son un enigma, porque eventualmente no tienen ninguna solución. Los problemas que un grupo de científicos trate de resolver deben tener asegurada una solución o, al menos, generar esa expectativa. Solo este tipo de problemas adquieren el rótulo de científicos. Aquellos que no encajan en esta descripción son calificados generalmente de “metafísicos” o “demasiado problemáticos”. Lo que conduce a un

científico a abordar los problemas de la ciencia normal es la convicción de que logrará resolverlos. Si no lo hace, se deberá únicamente a su falta de habilidad. Además, un enigma debe tener más de una solución asegurada, así como la existencia de reglas que limiten las soluciones y los pasos necesarios para llegar a estas.

3.3 La anomalía en la ciencia normal. El preámbulo del descubrimiento científico

A pesar de centrarse en los enigmas, la ciencia normal en algunos casos descubre fenómenos nuevos e inesperados. Este descubrimiento comienza con la percepción de una *anomalía*; es decir, un hecho que ha violado las expectativas de la ciencia normal y para el cual el investigador, por consiguiente, no estaba preparado (Kuhn, 1971: 100). En un determinado momento, el conjunto de herramientas disponibles en el paradigma no sirve para abordar ese problema. Más bien, su solución parece demandar un reordenamiento tanto metodológico como teórico, debido a que no entra en el marco de expectativas de la comunidad científica.

Posteriormente a la detección de la anomalía, se produce un ahondamiento en los problemas que esta plantea. Tal exploración va a concluir solo cuando la teoría del paradigma sea ajustada. Así, el hecho anómalo pasa a ser un hecho esperado. Sin embargo, si ello se logra, no es por un simple ajuste aditivo a la vieja teoría. Debe tener lugar una rearticulación. La ciencia debe aprender a ver la misma naturaleza de una manera diferente. Mientras tanto, el nuevo hecho no es completamente científico si no se produce la asimilación (Kuhn, 1971: 93).

Las consecuencias del párrafo anterior hacen evidentes las relaciones entre hecho y teoría. Necesariamente, el descubrimiento de un fenómeno es algo complejo, o más bien un proceso, “que involucra el reconocimiento, tanto de *que* algo existe como de *qué es*” (Kuhn, 1971: 97, cursiva en el original). Si no hay una teoría que defina y enmarque el nuevo hecho, este es precientífico. El descubrimiento no debe entenderse como un acto único y simple que tiene lugar

en un momento dado y que es realizado por una persona. Es un proceso con sus respectivas fases y requiere tiempo. En un primer momento, hay una percepción de que algo funciona mal con un determinado enigma, el cual no logra ser resuelto con las herramientas habituales del paradigma. Luego tiene lugar el reconocimiento conceptual y de observación, por parte de la comunidad científica, de esa anomalía para finalizar con el consiguiente cambio de las categorías y los procedimientos del paradigma que, por cierto, muchas veces producen resistencia al interior de la comunidad. Finalmente, cuando lo anómalo se convierte en lo previsto, el descubrimiento se completa (Kuhn, 1971: 110).

A modo de ejemplo de la relación entre hechos nuevos y teorías en el marco del descubrimiento, Kuhn (1971: 93) se refiere a lo sucedido con el oxígeno. Al menos son tres las personas que tuvieron la pretensión de atribuirse su descubrimiento: C. W. Scheele, Priestley y Lavoisier. Pero particularmente fue sobre estos dos últimos que recayó la polémica por la "autoría". Sin embargo, aunque no haya existido duda sobre quién descubrió el oxígeno, la pregunta sobre el momento de tal hallazgo puede ser inapropiada e inducir a un error en la percepción de la naturaleza del descubrimiento científico (Kuhn, 1971: 95).

En el caso de Priestley, la muestra de gas que logró obtener con su experimentación no era pura, y si obtener una muestra de oxígeno impuro es descubrirlo, todos quienes simplemente hubieran embotellado aire habrían también descubierto el oxígeno. Además, el mismo Priestley pensó, en un primer momento, que había obtenido muestras de óxido nitroso viendo después en el oxígeno solo aire deflogistizado, por tanto, no tenía conciencia de lo que realmente significaba su hallazgo.

En Lavoisier pueden observarse complicaciones similares. En una primera instancia, tampoco logró identificar el oxígeno en cuanto tal, y cuando finalmente pudo hacerlo, en los trabajos de 1776 y 1777, llegó a una concepción errónea. Pensaba que el oxígeno era un principio de acidez atómico, cuestión que fue eliminada de la química posteriormente. Esto muestra que la atribución a un momento dado a un descubrimiento no es del todo real si no se considera la asimilación del nuevo hecho en un marco que lo defina. Se necesitan nuevos

conceptos para analizar tales acontecimientos. La aseveración de que algo fue descubierto, aunque correcta, induce a un error, “debido a que sugiere que el descubrir algo es un acto único y simple, asimilable a nuestro concepto habitual de la visión (y tan discutible como él)” (Kuhn, 1971: 97).

Para desarrollar la idea de que el descubrimiento es un proceso, Kuhn (1971: 108-109) entrega un segundo ejemplo a través de un experimento psicológico. En una oportunidad, se le pidió a un grupo de sujetos que identificaran una serie de cartas de la baraja después de que estas habían sido expuestas brevemente ante sus ojos. La mayoría de las cartas eran normales, es decir, no tenían ningún cambio en su diseño que las diferenciara de las cartas de una baraja común. Había, sin embargo, una minoría de cartas que habían sido hechas intencionalmente anómalas. Después de cada exposición, se le preguntaba al sujeto qué había visto. Respecto de las cartas normales, su identificación era frecuentemente correcta. No obstante, la identificación de las cartas anómalas tuvo mayores dificultades, puesto que los sujetos solían identificarlas, casi siempre, como normales. El cuatro negro de corazones, por ejemplo, podía ser identificado como un cuatro de picas o de corazones. Los individuos, al actuar de esta manera, no hacían sino ajustar las cartas anómalas a las categorías establecidas por las experiencias previas. Por otro lado, cuando se aumentó el tiempo de exposición de las cartas anómalas, algunos sujetos comenzaron a darse cuenta de la diferencia. Al incrementarse aun más la exposición de estos elementos extraños, aumentaban las dudas respecto de lo que los sujetos creían percibir. Posteriormente, y de forma muy repentina, la mayoría de los sujetos comenzó a acertar en la identificación de las cartas anómalas, aunque hubo un grupo que nunca pudo darse cuenta de la anomalía, no siendo capaces de adaptar sus categorías a los nuevos hechos. Esto parece ser ilustrativo de cómo se comportan los miembros de una comunidad científica al enfrentarse con una anomalía y de las fases del descubrimiento científico: “En la ciencia, como en el experimento con las cartas de la baraja, la novedad surge sólo dificultosamente, manifestada por la resistencia, contra el fondo que proporciona lo esperado”

(Kuhn, 1971: 109. Para una profundización respecto del problema de la percepción y las teorías, véase González, 1999).

La crisis de un paradigma se inicia cuando la anomalía persiste y se agudiza. Se le presta, por parte de los científicos, más atención a un problema que presenta dificultades para su aclaración. Con el tiempo, se intentan soluciones divergentes entre sí que traen como corolario que las reglas que rigen la ciencia normal se hagan cada vez más confusas, y poco a poco los científicos comienzan a manifestarse en desacuerdo con su paradigma.

3.4 Características de las revoluciones científicas

Anteriormente, se definió revolución científica como aquel episodio en el cual ocurre un cambio de paradigma que es incompatible con el antecesor. Por lo tanto, los conocimientos y todo el conjunto de elementos que generalmente contiene un paradigma (teoría, métodos, reglas, ejemplos, etc.) *no constituyen una suma* de la tradición científica inmediatamente anterior. Más bien se trata de una reconstrucción del campo. Aunque es cierto que se puede llegar a manejar el “mismo” conjunto de datos, estos en el nuevo contexto cambian en sus características, porque se ubican en otro sistema de relaciones. Por otra parte, la defensa de cada paradigma, en el marco de una discusión científica, plantea un problema importante. Cuando se tienen dos candidatos para regir la práctica de una disciplina, el paradigma vigente y su posible sucesor, los partidarios de uno y otro emplean argumentos que responden a su propio esquema teórico. De ahí que no exista un punto de contacto en el debate, puesto que para cada grupo su paradigma se presentará como correcto. En este sentido, la aceptación de la comunidad científica es la norma más eficaz para la elección de un paradigma.

Eso no implica, por cierto, que una teoría no pueda persuadir a los miembros de la teoría rival. Si se intenta dar cuenta de los distintos tipos de argumentos para persuadir a los científicos del cambio, se tiene, en primer lugar, la solución que genera un paradigma para problemas que no pudieron resolverse con el anterior. Adicionalmente, existen argumentos de carácter más implícito

como las consideraciones que Kuhn (1971: 241) denomina “estéticas”. Con ello hace alusión a que, por ejemplo, una teoría será elegida en desmedro de otra por aspectos como su “simpleza” en la formulación o porque es más “apropiada”. En tercer lugar, está la fe, es decir, la elección puede basarse en las promesas de resolución de problemas futuros. El paradigma anterior, si bien aparece como más confiable, al enfrentarse a la anomalía o a un problema que presenta dificultades extraordinarias ve coartado su margen de aplicación. Y esto es precisamente lo que el paradigma reemplazante posee; es decir, a pesar de no estar consolidado todavía como una práctica efectiva de resolución de problemas, promete la solución de los venideros. Por lo tanto, el nuevo paradigma crea expectativas donde el paradigma anterior sólo ve límites.

En otro ámbito, el cambio de paradigma hace que la comunidad perciba el mundo de manera diferente: “Lo que ve un hombre depende tanto de lo que mira como de lo que su experiencia visual y conceptual lo ha preparado a ver” (Kuhn, 1971: 179). A modo de ejemplo, se puede presentar un caso ocurrido en el campo de la astronomía. Entre 1690 y 1781, lo que actualmente se conoce como el planeta Urano fue caracterizado como una estrella, ignorándose el movimiento que hubiese sugerido otra interpretación. Herschel observó que el tamaño del disco era inusual para que ese objeto fuese una estrella, postulando que esa forma correspondía más bien a un cometa. Posteriormente, sería un científico llamado Lexell quien describió como planetaria la órbita de ese objeto. El cambio en la visión de las cosas y el mundo no responde, como se pudiera pensar, a la interpretación diferente que se hace de los fenómenos en un momento dado. Los científicos de diferentes paradigmas reúnen datos distintos, debido a que los paradigmas condicionan desde un comienzo su visión. Este punto es desarrollado por Hanson (1977), para quien la visión es una experiencia que no puede ser reducida a un estado físico caracterizado como un estímulo que pasa a través del sistema visual, es decir, como un mero proceso óptico-químico. No existe una visión neutra, aséptica por parte de un científico o de cualquier individuo en general. Las teorías y las interpretaciones tienen lugar en la visión misma, apareciendo en el acto y no como un proceso posterior en el cual se interpretan

los datos. Cuando se observa un fenómeno, a menudo existe un contexto inherente al pensar sobre lo que se observa: “Estamos preparados para apreciar de ciertas maneras los aspectos visuales de las cosas. Los elementos de nuestra experiencia no se agrupan al azar” (Hanson, 1977: 93-94). Esto explica que dos científicos, como Herschel y Lexell, vieran, según Hanson, algo distinto cuando ambos están fijando su atención en la misma cosa, pues su contexto, su conocimiento, sus teorías, o su paradigma en términos kuhnianos, los han preparado a ver una determinada cosa de una determinada manera. Dicho con otras palabras, *ver* no consiste en tener una experiencia visual solamente, sino también en el modo en que esta es experimentada (Hanson, 1977: 94). Esta posición, sostenida también por Kuhn, constituye una diferencia a lo planteado por algunos filósofos tradicionales, que pueden ser adscritos al positivismo lógico o al empirismo en general (véase González, 1999: 96-104), según los cuales las teorías se aplican a la visión objetiva de los fenómenos, interpretándolos. De esta manera de concebir la actividad científica, se desprende erróneamente que hay dos momentos, a saber, primero la observación pura de los datos y, segundo, una interpretación de estos sobre la base de teorías o conocimiento previo. Pero, como se señaló más arriba, el hecho de ver algo implica que hay conocimiento en esa visión, por tanto, el conocimiento radica en la visión y no es un proceso posterior. Al observar un objeto, se podría afirmar que la interpretación es la visión (Hanson, 1977: 104), de manera que no hay dos operaciones cuando se observa un determinado fenómeno.

En cuanto a la visión acumulativa de los descubrimientos científicos, errónea a los ojos de Kuhn, es apoyada por los libros de texto de las diferentes disciplinas, ya que “disimulan sistemáticamente (...) la existencia y la significación de las revoluciones científicas” (Kuhn, 1971: 212). Los libros de texto son el medio por el cual un estudiante adquiere los conocimientos y ejemplos necesarios para resolver problemas que plantea la ciencia normal. Al parecer, este efecto de invisibilidad de las revoluciones obedece al intento de reafirmar la ciencia contemporánea. Si se recurre a la historia del trabajo científico, es sólo para rescatar aquellos aspectos que contribuyan al paradigma actual. Así, un momento

determinado de una disciplina aparece como consecuencia inmediata de una tradición.

3.5 La función de los experimentos mentales. Una visión de conjunto

Los experimentos mentales (en adelante EMs) han suscitado un largo debate en torno a su aplicabilidad y confiabilidad tanto en ciencia como en Filosofía. En esta última área, y con el desarrollo de la Filosofía de la Mente, ha habido también importantes contribuciones en cuanto a la creación de nuevos EMs y de su discusión. En este sentido, es fundamental mencionar a John Searle (1980) y la Pieza China, EM que alcanzó gran notoriedad y en el cual se discute la posibilidad de que un programa comprenda historias exactamente de la misma manera que los humanos lo hacen. Otros EMs en Filosofía de la Mente se han centrado en el aspecto fenoménico de la conciencia y la crítica al Fisicalismo, como los de Nagel (1974), Jackson (1986) y Chalmers (1996), entre muchos otros.

Dada la abundante bibliografía sobre los EMs, es pertinente describir sus características relevantes como método de investigación y debate tanto en ciencia como en Filosofía. Preguntas sobre su estructura, función y clasificación necesitan ser abordadas como un primer paso para una posterior discusión de la concepción kuhneana del desarrollo científico y el rol de los EMs en Ciencia Cognitiva.

Según Brown (2008), en el último tiempo ha habido principalmente dos posiciones divergentes sobre EMs en Filosofía de la Ciencia que se circunscriben a dos tradiciones filosóficas, como el racionalismo platónico y el empirismo clásico. Brown (1991) y Norton (2004) son, respectivamente, sus representantes más importantes.

Para Norton, los EMs son meros argumentos que comienzan con premisas basadas en el ámbito empírico y que siguen reglas de inferencia deductivas o inductivas para, posteriormente, establecer una conclusión. Aquellos aspectos “pintorescos” propios de la situación imaginada que le entregan a los EMs un carácter experimental serían redundantes, aunque podrían ser psicológicamente útiles. Con “pintoresco” Norton alude al elemento narrativo asociado a un EM, lo

que sólo sería un aspecto anecdótico que apoya una determinada tesis. Desde este punto de vista, un EM está imposibilitado de ir más allá de las premisas empíricas, de manera que no entregaría ningún descubrimiento ni podría hacer más que la argumentación ordinaria.

Una preocupación fundamental para Norton (2004) sobre los EMs es el problema epistemológico. En el caso de que los EMs mentales sean capaces de entregar conocimiento del mundo natural, ¿cuál es la procedencia de este conocimiento? La respuesta a esta pregunta estriba en las premisas tácitas o explícitas que sustentan un EM. Tal conocimiento es transformado, generalmente de forma tácita, por medio de argumentación deductiva o inductiva para entregar el resultado final.

La tesis de que los EMs son sólo argumentos pintorescos se fundamenta en dos planteamientos (Norton 2004). En primer lugar, los EMs siempre pueden reconstruirse de forma argumentativa, basados tanto en suposiciones explícitas como implícitas, que entregarían el mismo resultado de un EM. Para ello es necesario que un EM tenga una base empírica, es decir, que esté enraizado en la experiencia de mundo. Solo este criterio empirista da seguridad sobre el conocimiento que un EM pretende entregar. La argumentación es, entonces, la reformulación posterior de esa experiencia. Por lo demás, según Norton, no existe ningún EM que no pueda reconstruirse de la manera antes señalada. En segundo lugar, crear un EM consiste en la puesta en marcha de un argumento, esto es, el alcance epistémico de estos dos recursos es exactamente el mismo. Un EM no tiene ningún poder especial, sino que inferiría predicciones tomando como base la experiencia ordinaria.

Como se planteó más arriba, en una posición opuesta está Brown (1991 y 2008), quien define los EMs como mecanismos de la imaginación que investigan sobre la naturaleza de las cosas y que aportan nuevo conocimiento de estas, pese a su carácter *a priori*. Estos se clasifican, a grandes rasgos, en constructivos y destructivos considerando su relación con una determinada hipótesis. Los primeros son usados positivamente para establecer una hipótesis o explicar una teoría, mientras los segundos son empleados para socavar una teoría científica a

través de la puesta en evidencia de una fisura, incoherencia o contradicción. Adicionalmente, hay un tercer tipo: aquellos que destruyen una teoría pero al mismo tiempo fundan otra. Tales EMs son denominados platónicos. Uno de los famosos ejemplos que Brown entrega para ilustrar esta clase de experimentos es el de Galileo, que trata de la velocidad con que caen dos cuerpos de distinto peso. Concretamente, lo que hace el EM de Galileo es destruir la concepción aristotélica de que los cuerpos pesados caen más rápido que los livianos y, simultáneamente, establecer que todos los cuerpos, no importando su peso, caen a la misma velocidad. Para tal propósito, Galileo plantea una situación hipotética en que una bala de cañon (P) y una de mosquete (L), menos pesada que la anterior, son atadas para formar un objeto compuesto (P+L). Esta situación conduce a una contradicción, puesto que P+L debería caer más rápido que el objeto P solo, dado que es más pesado. Pero, curiosamente, P+L debería caer más lento también, ya que la parte más liviana L actuaría como freno respecto de P en la composición P+L. En consecuencia, la teoría aristotélica es falsa y, adicionalmente, se extrae una conclusión evidente: los objetos L, P y L+P deben caer a la misma velocidad.

Sin embargo, Brown no se queda en la definición de este último tipo de EM entendiéndolo sólo como la suma de las características destructivas y constructivas. Ya el hecho de que sea rotulado como “platónico” anuncia el fondo al que Brown quiere llegar. Un EM platónico, además de poseer las características anteriores, es *a priori*, en el sentido de que “no está basado en nueva evidencia empírica ni se deriva lógicamente de datos viejos; y significa un progreso en el sentido de que la teoría resultante es mejor que la teoría anterior” (Brown, 1991: 76, traducción mía). La cuestión de que un EM no entregue datos empíricos nuevos es algo bastante asimilado por la comunidad filosófica en general y no debiera ser un asunto controversial aquí tampoco; de hecho, el mismo Kuhn (1982) en su fundacional artículo sobre EMs lo reconoce de esta manera. En cuanto a que el conocimiento extraído de un EM no se deriva meramente de la deducción lógica de los datos antiguos, Brown sostiene que el EM de Galileo no constituye una verdad lógica porque no podría servir para refutar una situación en la que los objetos cayeran a una velocidad diferente según su distinto color, por

ejemplo (véase Brown 2004: 30-31). No es posible afirmar que dos objetos que tienen una misma propiedad de las características del color suman esa propiedad de la misma manera que se puede hacer con el peso. En otras palabras, dos objetos que pesan 20 kilos cada uno considerados como una unidad sumarían 40 kilos, en cambio, la suma de dos objetos azules no constituyen una unidad más azul. En este sentido, la contradicción dada entre dos objetos de distinto peso en el marco de la teoría aristotélica de la velocidad no se daría entre dos objetos con distinto color que cayeran a velocidades distintas.

Hay algo aún más importante que está en la concepción de Brown (1991): los EMs platónicos conducen a leyes de la naturaleza, verdaderas relaciones entre universales, que son, por su parte, entidades abstractas que existen independientemente de los objetos físicos, también de los individuos mismos e, incluso, desterradas del espacio y del tiempo. EMs como el de Galileo permitirían el acceso a este escurridizo mundo platónico y, concretamente, a la ley de la naturaleza *todos los cuerpos caen a la misma velocidad*.

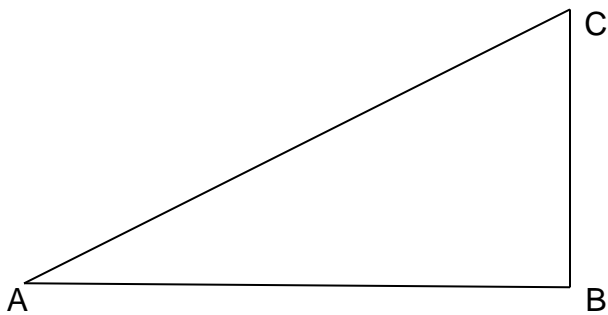
Anteriormente, en el marco de la teoría kuhneana del desarrollo científico, también ha habido un lugar para la discusión de los EMs. Primero, es necesario aclarar que cuando Kuhn (1982) expone sus ideas al respecto, tiene en mente, al igual que Brown y Norton, fundamentalmente disciplinas como la Física, la Química y la Astronomía. No obstante, sus ideas pueden ser aplicadas en el ámbito de la Ciencia Cognitiva para ayudar a la comprensión del desarrollo de esta disciplina y su relación con las otras ciencias.

Kuhn intenta dilucidar cuál es la función de los EMs. En primer lugar, piensa que los EMs no solo sirven para enfrentar al científico con una contradicción implícita en su manera de pensar. Si los EMs solo hacen eso, entonces la nueva manera de entender un fenómeno no correspondería a una comprensión de la naturaleza, sino de lo que es el aparato conceptual del científico (Kuhn, 1982: 264). Tal reconocimiento de la contradicción ayudaría a eliminarla y, al mismo tiempo, no requeriría de datos empíricos nuevos. Sin embargo, afirmar esto no es del todo exacto, según Kuhn. Más bien, los EMs mentales posibilitan al científico llegar a leyes y teorías diferentes respecto del paradigma anterior y, aunque no

entregan datos empíricos nuevos, se sitúan mucho más cerca de la experimentación real (Kuhn, 1982: 265).

Para desarrollar el punto anterior, Kuhn (1982) expone dos ejemplos, uno correspondiente a un experimento de laboratorio ejecutado por Piaget, y el otro al caso histórico del cambio conceptual de la velocidad en la tradición científica, desarrollado por Galileo a través de un EM. Este último será analizado a continuación.

Galileo logró hacer evidente la contradicción en la teoría aristotélica respecto del concepto de velocidad. Galileo, que es representado por *Salviati* en el EM, les pide a dos interlocutores que imaginen dos planos (CB vertical y CA inclinado) levantados sobre un plano horizontal (AB). Gráficamente se puede representar como sigue:



En el experimento dos cuerpos resbalan por los planos CB y CA sin fricción desde el punto C. Además, hay una concesión importante: cuando los dos cuerpos lleguen a su destino final, esto es, los puntos A y B, van a adquirir la misma velocidad. Posteriormente, *Salviati* les pregunta a sus interlocutores cuál de los dos cuerpos se mueve más rápido. Ambos señalan que el movimiento perpendicular CB es el más rápido.

Para llegar a esa conclusión, los interlocutores se valen de dos criterios. El primero de ellos es el de “llegar a la meta”. Al ser la distancia CB más corta que CA, es natural pensar para esos sujetos que esa era la respuesta porque el cuerpo que va desde C hasta B recorre una menor distancia y llega, por tanto,

“antes” a la meta, aunque el “más lento” (CA) haya recorrido más distancia. Por otra parte, el segundo criterio implicado es el de la “borrosidad perceptual”, que corresponde al juicio originado de la percepción inmediata de la velocidad del movimiento. Si un cuerpo aparece visualmente moviéndose “más rápido” que otro, será percibido como más rápido. En el contexto del EM de Galileo, el plano CB es percibido por los interlocutores como más rápido.

Sin embargo, pronto surgen los primeros problemas frente a la respuesta dada. Uno de los interlocutores señala que esta es incompatible con la concesión inicial, a saber, que al llegar a sus respectivos puntos de destino, tanto A como B, ambos movimientos adquirirían la misma velocidad. Si los dos cuerpos parten del reposo y alcanzan la misma velocidad final, deberían, por tanto, tener la misma velocidad media. Entonces, ¿cómo explicar que un elemento es más rápido que el otro?

El experimento es complejo porque se comparan dos cuerpos que recorren dos distancias distintas y, por otro lado, se define el movimiento más rápido como aquel que abarca la misma distancia en menor tiempo.

Luego, se les señala a los interlocutores que comparen los tiempos que los dos elementos necesitan para recorrer una distancia patrón. El patrón escogido corresponde al plano vertical CB. El problema se hace aún más complicado, puesto que el cuerpo que se mueve más rápido dependerá del lugar del plano CA considerado para medir la distancia patrón CB. Al hacer la medición desde la parte más alta del plano inclinado CA, el movimiento que se sitúa en el plano perpendicular CB ocupará menos tiempo en realizar el movimiento, y será considerado el más rápido. Si, por otra parte, la distancia patrón se mide considerando la parte más baja de CA, el cuerpo sobre el plano CB tomará más tiempo para abarcar la distancia a recorrer, lo que lo haría a este último más lento. Finalmente, si la distancia patrón CB no se mide considerando ni las partes alta ni baja, y se utilizan, por el contrario, puntos interiores, entonces los tiempos de CB y CA serán iguales, por lo que tanto la velocidad del plano inclinado como perpendicular no diferirá. Tal exposición de este EM pone en evidencia una contradicción, es decir, surgen tres respuestas posibles a la misma pregunta, y

cada respuesta es, a su vez, incompatible con las otras dos. En este sentido, el experimento mental de Galileo sirve para modificar los aparatos conceptuales de quienes lo enfrentan, dando cuenta de una paradoja implícita en la manera de pensar de un científico (Kuhn, 1982: 274).

El EM muestra que “más rápido” y “velocidad” no deben ser usados de la manera tradicional. Según Aristóteles, una cosa más rápida que otra recorre una magnitud mayor en un tiempo igual, una magnitud igual en menos tiempo así como una magnitud mayor en menos tiempo. En su concepción, la noción implícita de velocidad corresponde a lo que actualmente se considera como “velocidad promedio”, lo que corresponde al resultado de la distancia completa entre el total del tiempo transcurrido. Esto está enmarcado dentro de la concepción aristotélica del movimiento. Todo movimiento no celeste es visto como un acto completo y finito que se capta en conjunto. La medición de la cantidad y la velocidad de un movimiento se relaciona con los parámetros medievales que describen sus puntos terminales, los *termini a quo* y *ad quem*. Esta manera de considerar la velocidad es muy diferente de la concepción moderna de la que Galileo es un representante. Aristóteles no hace distinciones básicas que sí se hacen desde un enfoque moderno. No distingue velocidad instantánea, y los aspectos más relevantes de su física se condicionan por esta falta de distinción. La noción de “más rápido” da cuenta de algo diferente al ser aplicada en distintos contextos. Si se usa para referirse a la comparación instantánea del movimiento en instantes determinados, es muy distinto de aplicar ese mismo concepto de “más rápido” para comparar los tiempos cuando se completa el total de dos movimientos especificados. Por ello, un cuerpo puede ser más rápido en el primer sentido aquí expuesto, pero no en el segundo.

La reciente exposición del experimento de Galileo convierte a los EMs en herramientas que corrigen los errores conceptuales de un científico, pero, así considerado, es imposible que un EM enseñe algo sobre el mundo, o lo que es lo mismo, sobre nada que no sea ya conocido (Kuhn, 1982: 276). Desde una concepción tradicional, un EM es eficaz si presenta una situación normal en la que el científico se siente bien equipado para manejar. Nada de los elementos

imaginados en un determinado EM puede ser desconocido; por el contrario, debe haber familiaridad con la naturaleza. Por no contener información nueva, entonces el EM no enseña nada nuevo. No obstante, según Kuhn, y como fue adelantado al principio de esta sección, los EMs no se remiten a esta mera función correctora. Al ser expuestos frente a experimentos de esta índole, naturalmente la gente aprende de sus conceptos, pero también aprende algo acerca del mundo (Kuhn, 1982: 276). Cuando los interlocutores de *Salviati* comprenden de mejor manera el concepto de velocidad, aprenden adicionalmente la manera en la que se mueven los cuerpos en la naturaleza.

Además, el término “contradicción” o “confusión” en la que se encuentra un científico antes de ser expuesto a un EM merece al menos una clarificación. Es incorrecto decir que una situación que antecede a un EM es “contradictoria en sí”, puesto que para que sea considerado de esa manera debe aplicarse en todo mundo posible, como el ejemplo del círculo cuadrado, que es lógicamente imposible. El concepto aristotélico de velocidad no corresponde a un fenómeno de estas características. Es más, puede usarse para describir la mayoría de los movimientos que tienen lugar en la realidad, actuando como una verdadera ley de la naturaleza y formando parte de toda la teoría aristotélica del movimiento. Por el contrario, constituye un concepto contradictorio solo cuando se aplica a situaciones en que el criterio de velocidad instantánea y el criterio de velocidad promedio conducen a respuestas opuestas, como en el EM de Galileo, que expone respuestas incompatibles a la misma pregunta. No hay defectos intrínsecos en el concepto aristotélico mismo, pues no adolece de defectos lógicos. Simplemente no encaja en la estructura real del mundo a la que se pretendía aplicar. Aristóteles razonó como si en su mundo la velocidad hubiera tenido un carácter uniforme o cuasiuniforme, y hasta que no fueron puestas en evidencia las dificultades de aplicar semejante concepto de velocidad, no se podía calificar de “confuso” o “contradictorio”. De hecho, su aplicación había sido eficaz en la práctica investigativa precedente, pasando inadvertidas las posibles complicaciones. De esta manera, es posible concluir que el desajuste entre conceptualización y naturaleza fue la responsable de la confusión. Darse cuenta

de los errores que tal concepción de la velocidad conllevaba ayudó a aprender algo acerca del concepto mismo, pero también algo acerca de la realidad.

Según Kuhn (1982: 284), los EMs se asemejan a los experimentos de laboratorio en cuanto a, primero, la posibilidad de revelar que la naturaleza no se ajusta a un determinado conjunto de expectativas y, en segundo lugar, a constituirse potencialmente en métodos de revisión tanto de estas expectativas como de la teoría. Dicho de otro modo, ambos tipos de experimentos conducen a un conocimiento nuevo o a una nueva comprensión de la naturaleza. Si esto es posible, la pregunta es cómo se produce. Los experimentos de laboratorio sirven a ese propósito porque entregan información nueva. Mientras los EMs se basan por completo en información existente. ¿Cómo, entonces, tienen la misma función si la naturaleza de los datos que se manejan en uno y otro es completamente distinta? La respuesta a esta paradoja kuhneana estriba en que a veces los EMs “le dan al científico acceso a una información que, a la vez, tiene a mano y a pesar de eso le resulta de alguna manera inaccesible” (Kuhn, 1982: 284). Es inaccesible al científico porque su paradigma lo ha restringido a enfocarse en los enigmas, es decir, en aquellos problemas enmarcados en la ciencia normal, la cual no tiende hacia novedades fácticas o teóricas. La ciencia normal selecciona los problemas a abordar, los enigmas. Pero más de alguna vez un problema se presenta más complicado de lo común, y más que un simple enigma se puede transformar inesperadamente en una verdadera anomalía, esto es, en un hecho que no logra ser resuelto bajo la aplicación de las herramientas y teorías tradicionales. Tal anomalía al chocar persistentemente contra los procedimientos de un paradigma es el síntoma de la crisis y el preludio de la revolución científica. Lo que se conocía vagamente o estaba recluido en la periferia, logra, gracias a los EMs, ser enfocado y evidenciado con mayor precisión. A este respecto, Kuhn (1982: 286-287) señala que “los datos indispensables para que ocurra la revolución han estado existiendo en el borde de la conciencia científica”. El EM, en este sentido, es una herramienta analítica esencial que se emplea en un periodo de crisis y que contribuye a una reforma conceptual dentro de una disciplina, haciendo parte del ámbito de investigación un aspecto anteriormente ignorado o recluido en los

márgenes de las preocupaciones científicas. Ayuda al científico a considerar la importancia de un hecho inaccesible, por lo cual cambia el conocimiento que un científico tiene del mundo. Ese hecho periférico que no encajaba en el paradigma anterior es visto de una manera diferente. En conclusión, los EMs evidencian una anomalía en el marco de un paradigma científico.

Gendler (1998) desarrolla esta concepción kuhneana de los EMs. Al igual que Brown, fundamenta su teoría en el experimento de Galileo sobre dos cuerpos de distinto peso que caen libremente. Para Gendler, un EM de esas características le permite al científico darle sentido a un concepto previamente incomprendido, de manera que la creencia resultante es nueva. Por tanto, un EM mental puede conducir al científico a reconocer la inadecuación de su marco conceptual al presentarle fenómenos que él reconoce como parte de su mundo, ahora sistematizados por el EM, y que antes de la experiencia imaginaria tenían un carácter tácito e inarticulado, además de no estar bajo la organización de ningún paradigma. En otras palabras, un EM “puede conducirnos a rechazar compromisos teóricos débiles (y a final de cuentas falsos) a la luz de un *conocimiento* nuevamente sistematizado, pero anteriormente inarticulado acerca de cómo es el mundo” (Gendler, 1998: 415, cursiva en el original, traducción mía). Además, los EMs obligan a una participación constructiva del lector, lo que le da una fuerza justificativa adicional que un simple argumento no posee, constituyendo de esta manera una interesante crítica frente a lo que postula Norton (2004). El lector se ve impulsado a realizar, más que un EM, un *experimento-en-pensamiento* [*experiment-in-thought*], por medio del cual se pregunta a sí mismo qué esperaría o juzgaría si existen tales y cuales circunstancias. Dada esa situación inicial, el lector-científico intenta encontrar la respuesta, y a través del desarrollo de la situación imaginaria, llega a una creencia que él estima verdadera porque la experimentó personalmente. O dicho de forma más simple aún, la gente cree lo que hace. La fuerza justificativa se explica, entonces, porque las personas tienden a creer en cosas y situaciones que experimentan directamente, por ello un conjunto de premisas de un argumento no tiene este poder al no implicar una

participación constructiva del lector como sí ocurre en un EM. De esta manera, y en oposición a Norton, un EM no puede ser reducido a un argumento.

Esta concepción kuhneana-gendlereana de los EMs puede ser fructífera al ser aplicada en el marco del estudio de la conciencia fenoménica en la Ciencia Cognitiva y, particularmente, en los EMs en Filosofía de la Mente que abordan su problematización (Nagel, 1974; Jackson, 1986; Chalmers, 1996, entre otros). La descripción y explicación de la conciencia fenoménica ha presentado complejidades dentro del paradigma reduccionista dada su ontología de primera persona. Desde un criterio objetivista, la conciencia fenoménica se presenta inaprensible puesto que su punto de vista de aproximación es el de la tercera persona, y, entendida como elemento constituyente del mundo físico, no se conforma a un determinado conjunto de expectativas científicas. Un EM como el de Nagel (1974), que se expondrá en la siguiente sección, puede demostrarlo con contundente claridad, así como el EM de Galileo demostró que la concepción aristotélica de la velocidad no se ajusta al mundo real.

3.6 El rol de la experimentación mental en la Filosofía de la Mente: el caso del murciélago de Nagel

La experimentación mental en Filosofía de la Mente puede presentarse como una herramienta útil en el estudio de la conciencia para dar cuenta de la ineficacia de las aproximaciones reduccionistas como la del paradigma computacional-representacional de la mente (PCR), cuyo punto de vista de tercera persona tiende hacia la objetividad, ignorando el aspecto fenoménico. Particularmente ilustrador es el EM de Nagel (1974) *What is it like to be a bat?*

Se sabe que el murciélago es un tipo de organismo que percibe el mundo exterior a través de la *ecolocalización*, es decir, por el envío de ondas sonoras que rebotan en los objetos del mundo natural y que finalmente vuelven al murciélago para ser procesados. Su cerebro está diseñado para que la información recibida por este medio le permita hacer discriminaciones precisas de distancia, tamaño, forma, movimiento y textura similares a las que hacen las personas con su sistema

visual. Sin embargo, aunque se puede concluir esto a través de la descripción objetiva de este organismo, las personas no poseen, en su sistema perceptual, una estructura similar al sonar del murciélago. Por ello esto presenta una barrera insalvable para saber realmente qué es ser como un murciélago. La experiencia humana proporciona el material básico para la imaginación, y si no se tiene la experiencia de la ecolocalización, es imposible que la imaginación humana pueda concebir tal característica. De esta manera, el conocimiento que se obtiene desde una descripción objetiva siempre será incompleto, puesto que falta dar cuenta del aspecto *what it is like*. Por tanto, no se debe negar su realidad, aunque el punto de vista objetivo de la tercera persona es incapaz de abordarlo. El aspecto fenoménico de la conciencia, ignorado por enfoques reduccionistas, debe ser incluido en el mundo físico si se pretende estudiar la conciencia en toda su complejidad (Searle, 2004: 80-81).

Este EM tiene una enorme implicancia para el problema mente/cerebro en los estudios cognitivos. El aspecto fenoménico de la conciencia puede entenderse como una *anomalía* kuhneana, es decir, como un hecho inesperado para la comunidad científica, para el cual las herramientas y teorías disponibles dentro de un paradigma científico no dan ninguna solución ni expectativa de llegar a lograrla. El carácter subjetivo de la experiencia no es un fenómeno para el cual el paradigma reduccionista esté preparado. Un análisis así solo se plantea desde la objetividad de la tercera persona. El EM del murciélago ilustra la imposibilidad de dar cuenta de las características fenoménicas desde una perspectiva como esta. La razón estriba en que la subjetividad está esencialmente conectada con un solo punto de vista, el de la primera persona, y de esta manera “parece inevitable que una teoría física objetiva renunciará a semejante punto de vista” (Nagel, 1974: 220, traducción mía). Si el carácter fenoménico de la conciencia solo es comprensible desde el punto de vista de la primera persona, entonces un enfoque que tienda hacia la objetividad no se acercará al fenómeno que pretende estudiar, sino que, por el contrario, se alejará de él (Nagel, 1974: 223). El PCR de la mente, enmarcado en la Ciencia Cognitiva, pertenece a esta aproximación reduccionista de la realidad. Concretamente, lo que hace este paradigma es entender la

conciencia en términos de sus habilidades y funciones (Chalmers, 1995: 620), por tanto, solo abarca el *easy problem*. Pero la conciencia subjetiva excede estos planteamientos y no puede explicarse sobre la base de la realización de determinadas consideraciones causales. El problema de la conciencia subjetiva, el *hard problem*, “persiste incluso cuando la realización de todas las funciones relevantes es explicada” (Chalmers, *ibídem*, traducción mía). O como lo explica Güzeldere (1997: 28), una concepción computacionalista intenta explicar la conciencia desde una perspectiva de tercera persona, esto es, en términos de *lo que hace*, pero no necesariamente en cuanto a *lo que se siente*.

Por las razones esgrimidas en los párrafos anteriores, los EMs constituyen una herramienta analítica válida en el estudio de la conciencia, pues hacen ver como ineficaz la estrategia reduccionista, que se plantea desde la objetividad, para el estudio de la conciencia fenoménica. Kuhn (1982: 287) señala que los EMs se emplean durante la crisis de un paradigma y promueven la reestructuración conceptual básica al interior de una ciencia. Y esa reestructuración, se deduce, parte de la agudización de los problemas ocasionados por el hecho inesperado. La función de este EM no es más que evidenciar la *anomalía*. En esta misma dirección, es posible afirmar que los EMs en ocasiones “le dan al científico acceso a una información que, a la vez, tiene a mano y a pesar de eso le resulta de alguna manera inaccesible” (Kuhn, 1982: 285). O en palabras de Gendler (1998: 415), un EM puede sistematizar datos, pertenecientes al mundo real, que previamente existían de forma inarticulada y desorganizada. En el caso de la conciencia, la información disponible e inarticulada corresponde al aspecto subjetivo, el cual es inaccesible por la inadecuación de las herramientas y teorías para estudiarlo.

El carácter fenoménico es un dato “a mano” y, por tanto, “viejo”, pero al mismo tiempo es un conocimiento que no ha sido integrado por ninguna teoría científica como un objeto de investigación. Es decir, es una información familiar porque es parte constitutiva de la experiencia humana, pero que está en un estado precientífico o preteórico si se piensa que, debido al modo de aproximación reduccionista objetivo, no forma parte del ámbito en el cual la ciencia suele

aplicarse. El aspecto fenoménico de la conciencia, al constituirse en anomalía, lo hace en la medida en que el EM de Nagel lo presenta como un aspecto de la realidad que no debe ser ignorado, porque si se actúa de esta manera, lo que resulta es la eliminación de una característica esencial de la conciencia (Nagel, 1974 y 1986; Chalmers, 1996: xii-xiii; Searle, 2004: 84). Este EM presenta un reto enunciado en la pregunta *what it is like to be a bat?* La imaginación humana es limitada y no puede llegar a saber qué es como ser un murciélago, pero al menos posibilita el hecho de darse cuenta de que tal característica, aunque fenoménicamente inalcanzable desde la imaginación de una persona con una estructura biológica muy diferente a la de un murciélago, forma parte de la realidad. De esta manera, el ámbito de lo real debe también comprender lo subjetivo, o en palabras de Nagel (1986: 27): “La realidad no es solo realidad objetiva”, lo que convierte a la experiencia subjetiva en un hecho inesperado, una *anomalía*, para un paradigma reduccionista que estudia la conciencia y que se fundamenta en una concepción de objetividad física. En este sentido, si el EM de Nagel no entrega datos empíricos nuevos, como ha señalado la mayoría de la literatura sobre los EMs como género (Irvine, 1991), al menos sí permite el acceso a un fenómeno viejo que ha sido ignorado e inarticulado, el cual a través de la conceptualización y problematización se presenta como nuevo para la ciencia.

Al tener en cuenta la clasificación de Brown (1991), el EM de Nagel parece tener, al menos, una importante característica de un EM platónico: el poder socavar y al mismo tiempo fundar una teoría científica. Sin embargo, no posee todos los elementos de un EM platónico, como por ejemplo, el hecho de ser a priori o falible, debido a que está basado en la experiencia, en este caso subjetiva, y, por otro lado, posee un carácter infalible, dada la incapacidad humana de imaginar la experiencia fenoménica de ser otro organismo tan diferente como el murciélago.

La razón de que socave y al mismo tiempo funde una teoría científica estriba en que pone en jaque una concepción reduccionista de la conciencia, mientras que, por otra parte, establece como ley de la naturaleza que el aspecto fenoménico es esencial para su descripción. Si esta ley da cuenta de ese

escurridizo mundo platónico con el que Brown sueña, es algo que es difícil de aceptar en una primera instancia, y excedería con largueza los alcances del presente trabajo. Pero sí, al menos, se puede afirmar con toda propiedad que el EM de Nagel tiene una doble función, la cual consiste en destruir y crear una nueva teoría, como los EMs platónicos, en este aspecto al menos. Su rol es, entonces, bastante similar a la del EM de Galileo de los dos cuerpos en caída libre. La estrategia de Nagel consiste en evidenciar que las aproximaciones reduccionistas dejan escapar un elemento esencial de la conciencia, esto es, su dimensión subjetiva, sin la cual simplemente no hay conciencia. Al plantear la pregunta *what it is like* se intenta que el lector-experimentador participe en este EM y piense en la posibilidad de disociar la experiencia fenoménica de la conciencia y, por otro lado, compatibilizar la descripción objetiva de la misma conciencia con su aspecto fenoménico. Lo que se consigue es que el lector perciba una contradicción, al igual que en el EM de Galileo, aunque de diferentes características. Específicamente, la contradicción está en el intento de hacer una descripción objetiva (desde un punto de vista de tercera persona) de un fenómeno que tiene una ontología de primera persona. Es decir, la conciencia no puede entenderse a cabalidad sin incluir el punto de vista de la primera persona, y si se quiere describirla objetivamente, la real naturaleza del fenómeno deja de percibirse.

Otro punto interesante es que este EM posee características que lo distinguen de otros. Irvine (1991: 153) sostiene que una de las diferencias entre los EMs y los físicos es que estos últimos dependen de una intervención real en la naturaleza, mientras que los primeros adquieren la forma de un argumento basado en premisas hipotéticas. Pero, ¿en el EM de Nagel se cumple esta característica? Si se comprende que la conciencia y su dimensión fenoménica forman parte de la naturaleza, la diferencia parece no ser tan clara. Cuando se apela al lector para hacer el esfuerzo imaginativo de sentir como un murciélago, de alguna manera se experimenta de forma real con el aspecto subjetivo, por tanto, podría afirmarse que existe intervención directa en la naturaleza. A diferencia del EM de Galileo, en donde el razonamiento argumentativo de carácter contrafáctico ocupa un lugar

fundamental, con el EM del murciélago lo que se pretende es poner en el centro la experiencia subjetiva, la percepción misma del espectador, lector o experimentador, no como una situación hipotética, sino como un elemento constitutivo de la realidad de una persona. Pese a que adicionalmente se apele a una situación imaginaria, lo central es la imposibilidad de su percepción y no la situación contrafáctica en sí. Un experimento planteado de esta manera permite el acceso a la conciencia en un contexto de experimentación y no se queda solo en un razonamiento hipotético. Tal acceso privilegiado está dado por su modo de aproximación, el punto de vista de la primera persona, esencial característica de la conciencia subjetiva. Por tanto, si bien el experimento de Nagel posee elementos contrafácticos, que son una característica típica del género de los EMs, hay una intervención directa en la naturaleza, una intervención entendida no como una modificación, sino en términos de una exploración en la subjetividad que, por cierto, es un hecho perteneciente al mundo físico. Esa intervención estriba en la exploración de los límites de la imaginación humana, que irremediamente fracasa en alcanzar la experiencia fenoménica de un murciélago, lo que sitúa a esta experiencia fenoménica como un elemento irreductible desde la tercera persona objetiva, e inalcanzable, al mismo tiempo, desde la comprensión de otra mente con una estructura fisiológica muy diferente a la del murciélago. Por tanto, estos elementos hacen del experimento de Nagel un EM híbrido, dada sus características contrafácticas y también experienciales.

Quizá los términos de *experimento-en-pensamiento* (*experiment-in-thought*), como plantea Gendler (1998: 414), o de *experimento introspectivo* sean más apropiados para dar cuenta de lo planteado por Nagel. Aunque Gendler nada dice sobre EMs de la conciencia ni sobre lo argumentado por Nagel, plantea que los EMs son experimentos reales. El lector-experimentador participa constructivamente en un EM realizando un *experimento-en-pensamiento*. Quien guíe el experimento comienza preguntándose a sí mismo qué juzgaría si encontrara determinadas circunstancias. Posteriormente el lector se lanza a la búsqueda de la respuesta. Esta sería la técnica básica de la participación constructiva del lector. El punto que quiere establecer Gendler parece aceptable

en cierta medida si se aplica también al experimento del murciélago. Sin embargo, Gendler tampoco menciona que un *experimento-en-pensamiento*, además de las características contrafácticas, tenga que intervenir en la naturaleza, pero su propuesta podría ajustarse para describir el EM de Nagel como un experimento híbrido en el cual se da tanto la argumentación contrafáctica como una intervención en el mundo físico.

3.6.1 ¿Es relevante la conciencia para el paradigma computacional-representacional de la mente?

Una pregunta que puede surgir es si la conciencia subjetiva es relevante al paradigma actual en Ciencia Cognitiva, como una forma de desacreditarla en su calidad de problema científico. En otras palabras, si el paradigma cognitivista no considera la conciencia subjetiva un problema científico porque no es relevante a la inteligencia, ¿en qué se fundamenta su estatus de anomalía? Parte de la respuesta a estas preguntas surge del análisis de los conceptos de *ciencia normal* y *paradigma*. Como se definió más arriba (véase 3.1 y 3.2), *ciencia normal* es una investigación basada en una o más realizaciones científicas anteriores, es decir, en uno o más ejemplos, paradigmas, de cómo llevar a cabo la investigación científica en un determinado campo. Este significado de paradigma equivalente al de ejemplo, o *ejemplar* en términos kuhnianos, es importante, pero no el único, y en lo sucesivo se aludirá a esta acepción simplemente como *ejemplares* para evitar confusiones con el otro significado de paradigma más importante aquí. La otra acepción de paradigma, y que será referida como tal en lo que sigue, se refiere a la constelación de creencias, valores, métodos, técnicas, etc., que comparten los miembros de una comunidad científica. Sin embargo, no debe confundirse este concepto con el de teoría, que es mucho más simple. Para evitar tal confusión, Kuhn describe esta acepción más amplia de paradigma como una *matriz disciplinaria*. *Matriz* fue usado con la intención de denotar varios elementos diferentes contenidos en uno mayor, por otro lado, *disciplinaria* se refiere a los rasgos en común que tienen quienes practican una disciplina particular.

Los componentes del paradigma o matriz disciplinaria, como fue expuesto detalladamente en 3.1, son cuatro, a saber, las generalizaciones simbólicas, el componente metafísico, los valores y, finalmente los ejemplares. En lo sucesivo, se abordará el componente *metafísico* por la importancia para el tratamiento de la experiencia subjetiva como anomalía. Kuhn (1971: 282) define este componente metafísico en términos de compromisos de un grupo de científicos que están basados en determinadas creencias. Tales creencias se aplican a aquellos *modelos*, entendiendo estos como algún tipo de concepción teórica, que surgen en la historia de la ciencia aplicándose a un fenómeno de la naturaleza, y que los científicos aceptan como verdad. Ejemplos de esta clase de creencias son: “El calor es la energía kinética [sic] de las partes constituyentes de los cuerpos”, o “las moléculas de un gas actúan como minúsculas bolas de billar, elásticas, como un movimiento producido al azar” (Kuhn, *ibídem*). En el caso de la Ciencia Cognitiva tradicional, un ejemplo de este tipo de creencias se puede plantear refiriendo al supuesto computacional establecido por Von Eckardt (1993), a saber: *la mente es un mecanismo computacional, y las capacidades de la cognición humana, por tanto, constan de un sistema de capacidades computacionales*. Según Kuhn (1971: 283), compromisos de este tipo “dan al grupo sus analogías y metáforas preferidas o permisibles”. Respecto de la Ciencia Cognitiva, es evidente que la metáfora del computador ha sentado las bases de su investigación a lo largo de décadas, como fue abordado en el capítulo 2. Lo relevante aquí es que este tipo de elementos ayudan a determinar las explicaciones y soluciones aceptables para los enigmas. Los enigmas son aquella categoría de problemas que constituyen el objeto de la ciencia normal, para los cuales existe una expectativa de solución, y que, por tanto, ponen a prueba la capacidad y el ingenio de un científico. Es decir, son los problemas comunes a los que se enfrenta la mayoría de los científicos durante toda su vida. Respecto de las creencias, estas ayudan a determinar cuáles son los tipos de problemas que deben ser abordados, esto es, influyen en la selección de los enigmas (Kuhn, *ibídem*). En otras palabras, el paradigma, a través de su componente metafísico, señala cómo deben resolverse los problemas, cuáles son los métodos estandarizados para hacerlo y, más importante

aún, cuáles son los problemas mismos de interés para una disciplina. En este punto, entonces, la experiencia subjetiva no forma parte de esos problemas, porque no encaja en la concepción computacional de la mente, o en otras palabras, no ha formado parte de los enigmas del PCR de la Ciencia Cognitiva.

La ausencia de la experiencia subjetiva en el paradigma cognitivista tradicional se explica porque este posee una perspectiva de tercera persona, mientras que el *hard problem* de la conciencia tiene una ontología de primera persona. Esto hace que el enfoque cognitivista planteado en estos términos identifique a la conciencia respecto de *lo que hace* y no en cuanto a *lo que se siente* (Güzeldere, 1997: 28). En este contexto, la conciencia es individualizada, en cuanto al rol que desempeña, como un componente integral de una red mucho más amplia de estados mentales y procesos. Concretamente, es caracterizada como un proceso que cumple una tarea específica, un verdadero módulo con una función causal determinada en el sistema cognitivo general, con el aporte de las ideas computacionalistas y modelos de procesamiento de la información desarrollados en la Psicología Cognitiva (Güzeldere, *ibídem*). Por supuesto, el aspecto de la conciencia que se tiene en mente para desempeñar este rol causal es lo abarcado por el *easy problem*, esto es, la conciencia entendida como un mecanismo de procesamiento e integración sensorial, puesto que desde la perspectiva cognitivista tradicional el aspecto fenoménico de la conciencia, el *hard problem*, no tendría poderes causales. Visto de esta manera, no hay lugar para la conciencia fenoménica en el PCR, es irrelevante, pero sí hay lugar para la conciencia como un elemento integrador de los distintos sistemas sensoriales, es decir, para el *easy problem*.

Responder a la crítica de la irrelevancia de la conciencia subjetiva en la cognición inteligente no es fácil si se consideran, precisamente, los criterios del paradigma en el que no encaja como problema. Tales criterios, fundados en el componente metafísico de un determinado paradigma, convierten el debate entre paradigmas en una argumentación circular, donde cada bando argumenta considerando sus creencias en ciertos modelos sobre la naturaleza de las cosas. En este sentido, si una determinada comunidad científica adopta o rechaza un

paradigma, no se debe necesariamente a los argumentos expuestos por sus defensores o detractores (Kuhn, 1971: 151-152). Respecto de este punto, dos paradigmas en competencia son dos modos incompatibles de práctica científica en una comunidad. Por esta razón, la elección de uno u otro no puede estar determinada solo por la evaluación que se hace desde el paradigma predominante, pues estos criterios evaluativos dependen en parte de un paradigma que está siendo, precisamente, discutido. En este sentido, cuando dos paradigmas discuten cuál es el mejor “su función es necesariamente circular. Para argüir en la defensa de ese paradigma cada grupo utiliza su propio paradigma” (Kuhn, 1971: 152). La argumentación circular efectuada desde el paradigma discutido solo puede tener un carácter persuasivo, y “no puede hacerse apremiante, lógica ni probablemente, para quienes rehusan entrar en el círculo” (Kuhn, 1971: 152). En cuanto al triunfo mismo de un paradigma no hay más norma que la aceptación de la comunidad científica pertinente, a la manera de lo que ocurre en las revoluciones políticas (Kuhn, *ibídem*). Por tanto, la importancia de los argumentos que sostengan los miembros de un determinado paradigma debe relativizarse para explicar la adopción de un paradigma. Por supuesto, aunque pueda desarrollarse una argumentación circular en la defensa de un paradigma, no significa necesariamente que esos argumentos sean erróneos o inefectivos. Pero el estatus de un argumento de estas características es solo el de la persuasión (Kuhn, *ibídem*), y no obliga, desde un punto de vista lógico, a la elección de un determinado paradigma. Esto ocurre porque las premisas y valores compartidos por los partidarios de distintos paradigmas no son lo bastante amplios como para realizar una elección bajo parámetros lógicos, satisfaciendo, en menor o mayor medida, los criterios que dictan para sí mismos. En otras palabras, “no hay un algoritmo neutral para la elección de teorías, no hay ningún procedimiento sistemático de decisión que, aplicado adecuadamente, deba conducir a cada individuo del grupo a la misma decisión” (Kuhn, 1971: 305). Hay una diferencia muy clara entre la elección de un paradigma y una prueba de carácter lógico o matemático más objetiva. En esta última, se establecen las premisas y las reglas de inferencia. Eventualmente, al haber discrepancia respecto de las conclusiones,

las partes pueden volver sobre sus pasos revisando cada premisa anterior. En el proceso, uno de las dos partes tendrá que admitir la violación de una regla, en tanto la contraparte presentará las pruebas decisivas a su favor, lo que significará su triunfo, si son correctas. En cambio, en el proceso de elección de paradigmas científicos, el debate es acerca de las premisas, y se utiliza la persuasión sobre la verdad o falsedad de estas como prelude de la eventual demostración (Kuhn, 1971: 304).

¿Cuáles son las razones, entonces, para elegir y rechazar un paradigma? Kuhn (1971: 238-246) sostiene que no hay una sola, sino más bien un conjunto de factores. En primer lugar, una comunidad científica elige un paradigma nuevo porque resuelve problemas que para el anterior resultaron muy difíciles de asimilar y que lo condujeron, finalmente, a la crisis. Mientras más efectivo es el nuevo paradigma en resolver una mayor cantidad de problemas, más probabilidades hay de que se afiance. No obstante, esto es insuficiente por sí solo y no se da en todos los casos. A modo de ejemplo, la teoría ondulatoria de la luz, que venía a reemplazar a la teoría corpuscular, no tuvo éxito en explicar los efectos de la polarización, causante principal de la crisis de la óptica, sino hasta bien avanzado su desarrollo. Sin embargo, la habilidad comparativa de un paradigma por sobre otro, en cuanto a resolver problemas, es un argumento muy importante para los científicos, aunque no sea apremiante, como ya se estableció anteriormente. En segundo lugar, existe un factor más subjetivo que se relaciona con el carácter más apropiado y estético de una teoría. Particularmente, es más notorio en las Matemáticas, y aunque solo persuada a unos pocos, su efecto puede ser relevante por el hecho de la profundización del paradigma que realiza esa minoría de científicos. En tercer lugar, está la fe en los métodos y técnicas para resolver problemas venideros, en vez de la capacidad demostrada de resolver problemas anteriores. En otras palabras, una comunidad puede decidir por un paradigma nuevo que va a guiar la investigación futura, pese a que el viejo ha resuelto una gran cantidad de enigmas y fallado pocas veces. Tal decisión no obedece necesariamente a un criterio racional, ya que también se basa en consideraciones estéticas personales e inarticuladas. Por ejemplo, la teoría general de Einstein

produce un efecto estético en científicos matemáticos imposible de percibir para un investigador ajeno a esta comunidad. Por otro lado, es necesario establecer que los criterios más bien místicos no son el fundamento principal de la elección de un paradigma. Por el contrario, son muy pocos los científicos que se guían por tales consideraciones, pero al mismo tiempo ellos podrían constituir los primeros adeptos a un paradigma, desarrollándolo y profundizándolo de manera de crear argumentos para convencer a otros. Kuhn, en este sentido, no descarta el poder de los argumentos, sino más bien los relativiza como un criterio único e infalible de elección de un paradigma. Dicho de otro modo, no existe uno o varios argumentos que tengan el poder de persuadirlos a todos por igual.

Naturalmente, en el caso del aspecto subjetivo de la conciencia, considerado como una anomalía para el PCR, no ocurre exactamente la misma situación descrita antes, porque no hay un enfrentamiento entre dos paradigmas. Solo hay uno: el PCR. Las posiciones divergentes de este paradigma provienen principalmente desde la Filosofía de la Mente, no de un paradigma rival más o menos articulado. Pero la discusión anterior intenta demostrar que los argumentos argüidos por un determinado paradigma son necesariamente circulares, y en ese sentido no deben considerarse como pruebas irrefutables para los que no compartan los criterios de ese paradigma, debido a que, precisamente, son esos criterios los que están siendo cuestionados. Para que la conciencia adquiera un estatus científico es imprescindible la aparición de un nuevo paradigma que la cobije, usando otras técnicas, métodos, metáforas y teorías que revolucionen la forma de acercarse a tal fenómeno y la enmarquen en un contexto en donde su explicación sea relevante. Ese paradigma fenoménico no existe actualmente, pero eso no indica que no llegue a desarrollarse y menos que la conciencia no pueda considerarse una anomalía en el paradigma actual, puesto que el estadio histórico de esta es anterior a la aparición de un paradigma revolucionario.

Al revisar la historia de la Ciencia Cognitiva en el marco de la investigación en Psicología, se tienen más elementos para abordar el problema de la relevancia de un fenómeno respecto de un determinado paradigma. En el paradigma conductista que imperaba en la Psicología en gran parte del siglo XX, la mente,

podría decirse, era un fenómeno distractivo o metafísico, en términos de que se presentaba demasiado problemático para su estudio, puesto que era imposible obtener información desde la tercera persona, dada las herramientas y teorías de la época. El Conductismo nace como respuesta a la Psicología introspectiva decimonónica, que intentaba acceder a la conciencia a través del autoanálisis. La introspección que llevaban a cabo individuos entrenados podía, a través de reportes, dar a conocer información sobre las experiencias mentales (Bechtel, 1998: 14). El Conductismo metodológico, en oposición a este tipo de Psicología, sostendrá que solo la evidencia observable es relevante para la ciencia (Searle, 2004: 35-36). Ello derivó en un estudio del comportamiento observable concebido como una respuesta a un estímulo. En consecuencia, la mente se transformó en un hecho explicativamente irrelevante y la Psicología, por su parte, adquirió la forma de una teoría de la conducta. Desde un enfoque kuhniano, se podría afirmar que la mente era un fenómeno demasiado problemático, distractivo, incluso metafísico para el Conductismo. Sin embargo, con el advenimiento de la Ciencia Cognitiva, la mente vuelve a ser enfocada como un problema principal. ¿Cómo, entonces, un problema irrelevante para el Conductismo, como la mente, se constituye en un enigma de otro paradigma científico? Naturalmente, en el proceso se desarrollan nuevas herramientas y nuevas teorías que logran dar cuenta del nuevo hecho, de manera que toda la investigación en Psicología sufrió un replanteamiento de los problemas que debía abordar y de la metodología para estudiarlos. Lo interesante de esto es que la historia de la Ciencia Cognitiva hace ver como posible el hecho de que un problema desprestigiado por ciertas prácticas científicas puede, en algún momento, volverse un enigma para un paradigma determinado, y esto es necesariamente como consecuencia de un paso previo por el estado de anomalía, es decir, a través de la evidencia de que un determinado fenómeno que se desajusta a las explicaciones de las teorías predominantes en una ciencia, que, en el caso del Conductismo, correspondió a la mente. Si no fuera así, no se habría manifestado nunca la necesidad de cambiar de paradigma. La mente, se podría argüir, era un hecho demasiado problemático en cuanto a obtener datos confiables desde la tercera persona y, por ello, irrelevante desde un

punto de vista de la metodología científica. Sin embargo, el cognitivismo superó el reto conductista, que consistía en obtener evidencia de lo mental desde la tercera persona objetiva, estableciendo un nuevo estadio en la Psicología. La estrategia fue recurrir a nuevos métodos y nuevas teorías que se desmarcaran de los criterios estrictamente conductuales. También se usó una nueva metáfora, la del computador. La mente fue y es entendida hasta hoy en términos del procesamiento de la información, siendo útil a los propósitos de la Ciencia Cognitiva. Los datos que se pueden obtener de un dispositivo físico como el computador son controlados, por tanto, más confiables para ser analizados en comparación con los reportes de un individuo humano. El computador puede ser estudiado desde la tercera persona, y a través del diseño de programas y su explicación en términos de lo que pasa en el cerebro es posible una descripción más acabada y confiable de este. La Historia demuestra que la Ciencia Cognitiva fue capaz de cobijar, con una revolución en las metodologías y teorías, bajo un paradigma un fenómeno irrelevante según las preocupaciones científicas.

Hasta cierto punto, la Ciencia Cognitiva pudo adecuarse a los requerimientos del Conductismo. El cambio consistió en trasladar la evidencia desde la conducta a la mente, pero siempre en el marco de la observación objetiva. En este sentido, ambos paradigmas, tanto el viejo como el revolucionario, tenían puntos en común en cuanto a este aspecto de sus metodologías. Ninguno de los dos quería renunciar a la tercera persona como modo de aproximación a la realidad. Por tanto, ambos están comprometidos con el criterio común de la objetividad física. Desde este punto de vista, la incompatibilidad de ambos paradigmas se ve atenuada considerando este elemento. La pregunta a responder es si el nuevo paradigma, aquel que le daría cabida a la conciencia fenoménica, podría responder también al reto de la evidencia observable desde la tercera persona y, que al mismo tiempo, eso implique la necesidad de un paradigma revolucionario como sucedió con la Ciencia Cognitiva. Esta cuestión es particularmente problemática, puesto que está directamente relacionada con el estatus de la anomalía y con la necesidad de una eventual revolución.

¿Cuál sería el reto que la Ciencia Cognitiva plantearía a un eventual paradigma competidor? Al aceptar que la conciencia fenoménica constituye una anomalía para el PCR de la mente, entonces el nuevo paradigma debería explicarla o generar, al menos, una expectativa de solución. En este punto existe un peligro si se plantea la discusión en los términos del paradigma viejo, pues el debate puede hacerse circular, como lo ha señalado Kuhn. Si un nuevo paradigma, respondiendo al reto cognitivista, le otorgara una relevancia causal a la conciencia, probablemente la conciencia subjetiva perdería su estatus de anomalía, ya que el carácter causal es útil a la concepción computacional-representacional. Si la conciencia fenoménica puede ser reducida usando un criterio causal, entonces las viejas herramientas, métodos, técnicas y metáforas podrían explicar, sin alterar nada (no habría por qué hacerlo), su función en un sistema de procesamiento de la información. Se estaría en presencia de una rearticulación del viejo paradigma, a lo sumo. Por tanto, al responder esta vez en los mismos términos al reto cognitivista, como la Ciencia Cognitiva lo hizo con el Conductismo, se caería en la trampa de resituar a la conciencia en términos de un enigma científico, en el marco de la ciencia normal, y no habría revolución alguna. Es decir, la conciencia volvería a ser un fenómeno observable desde la tercera persona, y se la describiría en términos de *lo que hace*. Sin embargo, aquello sería del todo inútil, pues la conciencia tiene una ontología de primera persona, como se estableció en el capítulo 1, y no parece encajar en ninguna concepción de objetividad física, al menos tal como se la entiende tradicionalmente. Esta no logra enfocar el fenómeno, que se define estrictamente respecto de lo *que se siente*, esto es, subjetivamente. En este sentido, el reto del próximo paradigma va más allá de revolucionar los criterios y técnicas de una disciplina en particular. Más bien, si se quiere dar cuenta a cabalidad de la conciencia, debe renovarse por completo la concepción física de la realidad predominante en la tradición científica. Sea a través del desarrollo de una objetividad mental, en donde la mente sea explicada como una característica general de la naturaleza, pero en términos mentales (véase Nagel, 1986: 17-19), o por la creación de nuevos patrones de

reducción propuestos por futuros y revolucionarios enfoques científicos (Searle, 1997. Véase también 1.3 de esta tesis).

Resumiendo el punto establecido en esta sección, la crítica de la irrelevancia de la conciencia como problema científico se hace a partir de los criterios del PCR. El peligro de esta situación es que el debate pueda hacerse circular, ya que el viejo paradigma pone sus propias reglas para determinar los problemas científicos. En este sentido, no deben considerarse como pruebas irrefutables para los que no compartan los criterios de ese paradigma, debido a que, precisamente, son esos criterios los que están siendo cuestionados. Adicionalmente, existe el peligro de que, al plantearse la discusión en los términos exactos del paradigma en crisis, el nuevo fenómeno, la anomalía, *tenga que ser* asimilada por el viejo paradigma. Es decir, si la conciencia subjetiva fuera explicada exhaustivamente de acuerdo con la función que desempeña en la cognición, no sería necesario un cambio de paradigma. No habría razón para ello. Es más, si cualquier fenómeno tuviera que obedecer los requerimientos del paradigma viejo, no ocurriría nunca una revolución científica, cuestión, por cierto, que no se ajusta a la realidad.

Pese a que el debate sobre la irrelevancia de la conciencia corre el peligro de tornarse circular, es necesario de todas formas analizar por qué la experiencia subjetiva sería importante para un futuro paradigma. Naturalmente que una discusión así debe plantearse considerando la relevancia del aspecto subjetivo como un fenómeno de la realidad y no necesariamente situar el debate según los criterios del PCR, como ya se señaló más arriba. La cuestión es compleja, porque no existe todavía un paradigma que cobije a la conciencia subjetiva como fenómeno de estudio, por lo que no hay una situación en la que se tenga que elegir un paradigma entre dos candidatos. Establecido este punto, y sumado a la posibilidad de circularidad en el debate, el *conocimiento* que se obtiene en la experiencia subjetiva podría justificar su relevancia. Es decir, la experiencia subjetiva entrega un conocimiento imposible de obtener desde la tercera persona objetiva. Esta cuestión está implicada en el experimento de Nagel (1974), pero es abordada con particular énfasis por Jackson (1982 y 1986) en dos EMs. En el

primero de ellos, Jackson (1982) presenta a Fred, un hombre capaz de distinguir matices en los colores que para las personas comunes resulta imposible. Eventualmente, Fred realiza la tarea de clasificar tomates maduros en dos grupos, lo que hace de una forma consistente. Luego, se le vendan los ojos y se procede a desordenar los tomates previamente clasificados por él. Al quitarle la venda, Fred nuevamente comienza a clasificar los tomates de la misma forma como lo había hecho antes, y con la misma consistencia. Cuando se le pregunta cómo lo hace, Fred responde que no le parece que todos los tomates que ordenó tengan el mismo color rojo. Sin embargo, para el resto de la gente los tomates parecen tener idéntico color. En otras palabras, Fred ve dos colores, dos matices de rojo, donde el resto ve solo uno. Además, cuando se describe físicamente el funcionamiento de su cerebro y de su sistema óptico, y se descubre que los conos responden de una forma diferente respecto de las ondas de luz, o que tiene quizá un tipo de cono extra, como causa de su capacidad para distinguir colores, lo que se hace es explicar qué conduce a Fred a hacer tales distinciones. Sin embargo, no es posible saber acerca de la experiencia de percibir dos tipos de rojo. Aunque se describa todo su cuerpo físicamente, su conducta, su fisiología interna, incluso su historia y su relación con los otros, la experiencia es inaccesible desde estas aproximaciones. Por tanto, saber todo sobre la descripción física de Fred no significa saber todo sobre Fred. El aspecto imposible de describir en términos físicos son los *qualia*, término que Jackson usa para describir las propiedades cualitativas de los estados mentales, y cuya posesión o ausencia por parte de un individuo no hacen ninguna diferencia en el mundo físico, además de ser causalmente ineficaces, escapando a un criterio funcional y constituyéndose, por ello, en un verdadero epifenómeno (Jackson, 1982: 276). En otras palabras, los *qualia* señalan aquellas propiedades referidas por el concepto de conciencia fenoménica (Block) o el aspecto *what it is like* (Nagel, 1974). De ahí se sigue que si este tipo de propiedades no puede ser conocido a través de la información física que nos entrega, por ejemplo, la fisiología, entonces el Fisicalismo es falso. Jackson aquí usa el término Fisicalismo para referirse a lo que Nagel (1986) denomina la concepción física de la objetividad, esto es, una descripción completa

del mundo en términos puramente físicos (véase 1.3). También se puede concebir el hecho de que se logre hacer similar la fisiología de todas las personas a la de Fred, debido a adelantos médicos, o que se trasplante su sistema visual a otra persona, después de su muerte. De esta manera, sería posible saber qué se siente tener las experiencias de Fred y experimentar su diferenciación de los tipos de rojo. Por tanto, después de la operación se hace posible saber más sobre Fred y especialmente sobre sus experiencias. No obstante, antes de la operación se tenía toda la información que una descripción fisicalista podía dar de su cerebro, pero aún así no se podía tener toda la información. En conclusión, el Fisicalismo da una descripción incompleta del mundo, cuestión compatible con la crítica de Nagel (1986).

También Jackson (1982 y 1986) presenta un segundo EM para probar que la experiencia subjetiva es inaprensible desde una descripción planteada en términos físicos tradicionales. En este segundo EM se plantea una situación en que Mary, una científica brillante especializada en la neurología de la visión, ha sido puesta a investigar el mundo desde una habitación en blanco y negro, y usando una pantalla también de esas características. Mary adquiere toda la información física que hay que tener cuando las personas ven, por ejemplo, los tomates maduros, o el cielo, y usa términos para referirse a los colores como *azul*, *rojo*, *amarillo*, etc. Ella descubre qué combinaciones de longitudes de onda estimulan la retina. Adicionalmente, sabe cómo esto causa, a través del sistema nervioso central, la contracción de las cuerdas vocales y la expulsión de aire desde los pulmones que originan, finalmente, la oración *el cielo es azul*. Mary, al salir de la habitación o al poder ver imágenes a color en la pantalla, aprende la experiencia de los colores, cuestión imposible de obtener con la sola descripción física de cualquier sistema visual. En otras palabras, Mary aprende algo acerca del mundo, por tanto, antes de someterse a la experiencia subjetiva su conocimiento era incompleto.

Pese a que Jackson (1982: 275-276) no lo ve exactamente de la misma manera, el EM de Nagel (1974) tiene la misma función, aunque este último abarca más temas. Según Jackson lo fundamental en el EM de Nagel es la extrapolación

de una experiencia en otra, donde se discute la incapacidad de la imaginación humana para dar cuenta de la experiencia subjetiva de otro organismo como el murciélago, y por ello sería diferente de la discusión planteada en los EMs de *Fred* y *Mary*. Aunque esto es cierto en parte, con el experimento del murciélago también se sostiene que una concepción física de la objetividad no puede dar cuenta de la experiencia subjetiva, porque esta se plantea desde la tercera persona, siendo, por tanto, incapaz de abordar el aspecto fenoménico. En otras palabras, para Nagel el Fisicalismo hace una descripción incompleta de la realidad, en este caso, una descripción incompleta de un murciélago. Es imposible saber qué es como tener la experiencia de un organismo de esas características, aunque se tenga una descripción acabada de su estructura fisiológica. Por tanto, se puede concluir lo mismo que de los EMs de Jackson, a saber, que la experiencia subjetiva se deja fuera de una descripción fisicalista y, por tanto, no se puede acceder a todo el conocimiento sobre la realidad desde la tercera persona objetiva.

3.6.2 ¿Por qué la conciencia es una anomalía si es un fenómeno familiar y cotidiano?

Otra crítica que podría surgir, respecto de la consideración de la conciencia subjetiva como una anomalía en la Ciencia Cognitiva, es sobre su estatus de hecho inesperado. En otras palabras, si la conciencia subjetiva es un fenómeno tan familiar, no solo para los científicos sino para cualquier persona, ¿cómo se constituye en hecho anómalo? Es decir, ¿por qué es tan sorprendente? Para responder estas preguntas es necesario volver sobre las características del concepto mismo de *anomalía*. Kuhn (1971) lo describe como un hecho que ha violado las expectativas de una comunidad científica incapaz, con sus herramientas, métodos y teorías, de acercarse a un determinado problema. En otras palabras, la anomalía “es un fenómeno para el que el investigador no estaba preparado” (Kuhn, 1971: 100). ¿Acaso un científico no va a estar preparado frente a un elemento de la naturaleza tan familiar, tan cotidiano como la experiencia subjetiva?

En primer lugar, es pertinente remarcar que el científico aprende a ver la naturaleza desde su propio paradigma. En otras palabras, su paradigma establece cuáles deben ser los enigmas, es decir, los problemas en los que se fija la atención. Además, el paradigma señala cuáles son los métodos estandarizados para hacerlo y cómo deben resolverse. En este sentido, el paradigma le indica al científico cuál es la sección de la naturaleza que debe estudiar. Si el fenómeno no reside en el marco que se da desde un paradigma, entonces ese hecho no es científico, y la comunidad lo ignora. De alguna manera, el nuevo hecho no es visible, porque las herramientas de una disciplina son incapaces de afrontarlo, y es en este sentido que el científico no está preparado para describirlo. En *su* mundo no hay lugar para la anomalía. Pero no debe pensarse que el hecho anómalo es sorprendente dado que es algo nunca visto antes necesariamente, o un aspecto del que jamás se haya sabido. Por el contrario, la anomalía problematizada por los EMs es un dato a mano, familiar en algún sentido, pero es, al mismo tiempo, completamente inaccesible (Kuhn, 1982: 285). La anomalía, fundamental para que ocurra la revolución científica, corresponde a esa clase de datos “que han estado existiendo al borde de la conciencia científica” (Kuhn, 1982: 287). Lo que hace un EM, en el contexto de la crisis de un paradigma, es, precisamente, poner en el centro de atención ese fenómeno que había sido aislado de la ciencia normal (Kuhn, *ibídem*). En este punto, entonces, la experiencia subjetiva no ha formado parte de los problemas del PCR. La concepción computacional de la mente no le ha atribuido ningún rol especial a la conciencia fenoménica, y quizá tenga que hacerlo así porque la conciencia excede la explicación en esos términos (Thagard, 1996: 146-147). Hasta ahora, no habido razón, en el marco del PCR, para atribuir conciencia fenoménica a los computadores porque en estos se simula el pensamiento humano sin apelar a estados fenoménicos. Existen modelos computacionales para simular resolución de problemas, aprendizaje y lenguaje, por ejemplo. Considerando esto, la del PCR es una teoría inadecuada porque no alcanza a abordar fenómenos que sí tienen lugar en la vida mental humana, que exceden los alcances de cualquier simulación desarrollada actualmente. Por esta razón, la metáfora del computador no es suficiente, como estrategia metodológica,

para describir la mente. Por su parte, los partidarios del PCR podrían afirmar que la conciencia es irrelevante, ya que no es necesaria para la simulación del pensamiento humano. Pero el error en esta última consideración es hacer de la metáfora del computador quizá algo bastante literal, como señala Von Eckardt (1993: 116). Esta autora, aunque señala esta posibilidad, no toma una posición a favor o en contra de este hecho. Sin embargo, naturalmente el riesgo es que de esta forma los partidarios del PCR no se sientan impulsados a dar una explicación en términos mentales humanos, pues su interpretación de los procesos cognitivos humanos es probablemente que funcionan realmente como procesos computacionales.

Por otro lado, las reticencias a aceptar el carácter anómalo de la conciencia teniendo en cuenta su familiaridad surgen, sobre todo, al dejarse llevar por la intuición de que el descubrimiento científico, contexto en el que se da la anomalía, es algo que ocurre en un lugar y en un instante. Pero lo que se entiende comúnmente por descubrimiento no es un evento con una fecha, un lugar y un autor, necesariamente, ni es algo tan simple que pueda asimilarse al concepto cotidiano de la visión o del sistema perceptual humano (Kuhn, 1971: 97). Esto es, descubrir algo no es ver o tocar ese algo que nunca había aparecido antes. Por el contrario, el descubrimiento es un proceso, en cuyos pasos se va identificando el fenómeno y explicando su naturaleza, por tanto, “involucra el reconocimiento, tanto de *que* algo existe como de *qué* es” (Kuhn, *ibídem*, cursiva en el original). Y completar esa tarea necesariamente toma tiempo.

Para entender de una forma más global la discusión presente, el fenómeno de la anomalía debe analizarse en el contexto del descubrimiento científico, pues forma parte constitutiva de este. Como todo proceso, el descubrimiento tiene fases, y a juzgar por Kuhn (1971: 93 y 107) son tres, a saber:

1. *La percepción previa de la anomalía*: existe un reconocimiento de que la naturaleza ha violado las expectativas de la investigación en el contexto de la ciencia normal.

2. *La aparición gradual y simultánea del reconocimiento tanto conceptual como de observación del hecho anómalo*: se desarrolla una exploración de la zona de la anomalía, analizando sus características.

3. *El cambio consiguiente de las categorías y los procedimientos del paradigma, acompañados a menudo por resistencia*: se realiza el ajuste de la anomalía a través de una explicación dada en el contexto de un paradigma nuevo. De esta manera, lo que antes parecía no tener respuesta, en el marco de un paradigma viejo, adquiere una explicación a través de un nuevo enfoque científico. Sólo en esta fase el nuevo fenómeno adquiere el estatus de científico.

En definitiva, el proceso del descubrimiento comienza con la anomalía y se completa con el ajuste de las categorías conceptuales. Así, lo inicialmente anómalo se convierte en lo previsto (Kuhn, 1971: 110). Hasta que no se completa el descubrimiento, el nuevo hecho es precientífico. Por tanto, el nuevo hecho requiere de los conceptos adecuados que lo describan, y en este sentido se puede afirmar que el descubrimiento no ocurre en un momento. Es decir, las teorías, los conceptos, la observación y los hechos están entrelazados inseparablemente.

Por supuesto, si se considera lo que intuitivamente se entiende como descubrimiento, parece erróneo afirmar que el descubrimiento tiene lugar en un momento específico. Pero como se explicó en los párrafos precedentes, un nuevo hecho necesita de los conceptos, y la conciencia ha carecido de tales elementos que hubieran podido, eventualmente, situarla en un marco investigativo. Al considerar las fases que Kuhn describe para el descubrimiento científico, la conciencia recién está en la 2, es decir, en la exploración conceptual. Antes de eso, la conciencia no asomaba, al menos seriamente, en los debates científicos, y podía evitárselo como problema por su complejidad, ya que es un fenómeno que tiene una ontología de primera persona, inaprensible desde una concepción de objetividad física de la ciencia tradicional que se sitúa desde la tercera persona. Este tipo de problemas, que no encajan en un paradigma determinado, a veces

surgen en la investigación, pero se los descarta, y a menudo se les tilda de “metafísicos” o “demasiado problemáticos” (Kuhn, 1971: 71).

La incipiente teorización sobre la conciencia ha tenido lugar con especial énfasis en la Filosofía de la Mente, con la identificación de conceptos como la *conciencia fenoménica*, el aspecto *what it is like*, *el hard problem* y *qualia*, fundamentalmente (véase capítulo 1), que han intentado identificar y describir el aspecto subjetivo de la conciencia. En otras palabras, aunque se haya tenido la percepción intuitiva y familiar de la conciencia históricamente, no se había podido encauzarla como un problema relevante para la ciencia, porque no existían los conceptos, o al menos su discusión, para ello. Además ni siquiera existía, antes de la Ciencia Cognitiva, una disciplina que estudiara, de forma consolidada, lo mental. Un fenómeno para que se constituya como descubrimiento necesita de teorías que lo encaucen y lo expliquen, de esa manera se sabrá qué es, y solo de esa manera es un descubrimiento y un hecho científico. Al respecto Kuhn afirma: “Sólo cuando todas las categorías conceptuales pertinentes están preparadas de antemano, en cuyo caso el fenómeno no será de un tipo nuevo, podrá descubrirse sin esfuerzo *qué existe y qué es*, al mismo tiempo y en un instante” (Kuhn, 1971: 97, cursiva en el original). De esta manera, al hacer la distinción de la conciencia fenoménica o *qualia*, lo que se hace es identificar un concepto pero también identificar un aspecto de la naturaleza. Al plantearse tal distinción, es posible darse cuenta de que el PCR no da cuenta de ese fenómeno, que, a su vez sí forma parte de la realidad mental humana.

¿Por qué, entonces, la conciencia surgió como una anomalía en el marco de la Ciencia Cognitiva y no antes? El paradigma inmediatamente anterior al cognitivismo, es decir, el Conductismo, consideraba la mente explicativamente irrelevante. La Ciencia Cognitiva pudo, a través de una revolución metodológica, estudiar parte del fenómeno mental, lo que se conoce como el *easy problem*. Es en ese contexto que la conciencia subjetiva, situado en un marco en donde parte de los procesos mentales eran estudiados, que su importancia emergió de una forma que sin este paradigma hubiera sido imposible: “La anomalía sólo resalta contra el fondo proporcionado por el paradigma” (Kuhn, 1971: 111). En otras

palabras, el mismo paradigma que no puede abordar un fenómeno, que está en sus márgenes, es el contexto perfecto en que ese hecho anómalo resalta aún más. La Ciencia Cognitiva sí estudia la conciencia, pero aquella parte que Chalmers (1996) denominó el *easy problem*. Este marco hace posible notar que la conciencia trasciende dicho enfoque, porque es algo más que un elemento de integración sensorial y que también posee un aspecto subjetivo, esencial a su naturaleza.

Desde una perspectiva histórica, el desarrollo de la conciencia como problema puede entenderse en tres etapas fundamentales. En primer lugar, un periodo en que era un problema demasiado complejo para ser estudiado con las herramientas y teorías disponibles. Con propiedad se podría sostener que tales características lo convertían en un problema metafísico o en un fenómeno demasiado problemático. El Conductismo metodológico sería el último estadio de esta concepción. En segundo lugar, aparece una etapa definida por el estudio del *easy problem*, es decir, aquel componente de la conciencia que cumple la función de elemento integrador de los distintos sistemas sensoriales, que va desde Broca y Wernicke hasta los actuales correlatos neuronales de la conciencia en el marco de la Ciencia Cognitiva. Finalmente, está la etapa en que la otra parte de la conciencia, su aspecto fenoménico, que no puede ser ajustada en un enfoque representacional-computacional, se constituye como anomalía en el marco del paradigma cognitivista.

Conclusiones

El PCR de la mente, enmarcado en la Ciencia Cognitiva, pertenece a una aproximación reduccionista de la realidad. Concretamente, lo que hace este paradigma es entender la conciencia en términos de sus habilidades y funciones, por tanto, solo se encarga del *easy problem*. Pero la conciencia subjetiva excede estos planteamientos y no puede explicarse sobre la base de la realización de determinadas consideraciones causales. El *hard problem* persiste incluso cuando se describen todas las funciones relevantes de la cognición. En otras palabras, una concepción computacional y representacional intenta explicar la conciencia desde una perspectiva de tercera persona, esto es, en términos de *lo que hace*, pero no necesariamente en cuanto a *lo que se siente* (Güzeldere, 1997).

En este contexto, la concepción del desarrollo de la ciencia de Thomas Kuhn (1971 y 1982) ha resultado ser adecuada para identificar el estatus del problema de la conciencia en el PCR de la Ciencia Cognitiva. Desde esta perspectiva, la experiencia subjetiva puede entenderse como una anomalía, es decir, como un hecho inesperado para la comunidad científica, para el cual las herramientas y teorías disponibles dentro del PCR no dan ninguna solución ni expectativa de llegar a lograrla. El carácter subjetivo de la experiencia no es un fenómeno para el cual una concepción reduccionista esté preparada. La función del EM de Nagel (1974) es identificar y problematizar dicha anomalía, porque pone en un lugar central el aspecto subjetivo. La estrategia de la que se vale este experimento en particular es explorar los límites de la imaginación humana para concluir que, aunque se tenga una descripción fisiológica y funcional de un organismo, en este caso un murciélago, es imposible saber qué se siente ser ese organismo. Por tanto, una descripción desde la tercera persona, metodología de la ciencia tradicional, y específicamente del PCR en este caso, es insuficiente para abordar el problema de la subjetividad. Por esta razón, se hace necesaria la aparición de un nuevo paradigma, a través de una revolución científica, que logre explicar la conciencia a partir de la creación de nuevas teorías, metodologías y

metáforas. Respecto de estas últimas, la del computador, fructífera en las explicaciones de la mayoría de los procesos cognitivos, ha demostrado ser insuficiente para dar cuenta de la conciencia en toda su complejidad. En este sentido, para el EM de Nagel se aplica lo expuesto por Kuhn (1982) y Gendler (1998) respecto de la función que tiene este tipo de experimentos en la crisis de un paradigma. Según Kuhn (1982), los EMs, como los experimentos de laboratorio, conducen a un conocimiento nuevo o a una nueva comprensión de la naturaleza, aunque se basen en información conocida. La respuesta a esta paradoja kuhneana estriba en que los EMs le entregan al científico acceso a una información que tiene a mano, pero que de alguna manera le resulta inaccesible. Su inaccesibilidad se explica debido a que el paradigma científico determinado, en este caso el PCR, obliga al investigador a centrarse en los enigmas, es decir, en aquellos problemas enmarcados en la ciencia normal. En otras palabras, el PCR determina los enigmas respecto de su adecuación a un criterio causal y funcional. De esta manera, la conciencia subjetiva no tiene lugar en el conjunto de problemas científicos que la Ciencia Cognitiva estudia, aunque sea un fenómeno familiar, o “a mano” en palabras de Kuhn (1982: 284). Gracias a los EMs, como el de Nagel o los de Jackson (1982 y 1986), un aspecto anteriormente ignorado, inarticulado o recluso en los márgenes de las preocupaciones científicas, como la conciencia subjetiva, logra ser sistematizado bajo una nueva perspectiva (Gendler 1998). Los EMs ayudan al científico a considerar la importancia de un hecho inaccesible, por lo cual cambia su conocimiento del mundo. Ese hecho periférico, que no encajaba en el paradigma anterior, es visto de una manera diferente. En conclusión, los EMs evidencian la anomalía en el marco de un paradigma científico, lo que demuestra que la función de los EMs postulada por Kuhn, y desarrollada después por Gendler, se ajusta plenamente al papel que tiene el EM de Nagel, y también los de Jackson, en el contexto del PCR de la mente.

Respecto de la discusión sobre si los EMs son meros argumentos (Norton, 2004), o mecanismos de la imaginación que aportan un nuevo conocimiento de la naturaleza (Brown, 1991), la presente tesis aboga por lo segundo. Los EMs obligan a una participación constructiva del lector (Gendler, 1998), lo que le da una

fuerza justificativa adicional que un simple argumento no posee, constituyendo de esta manera una crítica frente a lo que postula Norton. El lector se ve impulsado a realizar, más que un EM, un *experimento-en-pensamiento*, como sucede en el EM de Nagel (1974), por medio del cual se pregunta a sí mismo qué esperaría al existir determinadas circunstancias. Dada esa situación inicial, el lector-científico intenta encontrar la respuesta, y a través del desarrollo de la situación imaginaria, llega a una creencia que él estima verdadera porque la experimentó de forma directa. O dicho de una forma más simple aún, la gente cree lo que hace. La fuerza justificativa se explica, en consecuencia, porque las personas tienden a creer en cosas y situaciones que experimentan directamente, por ello un conjunto de premisas de un argumento no tiene este poder al no implicar una participación constructiva del lector, como sí ocurre en un EM. De esta manera, un EM no puede ser reducido solo a un argumento.

Por otro lado, en cuanto al estatus mental del experimento de Nagel (1974), es posible sostener que este posee características que lo distinguen de otros EMs, porque también tiene características de los experimentos físicos de laboratorio. Irvine (1991: 153) sostiene que una de las diferencias entre los EMs y los físicos es que estos últimos dependen de una intervención real en la naturaleza, mientras que los primeros adquieren la forma de un argumento basado en premisas hipotéticas. Si se comprende que la conciencia y su dimensión fenoménica forman parte de la naturaleza, la diferencia parece no ser tan clara al considerar el experimento de Nagel. Cuando se apela al lector para hacer el esfuerzo imaginativo de sentir como un murciélago, de alguna manera se experimenta de forma real con el aspecto subjetivo, por tanto, podría afirmarse que existe intervención directa en la naturaleza, aunque no haya una transformación o modificación propiamente tales, como ocurriría en un experimento de laboratorio. A diferencia del EM de Galileo sobre los cuerpos que recorren dos planos distintos, en donde el razonamiento argumentativo de carácter contrafáctico ocupa un lugar fundamental, con el EM del murciélago lo que se pretende es poner en el centro la experiencia subjetiva, la percepción misma del experimentador, no como una situación hipotética, sino como un elemento constitutivo de la realidad de un

organismo. Pese a que adicionalmente se apele a una situación imaginaria, lo central es la imposibilidad de la percepción de ese hecho y no la situación contrafáctica en sí. Un experimento planteado de esta manera permite el acceso a la conciencia en una situación de experimentación y no se queda solo en un razonamiento hipotético. De ahí la diferencia del experimento de Nagel respecto de otros EMs. Tal acceso privilegiado está dado por su modo de aproximación, el punto de vista de la primera persona, esencial característica de la conciencia subjetiva. Por tanto, si bien el experimento de Nagel posee elementos contrafácticos, que son una característica típica del género de los EMs, hay una intervención directa en la naturaleza, una intervención entendida no como una modificación, sino en términos de una exploración de la subjetividad, que, por cierto, es un hecho perteneciente al mundo físico. Esa intervención estriba en la exploración de los límites de la imaginación humana, que irremediamente fracasa en alcanzar la experiencia fenoménica de un murciélago, lo que sitúa a esta experiencia fenoménica como un elemento irreductible desde la tercera persona objetiva, e inalcanzable, al mismo tiempo, desde la comprensión de otra mente con una estructura fisiológica muy diferente a la del murciélago. Por tanto, estos elementos hacen del experimento de Nagel un EM híbrido, debido a sus elementos tanto contrafácticos como experienciales. En este sentido, los términos de *experimento-en-pensamiento*, como plantea Gendler (1998: 414), o de *experimento introspectivo* podrían ser más apropiados para dar cuenta de experimentos como el de Nagel.

Respecto de la relevancia del aspecto subjetivo de la conciencia, se presenta como una cuestión problemática dada la posibilidad de caer en un debate circular si se consideran sin cuestionamientos los criterios que intenta imponer del viejo paradigma. Desde el punto de vista del PCR, la conciencia es irrelevante porque no posee un carácter causal ni funcional, y así se intenta desacreditar la conciencia subjetiva como problema científico. Esta es la manera en que el PCR de la mente pone sus propias reglas para determinar los enigmas a ser resueltos. La circularidad estriba en que, al obedecer a esos criterios, no habría la necesidad de una revolución científica. Si se encuentra una función a la

conciencia que encaje en un criterio computacional y representacional, no sería necesario un nuevo paradigma. Aquí es evidente que cada bando argumenta considerando sus propias creencias en ciertos modelos sobre la naturaleza de las cosas. Según Kuhn (1971), los criterios provenientes de un paradigma en crisis no deben considerarse como pruebas irrefutables para los que no compartan los criterios de ese paradigma, debido a que, precisamente, son esos criterios los que están siendo cuestionados. Por supuesto, aunque pueda desarrollarse una argumentación circular en la defensa de un paradigma, no significa necesariamente que esos argumentos sean erróneos o inefectivos. Pero el estatus de un argumento de estas características es solo el de la persuasión, y no obliga, desde un punto de vista lógico, a la elección de un determinado paradigma. Ahora bien, una forma de responder a la relevancia de la experiencia subjetiva, fuera de la discusión circular del PCR, podría incluir el *argumento del conocimiento* (Jackson, 1982). La información de la experiencia subjetiva no se obtiene de la descripción fisiológica de un organismo, por tanto, esta última es incompleta. En otras palabras, una concepción física de la objetividad puede entregar una descripción fisiológica de un individuo, pero fracasa en explicar los *qualia* o el aspecto *what is it like*, esto es, la experiencia subjetiva que tiene lugar en la percepción, como ya lo había demostrado Nagel (1974) además de Jackson (1982 y 1986) con sus dos EMs, el de *Fred y Mary*. Este último presenta a una científica, Mary, experta en la neurofisiología de la visión pero que solo ve la realidad desde una habitación en blanco y negro. Mary obtiene toda la información que una descripción física de los procesos visuales puede dar. Sin embargo, cuando esta misma científica vuelve a ver la realidad en colores se da cuenta que hay un aspecto que la fisiología de la visión no le entregó, a saber, los *qualia* de los colores. Por tanto, una concepción física de la objetividad no enseña todo lo que hay que saber acerca de la realidad. En consecuencia, la descripción que se realiza respecto de lo mental en el PCR, en el contexto de aquella concepción, es incompleta, porque desconoce la experiencia subjetiva de los fenómenos mentales.

Bibliografía

Bechtel, W. et al. (1998) "The Life of Cognitive Science", en W. Bechtel y G. Graham (eds.) *A Companion to Cognitive Science*. Malden, Mass., Blackwell, pp. 1-104.

Blackmore, S. (2005) *Consciousness. A very Short Introduction*. New York, OUP.

Block, N. (2002) "Concepts of Consciousness", en D. J. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford, OUP, pp. 206-18.

Boden, M. (1981) *Minds and Mechanisms: Philosophical Psychology and Computational Models*. Ithaca, N. Y., Cornell University Press.

Boyd, R. (1979) "Metaphor and theory change: What is 'metaphor' a metaphor is?", en A. Ortony (ed.) *Metaphor and Thought*. Cambridge, Cambridge University Press, pp. 481-532.

Brewer, W. (1999) "Schemata", en R. Wilson y F. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass., MIT press, pp. 729-730.

Brown, J. (1991) *The Laboratory of Mind*. Londres, Routledge.

_____. (2004) "Why Thought Experiments transcend Empiricism", en C. Hitchcock *Contemporary Debates in Philosophy of Science*. Malden, Blackwell, pp. 23-43.

_____. (2008) "Thought Experiments", en: Edward N. Zalta (2008, ed.) *The Stanford Encyclopedia of Philosophy*. Disponible en línea en: <http://plato.stanford.edu/archives/win2008/entries/thought-experiment>

Chalmers, D. (1995) "Facing up to the Problem of Consciousness", *Journal of Consciousness Studies* 2, 200-219. Reimpreso en J. Heil (ed.) *Philosophy of Mind. A Guide an Anthology*. Nueva York, OUP, pp. 617-640.

_____. (1996) *The Conscious Mind*. Oxford, OUP.

Copeland, J. (1993) *Artificial Intelligence. A Philosophical Introduction*. Oxford, Blackwell.

Crick, F. y C. Koch (2002) "The problem of consciousness", *Scientific American*, 10-17.

Damasio, A. (2002) "How the Brain Creates the Mind", *Scientific American*, 4-9.

_____. (2003) "The person within", *Nature* 423, 227.

Davis, M. (1999) "Consciousness", en R. Wilson y F. Keil (eds.) *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass., MIT press, pp. 190-193.

Feldman, J. (1985) "Connectionist Models and their Applications: Introduction", *Cognitive Science* 9, 1-2.

Fodor J. y Z. W. Pylyshyn. (1988) "Connectionism and cognitive architecture: A critical analysis", en S. Pinker y J. Mehler (eds.) *Connections and Symbols*. Cambridge, Mass., MIT press, pp.3-71.

Gendler, T. (1998) "Galileo and the Indispensability of Scientific Thought Experiment", *The British Journal for the Philosophy of Science* 49, 397-424.

González, R. (1999) *Percepción - Teorías: Desde el Positivismo Lógico hacia la Ciencia Cognitiva*. Tesis para optar al Grado de Magister en Filosofía con mención en Epistemología. Santiago de Chile. Facultad de Filosofía y Humanidades, Universidad de Chile.

Güzeldere, G. (1997) "The Many Faces of Consciousness: A Field Guide", en N. Block et al. (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, Mass., MIT press, pp. 1-67.

Hanson, N. R. (1977) *Patrones de descubrimiento. Observación y explicación*. Madrid, Alianza Editorial.

Haugeland, J. (1981) *Mind Design*. Cambridge, Mass., Bradford/MIT Press.

_____. (1985) *Artificial Intelligence: The Very Idea*. Cambridge, Mass., Bradford/MIT Press.

Irvine, A.D. (1991) "Thought experiments in scientific reasoning.", en T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Maryland, Rowman and Littlefield Publications, pp. 149-65.

Jackson, F. (1982) "Epiphenomenal Qualia", *Philosophical Quarterly* 32: 127-136. Reimpreso en D. J.Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford, OUP, pp. 273-280.

_____. (1986) "What Mary didn't know", *The Journal of Philosophy* 83 (5): 291-295. Reimpreso en T. O'Connor and D. Robb (eds.) *Philosophy of Mind: Contemporary Readings*. Londres, Routledge, pp. 458-63.

Johnson-Laird, P. (1988) *The Computer and the Mind: An Introduction to Cognitive Science*. Cambridge, Mass., Harvard University Press.

Kuhn, T. (1971) *La estructura de las revoluciones científicas*. México D.F., FCE.

_____. (1982) "La función de los experimentos imaginarios", reimpreso en *La tensión esencial*. México D.F., FCE , pp. 263-289.

Logothetis, N. (2002) "Vision: a window on consciousness", *Scientific American*, 18-25.

McCauley, R (1999) "Reductionism", en R. Wilson y F. Keil (eds.) *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass., MIT press, pp 712-714.

Nagel, T. (1974) "What is it like to be a bat?", *Philosophical Review* 83, 435-50. Reimpreso en D. J.Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford, OUP, pp. 219-26.

_____. (1986) *The View from Nowhere*. Nueva York, OUP.

Nebel, B. (1999) "Frame-Based Systems", en R. Wilson y F. Keil (eds.) *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass., MIT press, pp. 324-326.

Newell, A. (1980) "Physical symbol systems", *Cognitive Science* 4, 135-183.

_____. (1990) *Unified Theories of Cognition*. Cambridge, Mass., Harvard University Press.

Norton, J. (2004) "Thought Experiments: Is There More To The Argument?", *Philosophy of Science* 71, 1139-1151.

Rosenthal, D. (1997) "A Theory of Consciousness", en N. Block et al. (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, Mass., MIT press, pp. 729-753.

Rumelhart, D. (1989) "The Architecture of Mind: A connectionist Approach", en J. Haugeland (ed.) *Mind Design II*. Cambridge, Mass., MIT press, pp. 205-232.

Roskies, A. (1999) "The Binding Problem", *Neuron* 24, 7-9.

Searle, J. (1980) "Minds, brains and programs", *Behavioral and Brain Sciences* 3, 417-24. Reimpreso en M. Boden (ed.) *The Philosophy of Artificial Intelligence*. Oxford, OUP, pp. 67-88.

_____. (1997) "Reductionism and the Irreducibility of Consciousness", en N. Block et al. (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, Mass., MIT press, pp. 451-459.

_____. (2004) *Mind: A Brief Introduction*. Oxford, OUP.

Sloman, S. (1999) "Cognitive Architecture", en R. Wilson y F. Keil (eds.) *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass., MIT press, pp. 124-126.

Smolensky, P. (1989) "Connectionist Modeling: Neural Computation/Mental Connections", en J. Haugeland (ed.) *Mind Design II*. Cambridge, Mass., MIT press, pp. 233-250.

Stillings, N., et al. (1995) *Cognitive Science. An Introduction*, Cambridge, Mass., MIT press.

Thagard, P. (1996) *Introduction to Cognitive Science*. Cambridge, Mass., MIT press.

Tononi, G. y G. Edelman. (1998) "Consciousness and Complexity", *Science* 282, 1846-1851.

Treisman, A. (1999) "Solutions to the Binding Problem: Progress through Controversy and Convergence", *Neuron* 24, 105-110.

Vallejos, G. (2008) *Conceptos y Ciencia Cognitiva*. Santiago, Bravo y Allende Editores.

Von Eckardt, B. (1993) *What Is Cognitive Science?*, Cambridge, Mass., MIT press.

Zeman, A. (2001) "Consciousness", *Brain* 124, 1263-1289.

_____. (2002) *Consciousness. A User's Guide*. New Haven y Londres, Yale University Press.