

La función de los experimentos mentales en las ciencias físicas.

Autor:

Mettini, Guadalupe

Tutor:

Cassini, Alejandro

2022

Tesis presentada con el fin de cumplimentar con los requisitos finales para la obtención del título Doctora de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires en Filosofía.

Posgrado



FILO:UBA
Facultad de Filosofía y Letras

**La función de los experimentos mentales
en las ciencias físicas**

TESIS PARA OPTAR POR EL TÍTULO DE DOCTORA

EN FILOSOFÍA

Doctoranda

Guadalupe Mettini

Director

Alejandro Cassini

2022

Agradecimientos:

Esta tesis es el fruto de un largo proceso de aprendizaje. Muchas fueron las personas que me acompañaron en este camino que concluye ahora con este trabajo pero que espero continúe y se diversifique. No menos importantes fueron las instituciones que proporcionaron los espacios y los recursos para formarme. Las palabras que me es posible expresar en este espacio son insuficientes para agradecer el afecto, la generosidad, la paciencia y la comprensión de todas y todos los que me han impulsado en estos años.

En primer lugar quiero agradecer a mi familia que es la raíz de todos mis crecimientos, por respaldar mis elecciones, especialmente por apoyarme en la singular decisión de estudiar filosofía y vivir de y para ella. A mis abuelos, mis padres y mis hermanos agradezco su ejemplo de perseverancia, trabajo y compromiso por el bien común.

A mis amigas agradezco su incondicional cariño, compañía y paciencia. Pero más que ninguna otra cosa les agradezco la lucha cotidiana que sostienen desde la literatura, la filosofía, el arte, la ciencia, el derecho y la política que hace al mundo un lugar más habitable para todas.

A mis amigos les agradezco por elegir siempre tratarme con cariño y respeto, por alentarme y estar dispuestos a la disputa fraterna.

A mis compañeras y compañeros de trabajo en la docencia les agradezco su optimismo y su invaluable tarea en la Universidad Pública. A los estudiantes con los que comparto aulas reales o virtuales les agradezco el privilegio de aprender mientras enseño.

A mi director, Alejandro Cassini le agradezco su generosidad intelectual, su compromiso con la rigurosidad filosófica y su meticulosa atención. Sepa que este trabajo es la consecución del anhelo de estudiar en la Universidad de Buenos Aires, una aspiración que su consejo y compromiso hicieron posible.

Finalmente quiero agradecer muy especialmente a mi amigo Bruno Borge, a quien conocí en el primer seminario de doctorado que cursé en la Facultad de Filosofía y Letras y quien desde entonces ha sido un maravilloso compañero de trabajo y un soporte insustituible en este proceso. Gracias por hacer de la academia un lugar de pertenencia.

ÍNDICE

INTRODUCCIÓN	6
PRIMERA PARTE. Naturaleza de los experimentos mentales	16
CAPÍTULO 1. La Filosofía de los experimentos mentales	17
1.1 Introducción	17
1.2 Los experimentos mentales en las ciencias físicas	19
1.3 El problema epistemológico de los experimentos mentales.....	29
1.4 Los experimentos mentales en la filosofía de la experimentación.....	39
1.5 El estado actual de la discusión	44
1.6 Definición provisoria y prospectiva	54
1.7 Conclusiones.....	61
CAPÍTULO 2. Experimentos mentales y experimentos reales. Analogías positivas y negativas.....	64
2.1 Introducción	64
2.2 Observación e intervención en el mundo natural.....	66
2.3 Fenómenos y datos	72
2.4 La vida propia de los experimentos	75
2.5 Visualizaciones y manipulación de escenarios imaginarios	77
2.6 Fenómenos y experimentos mentales.....	87
2.7 La autonomía de los experimentos mentales	91
2.8 ¿Son experimentos los experimentos mentales?	98
2.9 Conclusión.....	107
CAPÍTULO 3. Experimentos mentales y razonamiento científico	111
3.1 Introducción	111
3.2 Reconstrucción y esquemas de inferencia	114
3.3 Dispensabilidad y reductivismo.....	124
3.4 Inferencialismo	132
3.5 Razonamiento contrafáctico	143
3.6 Conclusiones.....	154
SEGUNDA PARTE. Los fundamentos del conocimiento en los experimentos mentales	157

CAPÍTULO 4. El origen del conocimiento en los experimentos mentales	158
4.1 Introducción	158
4.2 Perspectivas empiristas.....	162
4.3.1 Perspectivas Aprioristas I: Racionalismo.....	176
4.3.2 Perspectivas aprioristas II: Trascendentalismo	188
4.3 Perspectivas constructivistas	191
4.4 La razón y la experiencia como fuentes de conocimiento.....	198
4.6 Conclusiones.....	202
CAPÍTULO 5. Experimentos mentales, representación y modelos científicos... 205	205
5.1 Introducción	205
5.2 Sobre la noción de modelo científico	208
5.3 Los modelos mentales.....	223
5.4 Las simulaciones computacionales	234
5.5 Experimentos mentales y estrategias de idealización	241
5.6 Conclusiones.....	250
CAPÍTULO 6. La epistemología de los experimentos mentales. Condiciones de adecuación	253
6.1 Introducción	253
6.2 La descripción de los escenarios imaginarios: condiciones iniciales y marco teórico	254
6.3 Extractores de intuiciones	264
6.4 Modelado y representación	273
6.5 Interpretaciones teóricas	278
6.6 Los resultados experimentales	281
6.7 Conclusiones.....	289
TERCERA PARTE. Las funciones de los experimentos mentales	297
CAPÍTULO 7. Las funciones de los experimentos mentales	298
7.1 Introducción	298
7.2 Las tipificaciones existentes en la epistemología de los experimentos mentales	301
7.3 Problemas en torno a la clasificación de los experimentos mentales. Usos y funciones.....	316

7.4 Clasificación de los experimentos mentales	325
7.4.1 Usos heurísticos	328
7.4.2 Usos críticos de los experimentos mentales. Su rol en la elección de teorías.....	336
7.5 Conclusiones.....	348
CAPÍTULO 8. Los experimentos mentales en la formulación y las soluciones de la Paradoja de Olbers.....	349
8.1 Introducción	349
8.2 El surgimiento de la paradoja	351
8.3 La formulación inicial y las primeras soluciones a la paradoja.....	360
8.4 La paradoja de Olbers y la velocidad de la luz	367
8.5 Reformulación del viejo problema. La paradoja como evidencia de la expansión del universo	371
8.6 La caja cósmica y la cosmología relativista	373
8.7 Conclusiones.....	380
CONCLUSIONES GENERALES	384
BIBLIOGRAFÍA.....	404

INTRODUCCIÓN

"One often finds that thought experiments (or gedanken experiments) lie at the root of the really major advances in physics, the revolutions. Such thought experiments usually probe some foundational assumption, concerning space, time, locality, causality, determinism, matter, force, and so on." (Rickles, 2016: 3)

Los experimentos mentales constituyen una estrategia argumentativa ampliamente utilizada en la reflexión filosófica y en la investigación científica. La exploración de una situación imaginaria como punto de partida para establecer conclusiones normativas en ámbitos metafísicos, éticos o políticos ha sido un ejercicio recurrente en la filosofía clásica y moderna y una herramienta que junto al análisis conceptual caracterizan la metodología de la filosofía analítica contemporánea. Los pensadores presocráticos emplearon conjeturas de carácter contrafáctico y razonamientos analógicos para apoyar la plausibilidad de supuestos metafísicos. Así, por ejemplo, Heráclito (Frag. 42) evoca la imagen del caudal de un río para ilustrar la contingencia del mundo empírico, en el cual "todo fluye". En la filosofía platónica, la descripción de un escenario imaginario adquirió especial relevancia como estrategia metodológica en el marco de su teoría del conocimiento. El esclavo que, sin tener conocimientos de geometría, prueba con ayuda de Sócrates el teorema de Pitágoras en el *Menón* (82b-85c) es evidencia de que el conocimiento emerge por introspección. La filosofía aristotélica y su legado en la filosofía medieval son prolíficas en ejemplos de esta clase.

En la modernidad pueden identificarse algunos casos que cautivaron la imaginación de los filósofos posteriores. El ejemplo de la cera propuesto por Descartes en la segunda de sus *Meditaciones metafísicas* para explorar las

propiedades de la sustancia extensa y los casos propuestos por Locke en el libro XX, sec. xxvii de su *Ensayo sobre el entendimiento humano* para evaluar diversas tesis acerca de la identidad personal fueron ampliamente examinados y debatidos por la tradición filosófica. En la filosofía contemporánea, los experimentos mentales han sido empleados como estrategias argumentativas asociadas frecuentemente a la ilustración de teorías. Son ejemplos de este uso: el argumento de la tierra gemela formulado por Putnam (1975) para ilustrar su teoría externalista acerca del significado y el caso de la habitación china presentado por Searle (1980) contra el funcionalismo.

La manipulación de un fenómeno en una situación controlada, esto es, la experimentación en sentido habitual, ha sido considerada desde la Revolución Científica del siglo XVII como la principal fuente de conocimiento del mundo natural. No obstante, los experimentos mentales o imaginarios han ocupado un lugar significativo en los episodios de cambio conceptual, en la formulación de nuevos conceptos y en el desarrollo de nuevas teorías físicas. Reportaron, así, un enorme beneficio en ocasiones en las que la experimentación ordinaria estuvo limitada por razones técnicas, y fueron sumamente fértiles para la exploración de hipótesis teóricas en campos en los que la experimentación real no era posible. Galileo Galilei (1638) elaboró algunos de los ejemplos más representativos de esta clase de experimentos, entre ellos, el famoso experimento de los cuerpos en caída libre, dirigido a mostrar las inconsistencias que se siguen del empleo del concepto aristotélico de velocidad natural. Aunque existe una diversidad de ejemplares que pueden incluirse en esta categoría, todos los casos presentan algunos rasgos comunes. Consisten en la descripción de un escenario y/o de un aparato experimental y en la especificación de una serie de instrucciones sobre cómo introducir variaciones en la situación representada. La narrativa incluye, además, una interpretación de los resultados de estas acciones en el marco de algunos principios teóricos.

Desde el punto de vista filosófico los experimentos mentales presentan una propiedad asombrosa: su capacidad de proporcionar, al menos en principio, nuevo conocimiento del mundo natural prescindiendo de la introducción de datos empíricos nuevos. No obstante su importancia para la historia de la ciencia, esta clase de experimentos no constituyó un objeto de estudio genuino para la filosofía de la ciencia hasta finales del siglo pasado. Si bien una parte significativa de la reflexión sobre el conocimiento científico se centró, desde la década de 1980, en la actividad experimental, solo muy recientemente se emprendió un estudio sistemático de los experimentos mentales. Puede pensarse que las razones para dicha omisión no son muy diferentes de aquellas por las que la filosofía clásica de la ciencia postergó el análisis de los experimentos ordinarios, a saber, la vigencia de una imagen de la ciencia de acuerdo con la cual la experimentación es siempre precedida por la teoría y solo adquiere sentido a partir de ésta. Dicha imagen fue desafiada por filósofos como Hacking (1983), quienes argumentaron a favor del carácter no subordinado de las prácticas experimentales fundándose en la multiplicidad de funciones que desempeñan y en la objetividad de las entidades o fenómenos que son capaces de producir.

Esta línea de investigación, en su momento llamada *Nuevo Experimentalismo*, centró la atención en las prácticas experimentales, a las que adjudicó amplia autonomía respecto de las teorías. Desde esta perspectiva, aunque los científicos a veces elaboran y discuten experimentos mentales, estos son concebidos como tipos particulares de argumentos teóricos. Aquello que distingue los experimentos mentales de los experimentos reales consiste en que los primeros carecen de un elemento esencial a las prácticas experimentales, a saber, la intervención en el mundo material. En general, los filósofos convienen en que para llevar a cabo un experimento, los científicos tienen que intervenir activamente en el mundo material y que, al hacerlo, producen nuevos objetos, sustancias, fenómenos y procesos (Radder, 2003: 4). En este sentido, el experimento se diferencia de la teoría incluso si el trabajo teórico es siempre

acompañado de actos materiales como escribir una fórmula. Asimismo, prevalece la opinión de que no cualquier tipo de intervención en el mundo cuenta como un experimento. En general, los experimentos requieren por lo menos de cierta estabilidad y reproducibilidad de sus resultados. Satisfacer estos requisitos presupone un control tanto del sistema experimental como de su medio ambiente, y, además, cierta disciplina por parte de los experimentadores. Asimismo, esta perspectiva asume que los experimentos ordinarios tienen ciertas características que permiten justificar su autonomía respecto de las teorías. En palabras de Hacking (1992: 307), los experimentos tienen “vida propia” porque pueden madurar, evolucionar, adaptarse o reestructurarse. Son capaces de involucrar adelantos instrumentales para producir mejores resultados, desarrollar las habilidades de los experimentadores o ser llevados a cabo a la luz de una mejor comprensión de las teorías. Pero los experimentos mentales permanecen fijos a un marco conceptual. Esto significa que están limitados a un contexto y no pueden ser empleados en otros. Al depender de una serie de supuestos básicos tomados de un marco nomológico, los poderes epistémicos de los experimentos mentales se limitan, desde este punto de vista, a su capacidad de exhibir inconsistencias o contradicciones dentro de un marco conceptual. En síntesis, la dependencia teórica y la ausencia de intervención fueron las principales razones para excluir a los experimentos mentales del tratamiento sistemático de las prácticas experimentales.

Como contraparte a la interpretación de Hacking y de los filósofos pertenecientes a la corriente experimentalista, se puede señalar que los experimentos mentales científicos se asemejan a los experimentos reales al menos en tres aspectos importantes. En primer lugar, poseen una fuerza demostrativa que, al menos en principio, aventaja a la de los argumentos teóricos. En segundo lugar, en ellos el experimentador (o el lector del experimento) desempeña un papel activo, en algunos aspectos similar al que desempeña en las narrativas de los experimentos reales. El lector debe asumir la perspectiva de la

primera persona e imaginarse a sí mismo introduciendo modificaciones en el escenario imaginario. En tercer lugar, es posible argumentar que en los experimentos mentales, del mismo modo que en los reales, se puede distinguir el proceso experimental (equivalente a la *realización material*, al menos en relación con las fases de preparación e interacción) de su interpretación teórica.

Si bien las razones mencionadas pueden considerarse como argumentos a favor de no reducir los experimentos mentales a puros argumentos, estas no aportan elementos para esclarecer el núcleo del problema epistemológico en el que radica el interés filosófico en estas prácticas. El mismo consiste en determinar si a partir de un ejercicio especulativo es posible obtener nuevo conocimiento del mundo físico real. Se puede afirmar provisionalmente que la manipulación mental de escenarios imaginarios hace posible derivar ciertas consecuencias acerca del comportamiento de fenómenos físicos, de manera análoga a como la manipulación de un modelo a escala permite realizar inferencias sobre el sistema que representa. Asimismo, es un hecho que algunos experimentos mentales han producido transformaciones en los marcos conceptuales. Algunos experimentos mentales -como el de los cuerpos en caída libre de Galileo Galilei ([1638] 1974) o el del rayo de luz de Einstein (1949)- han permitido identificar inconsistencias en una teoría y han conducido al abandono de una hipótesis y a su reemplazo por otra. Pero esto no equivale a afirmar que los experimentos mentales pueden aportar el mismo tipo de evidencia que la que los experimentos reales pueden proporcionar respecto de una hipótesis científica (en particular, evidencia confirmatoria). La ausencia de intervención en el mundo natural priva a los experimentos mentales de la base objetiva de conocimiento en la que se apoya el conocimiento que los experimentos reales son capaces de producir. Esta diferencia con la experimentación real plantea preguntas sobre los criterios para identificar casos espurios en los experimentos mentales y pone de manifiesto la necesidad de dar una explicación satisfactoria de cómo el ejercicio imaginativo permite derivar conclusiones que se aplican a fenómenos físicos.

Otra cuestión vinculada con el problema epistemológico de los experimentos mentales que requiere de un examen más detallado es la de su poder heurístico. Si estos experimentos funcionan en base a representaciones imaginativas, una explicación de su funcionamiento debe establecer por qué la representación de escenarios imaginarios permite derivar conclusiones sobre el mundo natural de manera más fácil y rápida que mediante la realización de una inferencia. Considerando, además, que las idealizaciones y la postulación de situaciones contrafácticas son una característica fundamental de la narrativa de estos experimentos, resulta conveniente, para dar una explicación de su funcionamiento, aclarar la naturaleza, funciones y usos legítimos de las idealizaciones y de las hipótesis contrafácticas dentro de los experimentos mentales.

El examen sistemático de la naturaleza y el funcionamiento de los experimentos mentales en ciencias fácticas es relativamente reciente en la filosofía de la ciencia. Por otra parte, el análisis restringido al uso de esta clase de experimentos en el contexto de justificación de las teorías físicas es apenas incipiente. Las perspectivas que se han desarrollado en relación con esta temática se encuentran diseminadas en artículos y apéndices de textos canónicos pero los estudios monográficos enteramente dedicados a esta cuestión son escasos. La polémica acerca del papel epistémico de los experimentos mentales puede ser comprendida en función de dos grandes perspectivas en pugna. Existen, por una parte, interpretaciones empiristas acerca del conocimiento involucrado en los experimentos mentales y, por otra parte, hay posiciones aprioristas acerca del origen de este conocimiento. Mach (1896), pionero en el examen de la función de los experimentos imaginarios, abogó por una interpretación empirista de acuerdo con la cual la ejecución de un experimento mental depende de formas de conocimiento instintivo. Por su parte, Koyré (1968) propuso, a partir de un estudio histórico de los experimentos de Galileo, la primera interpretación apriorista de las funciones de los experimentos mentales. En esta línea interpretativa, Brown

(2001) desarrolló una interpretación completamente platonista del conocimiento proporcionado por los experimentos mentales, según la cual estos nos permiten captar intuitivamente propiedades que constituyen clases naturales. En contra de esta perspectiva, la propuesta empirista de Norton (1991) sostiene que los experimentos mentales son reducibles a argumentos que recuperan información empírica y la reorganizan. Otras posiciones no pueden situarse dentro de esta dicotomía. Kuhn (1964), en el marco de su teoría general acerca del desarrollo de las disciplinas científicas, identificó los experimentos mentales con herramientas que permiten detectar anomalías y cumplen una importante función en los procesos de reconceptualización en la historia de la ciencia. Popper (1959), por su parte, solo les reconoció un papel heurístico y crítico en el marco de la argumentación científica.

Los primeros tratamientos sistemáticos acerca del papel epistémico de los experimentos mentales fueron las obras monográficas de Brown (1991) y Sorensen (1992). Aunque en los últimos años el volumen de contribuciones sobre esta temática ha aumentado considerablemente, no hay casi ningún consenso sobre la definición de la noción de experimento mental ni tampoco hay acuerdos sobre su taxonomía y funciones legítimas. Aunque existe una extendida coincidencia en la idea de que los experimentos mentales involucran formas de razonamiento hipotético que tienen como objetivo principal examinar principios teóricos, no hay consenso acerca de si es conveniente abordar su estudio filosófico considerando estas prácticas como experimentos o como argumentos. Tampoco se aprecia, en la gran mayoría de las contribuciones, un esfuerzo por enmarcar la explicación del funcionamiento de los experimentos mentales en la filosofía analítica de la experimentación. Debido a que la mayoría de las perspectivas emplean la analogía con los experimentos reales, es razonable suponer que la interpretación filosófica de los experimentos reales puede resultar esclarecedora en la investigación de la naturaleza y las funciones de los experimentos mentales. Por otra parte, las posiciones teóricas desarrolladas

hasta el momento examinan casos que pertenecen exclusivamente al dominio de la física, como la mecánica (newtoniana, relativista o cuántica), donde la experimentación real no solo es posible, sino que ha prevalecido sobre la experimentación mental. Otras cuestiones aparecen en dominios en los cuales la experimentación real no es posible o resulta sumamente difícil. En síntesis, aunque en los últimos años los experimentos mentales científicos se convirtieron en un tema de interés de la filosofía de la ciencia, los estudios dedicados a esta cuestión carecen en general de sistematicidad y se han abocado al estudio de un conjunto de casos canónicos procedentes en su mayoría de la mecánica. En el estado actual de la cuestión no hay consensos sostenidos sobre la manera más adecuada de abordar el estudio de los experimentos mentales y por consiguiente tampoco hay acuerdos sobre su definición y clasificaciones.

El objetivo general de este trabajo es ofrecer un análisis sistemático de la naturaleza y funciones de los experimentos mentales en las ciencias físicas. La hipótesis general que articula esta indagación es que si los experimentos mentales proporcionan conocimiento acerca del mundo natural, dicho conocimiento debe tener origen en la experiencia y debe ser el producto de algún mecanismo de generación de creencias especificable y compartido por otras herramientas del razonamiento científico. De manera más específica, el propósito de esta investigación es examinar el problema del valor evidencial de los experimentos mentales, determinar su relación con otras formas de experimentación, analizar sus alcances y empleos legítimos. Se propone que interpretar los experimentos mentales como una clase de modelos científicos permite esclarecer el mecanismo de formación de nuevas creencias que estos experimentos emplean y hace posible, además, establecer condiciones generales de adecuación e interpretar sus diferentes usos en los contextos de descubrimiento y justificación. Para explorar el alcance de estas hipótesis de trabajo se examinan varios casos tomados de la historia de la ciencia. Para mostrar la plausibilidad de las conclusiones alcanzadas se analiza el papel

concreto de los experimentos mentales en la formulación y las soluciones propuestas de la Paradoja de Olbers. El caso de estudio pertenece a un dominio de fenómenos en los que no es posible la experimentación real por lo que, su examen permite revelar algunos puntos ciegos en el examen del funcionamiento de esta clase de experimentos. La tesis se estructura en tres partes en las cuales se examinan sucesivamente a) la naturaleza de los experimentos mentales, b) los fundamentos del conocimiento que pueden proporcionar y c) las funciones que pueden desempeñar legítimamente en el contexto de justificación de las teorías físicas.

En la primera parte de la tesis se presenta el problema epistemológico de los experimentos mentales, se examinan las principales posiciones acerca de su papel epistémico y se evalúan las dos principales líneas interpretativas vigentes en la actualidad, a saber: la experimental y la argumental. En general la posición argumental se asocia a perspectivas empiristas en lo que refiere al fundamento del conocimiento. Dentro de las interpretaciones experimentales en cambio se sitúan posiciones aprioristas, empiristas y constructivistas respecto del origen del conocimiento. El primer capítulo ofrece una delimitación del objeto de estudio. Se presenta allí una caracterización de los experimentos mentales en las ciencias físicas a partir de la presentación de algunos casos representativos, una introducción al debate filosófico desde las primeras teorizaciones hasta el estado actual de la discusión y una definición prospectiva de esta clase de experimentos. El segundo capítulo se propone establecer los alcances de la interpretación de acuerdo con la cual los experimentos mentales son propiamente experimentos. Se intentan establecer los límites de la analogía con los experimentos reales a partir de una aproximación a la caracterización de las nociones de observación y experimentación elaboradas por la filosofía de la experimentación. El objetivo del tercer capítulo es explicitar y evaluar un conjunto de tesis sobre la naturaleza de los experimentos mentales con las que las posiciones argumentales se comprometen. El examen de estas posiciones se centra en la función que

atribuyen a la representación de un escenario y a la ejecución en la imaginación de ciertas acciones que conducen a la generación de nuevas creencias.

La segunda parte de la tesis se aboca a la revisión del problema del fundamento del conocimiento y a los criterios de corrección en el uso de los experimentos mentales. El cuarto capítulo examina las principales posiciones actuales acerca de la fuente del nuevo conocimiento que los experimentos mentales pueden proporcionar y las facultades involucradas en la obtención de dicho conocimiento. Se evalúan las propuestas disponibles en virtud de sus respectivas ventajas explicativas. En el capítulo quinto se argumenta que interpretar los experimentos mentales como modelos permite dar sentido a las características que estos comparten con los experimentos de laboratorio, las simulaciones computacionales y los argumentos científicos. Se muestran en el capítulo sexto cómo las semejanzas entre experimentos mentales, experimentos reales y simulaciones computacionales hacen posible elaborar exitosamente una serie de condiciones de adecuación que permiten distinguir entre usos exitosos y usos espurios de los experimentos mentales.

En la tercera parte se aplican algunas de las conclusiones parciales de las dos partes anteriores al estudio de casos y se propone una taxonomía de los experimentos mentales. En el capítulo séptimo se ensaya una tipificación de los experimentos mentales y en el octavo se procede a mostrar, a partir del estudio de los planteos y las diferentes soluciones propuestas de la paradoja de Olbers, que la interpretación de los experimentos mentales como modelos permite elucidar las funciones de estos dispositivos en los contextos de descubrimiento y justificación de las teorías científicas.

PRIMERA PARTE

Naturaleza de los experimentos mentales

CAPÍTULO 1

La filosofía de los experimentos mentales

1.1 Introducción

El empleo de escenarios imaginarios ha formado parte importante de la argumentación científica. En los orígenes de la física moderna abundan ejemplos en los que una situación contrafáctica se postula para plantear dudas acerca de la aplicabilidad de un principio teórico, establecer inconsistencias en una teoría o incluso dar sustento a una nueva hipótesis. El escolástico francés, Jean Buridan (1509), por ejemplo, examinó un conjunto de casos hipotéticos acerca del desplazamiento de los proyectiles para mostrar que los efectos del movimiento violento no pueden ser explicados por la noción aristotélica de *antiperístasis*. En sus *Diálogos*, Galileo (1638) elaboró algunos de los ejemplos más representativos de esta clase de experimentos, entre ellos, el famoso caso de los cuerpos en caída libre dirigido a mostrar las inconsistencias que se siguen del empleo del concepto aristotélico de movimiento natural. Algunos de estos experimentos ocuparon un lugar de importancia en el desarrollo de nociones fundamentales para la ciencia moderna. El experimento del balde giratorio empleado en el marco de la controversia entre Newton (1687) y Leibniz acerca de la naturaleza del espacio, fue posteriormente discutido y reinterpretado por Mach (1919) y Einstein (1916). Más recientemente, experimentos mentales vinculados a la Teoría de la Relatividad y la Mecánica Cuántica han estimulado intensas discusiones acerca de la plausibilidad de sus principios fundamentales. Schrödinger (1935) dirigió el experimento del gato en la caja contra la interpretación de Copenhague de la mecánica cuántica, en particular contra el principio de superposición de estados. En termodinámica, uno de los casos más conocidos es el famoso experimento del

demonio que Maxwell (1871) empleó para mostrar el carácter estadístico del principio formulado por primera vez por Carnot. Asimismo estos experimentos han cumplido importantes funciones en la exploración de hipótesis astrofísicas y cosmológicas. Este es el caso de los experimentos utilizados para formular y resolver la llamada Paradoja de Olbers.

De manera general y provisoria, podría definirse a un experimento mental como *un relato que describe una situación que el lector debe representarse en la imaginación y en la que debe introducir variaciones con la finalidad de evaluar una hipótesis o principio teórico*. No obstante, resulta difícil ofrecer una caracterización que permita incluir adecuadamente la heterogeneidad de los ejemplares que generalmente se reúnen bajo esta denominación. Aunque su importancia en el razonamiento científico ha sido reconocida ampliamente en las teorizaciones filosóficas, los intentos de proporcionar una descripción de su naturaleza que sea consistente con el desarrollo de una tipología y una elucidación de las funciones que pueden desempeñar han sido insuficientes. Las caracterizaciones existentes resultan poco fértiles debido a que son demasiado estrechas, o bien demasiado amplias o ambiguas en su formulación. El propósito de este capítulo es suministrar una caracterización del objeto de estudio de este trabajo que ofrezca un criterio suficiente para identificar los experimentos mentales, que sea lo bastante flexible como para incluir una amplia variedad de casos y que allane el camino hacia su estudio sistemático. Con este fin, en primer lugar se delimita el objeto de estudio de la investigación a partir de la descripción de una serie de casos de experimentos mentales considerados paradigmáticos. La selección de casos se justifica en virtud de un rasgo central que estos exhiben de manera conspicua y que resulta relevante para su análisis filosófico: se proponen proveer conocimiento nuevo del mundo natural en ausencia de la ejecución efectiva del experimento. Esta característica es determinante porque su presencia en una multiplicidad de casos históricamente notorios motorizó reflexiones que tendieron a agruparse en un campo más o menos unificado. Las primeras

conceptualizaciones de los experimentos mentales coinciden con el origen de la filosofía de la ciencia como disciplina independiente, hacia fines del siglo XIX. Estas contribuciones iniciales acerca de la naturaleza y funcionamiento de los experimentos mentales tendieron a expresar, de modo no articulado, la paradoja central que motiva todavía hoy el tratamiento metateórico de estas prácticas: ¿cómo es posible obtener conocimiento nuevo del mundo natural solo razonando acerca de una situación imaginada? Estas reflexiones tempranas abordaron el problema de los experimentos mentales aunque no lograron conformar un campo autónomo de debate filosófico. En el segundo apartado se reconstruyen y analizan brevemente estas teorizaciones, se enuncia el problema epistemológico al que se abocaron y se trazan las perspectivas que se fueron delineando a partir de estos tratamientos inaugurales. A continuación, se examina brevemente el tratamiento de esta temática en el contexto de la recientemente desarrollada filosofía de la experimentación. Luego, se presenta el estado de la discusión actual en la filosofía de los experimentos mentales. Se revisan allí las definiciones y caracterizaciones que se han realizado hasta el momento. Asimismo, se brinda un panorama de algunas cuestiones abiertas. Finalmente, se argumenta que las definiciones y tipologías existentes de los experimentos mentales resultan insuficientes y se ensaya una definición con el fin de tratar los problemas filosóficos a los cuales está abocado este trabajo.

1.2 Los experimentos mentales en las ciencias físicas

La manipulación de un fenómeno en una situación controlada ha sido considerada desde la Revolución Científica como la principal fuente de conocimiento del mundo natural, no obstante, la discusión de principios teóricos a partir de la elaboración de situaciones hipotéticas ha ocupado un lugar significativo en los episodios de cambio conceptual, en la formulación de

conceptos y en el desarrollo de nuevas teorías físicas. Es posible argumentar que la ubicuidad de los experimentos ordinarios es una razón para considerar que los experimentos mentales han sido menos frecuentes y que por lo tanto no pueden ser valorados como prácticas habituales en el razonamiento científico, al menos no a la par de los experimentos concretos. Reiss (2002: 4) sostiene que el número de experimentos mentales que es posible identificar es mucho menor que el de los experimentos reales y que, por tanto, su significación para la historia de la ciencia es despreciable. No se parte aquí de la suposición de que los experimentos mentales son relevantes por la frecuencia en la que han sido usados sino por las implicaciones que ha tenido principalmente en la formulación de nuevos conceptos y en la exploración de los límites de los principios teóricos. Han resultado de utilidad en ocasiones en las cuales la experimentación ordinaria estuvo limitada por motivos técnicos y han sido sumamente fructíferos para exploración de hipótesis teóricas. Por otra parte, los experimentos mentales resultan filosóficamente interesantes por derecho propio. Algunos de estos casos han obrado lo que algunos denominan “magia epistémica” (Norton, 2004: 44), esto es, han logrado trascender lo que la mera suposición hace posible proporcionando nuevo conocimiento acerca de fenómenos naturales sin la introducción de datos empíricos nuevos. Este rasgo ha producido la perplejidad de muchos filósofos. Para establecer si el asombro de los filósofos es producto de un espejismo, de una argucia heurística incapaz de aventajar a un argumento o si es motivado por la existencia de un mecanismo prodigioso, aunque infrecuente que los científicos tienen en su poder para investigar la naturaleza, es preciso esclarecer el funcionamiento de esta clase de experimentos. La selección de casos que se presenta a continuación se fundamenta en el carácter representativo de los mismos: exhiben de manera ejemplar los dos principales aspectos que estimularon el análisis filosófico de estas prácticas, a saber: la falta de intervención en el mundo natural y la “instantaneidad” con la que los resultados son obtenidos. Estas características motivaron la indagación sobre sus poderes epistémicos. Asimismo, estos ejemplos hacen patente el uso de dos estrategias

características de los experimentos mentales: los supuestos contrafácticos y las idealizaciones. La elección de los casos está motivada también por la importancia que estos experimentos mentales en particular han tenido para el desarrollo de algunas nociones fundamentales para la física moderna. Adicionalmente los ejemplos a continuación presentados fueron objeto de análisis en algunos de los trabajos más importantes acerca del estatus epistémico de los experimentos mentales (por ejemplo, Mach, 1896-1897; Koyré, 1960; Brown, 1991; Norton, 1996; Sorensen, 1991) y continúan recibiendo mucha atención por parte de científicos y filósofos (por ejemplo, Thagard, 2014; Peacock, 2018; Palmieri, 2018 y Arthur, 2018).

Sin duda el caso más célebre entre los experimentos mentales en física es el empleado por Galileo en su *Discorsi* (1638) con el objetivo de exhibir contradicciones en la física aristotélica. El caso trata acerca del comportamiento de los cuerpos en caída libre y anticipa el principio de equivalencia entre masa inercial y gravitatoria, de acuerdo con el cual todos los cuerpos caen con la misma aceleración independientemente de su masa y composición. La tesis aristotélica en cuestión afirma que el movimiento en el vacío es imposible dado que en un medio de estas características la velocidad de los cuerpos en caída libre sería infinita. Esta hipótesis descansa en dos supuestos: el primero establece que dos cuerpos de distinto peso se mueven en el mismo medio con velocidades proporcionales a sus pesos y el segundo, que un cuerpo en diferentes medios se mueve en cada uno de ellos a una velocidad inversamente proporcional a la resistencia del medio. (Galileo, [1638] 1974: 61). El experimento plantea siguiente cuestión: si se lanzaran unidas una piedra grande que se mueve a una velocidad determinada, por ejemplo, de ocho grados y una pequeña que se mueve a una velocidad de cuatro grados, ¿cómo debería interpretarse, de acuerdo con la doctrina aristotélica, el movimiento del sistema? (Figura 1). Por un lado, podría suponerse que la velocidad de caída del complejo debería ser menor a los ocho grados, debido a que la piedra más pequeña cae a una velocidad menor y

desacelera a la piedra más grande. Por otro lado, dado que las dos piedras unidas dan como resultado una piedra más pesada, debería asumirse que el sistema se movería con una velocidad mayor que la de la piedra más grande tomada aisladamente (como se muestra en la Figura 1). De esta contradicción se sigue la falsedad de la primera hipótesis aristotélica. Para mostrar que el segundo de los principios aristotélicos es falso el argumento incluye otros ejemplos, con algunos de ellos muestra que el mismo cuerpo no adquiere, cayendo en medios de distinta densidad (como el agua y el aire), velocidades inversamente proporcionales a la resistencia de los medios que atraviesa. En un medio material, aunque sea sutil como el aire, la resistencia aumenta con la aceleración y se retardan por ello los grados de velocidad adquiridas. Galileo supuso que en un medio completamente vacío de resistencia todos los cuerpos caerían con la misma velocidad (Galileo, [1638] 1974: 172).

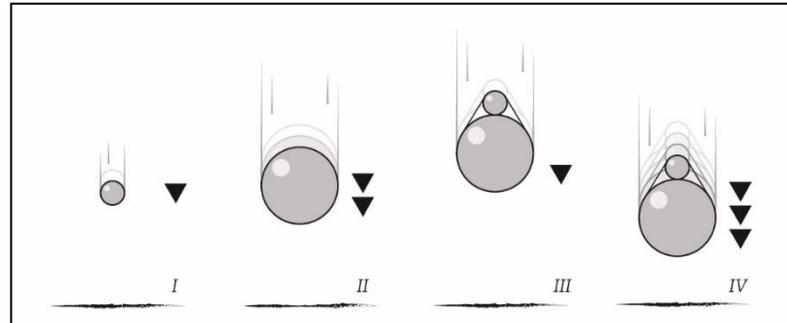


Figura 1.1

Los experimentos empleados por Galileo para establecer su principio de equivalencia¹ no incluyen la realización material del sistema experimental. Las

¹ Se denomina aquí *principio de equivalencia* a la afirmación de que en un medio vacío de resistencia cuerpos de diferente masa caen con la misma aceleración. Se adopta esta designación por razones de economía, aunque su denominación como su formulación en términos de masa y aceleración son contemporáneas y no corresponden con los términos en los que el principio fuera formulado por Galileo.

conclusiones acerca de las propiedades del movimiento de los cuerpos en caída libre, son el resultado de manipulación mental del escenario imaginado. El experimento mental del sistema de pesos combinados impugna la hipótesis aristotélica, mostrando las contradicciones que se derivan de aplicar dicha conjetura para explicar el caso. Esta conclusión adicionada a la suposición acerca de cómo se comportarían los cuerpos en un medio sin resistencia, permite derivar el principio de equivalencia. El nuevo conocimiento obtenido como resultado de estas operaciones no depende de la introducción de nueva información empírica. Esta característica ha constituido una de las principales motivaciones del estudio metacientífico de los experimentos mentales.

El experimento mental elaborado por Stevin para indagar las propiedades del plano inclinado es otro ejemplo en el que puede verse claramente este rasgo. El caso apela a cierto conocimiento intuitivo que el experimentador posee acerca de cómo se comportaría un peso depositado sobre un plano: en un plano horizontal el peso permanecerá en reposo y en un plano vertical se deslizará libremente. El experimento pone en cuestión lo que ocurre en los dos casos intermedios (Figura 2). En él se describe un soporte con forma de prisma, constituido por dos planos inclinados, que se suponen sin fricción, de diferentes alturas, sobre los cuales descansa una cadena o una serie de pesos unidos en forma de cadena cerrada. En este escenario la pregunta relevante es si la cadena se mantendrá o no en equilibrio estático. Ante esta cuestión existen tres posibilidades: i) los pesos se moverán, hacia un lado quizás porque la pendiente es más pronunciada, ii) se moverán hacia el otro lado, quizás porque hay más masa en esa porción de la cadena, o iii) permanecerán en equilibrio estático. Stevin concluye que la cadena permanecerá en reposo (Figura 3). Si a causa de la diferencia de pesos la cadena se desplazara hacia alguno de los lados, seguiría moviéndose de manera continua, generándose así una máquina de movimiento perpetuo. La suposición de que esto es imposible es esencial para el experimento. La conclusión de Stevin es que dos pesos cualesquiera en un plano inclinado

unidos entre sí, se equilibran siempre que sus respectivos pesos sean proporcionales a las longitudes de los planos. (Mach, [1893] 1960: 25).

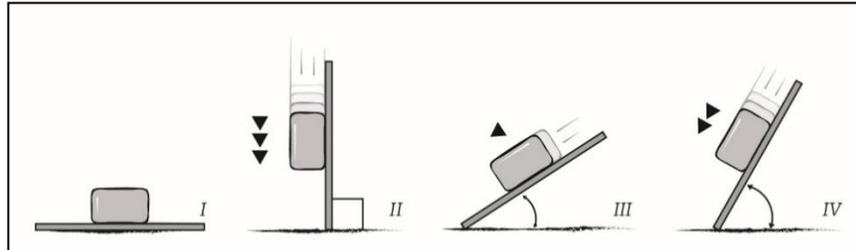


Figura 1.2

El experimento mental de Stevin exhibe otra particularidad que ha motivado el análisis filosófico. La evidencia que proporciona a favor del principio enunciado hace que su aceptación resulte casi forzosa. Asimismo, la visualización del escenario en cuestión hace superflua la ejecución material del experimento. La imposibilidad de dar sentido a las demás alternativas acerca de cómo se comportaría el sistema de pesos unidos, compele a aceptar por *default* la idea de que permanece en equilibrio estático.

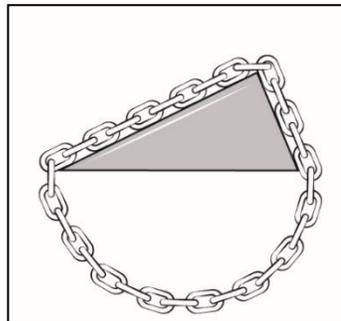


Figura 1.3

Aunque no todos los experimentos mentales de importancia en la historia de la física se asemejan a los casos de Galileo y Stevin en cuanto a su funcionamiento y al tipo de apoyo evidencial que proporcionan a una hipótesis,

todos ellos fueron elaborados con el fin de derivar conclusiones acerca del mundo natural a partir de la mera exploración de una situación imaginaria.

El experimento del balde en rotación, fue empleado por Newton como evidencia acerca de la existencia del espacio absoluto. Dado que el espacio en sí mismo es inobservable, la construcción del experimento mental resultó fundamental para su argumentación. Para Newton, las personas comunes conciben las cualidades temporales y espaciales bajo las relaciones que se dan entre los objetos sensibles (Newton, [1687] 1974: 6). Dado que las partes del espacio no pueden ser diferenciadas unas de otras por los sentidos, en su lugar se emplean mediciones sensibles para distinguirlas. De la posición y la distancia de las cosas consideradas respecto de cualquier cuerpo que se tome como inmóvil se definen todos los lugares (Newton, [1687] 1974: 8). Relativamente a tales lugares se estiman los movimientos considerados como la traslación de los cuerpos de un lugar hacia otro. Por lo que, en vez de lugares y movimientos absolutos para la medición del espacio se emplean los relativos. El espacio relativo es alguna relación móvil o medida del espacio absoluto, que los sentidos determinan por la posición de los cuerpos y que es tomada erróneamente por el espacio absoluto. Newton estaba convencido de que el espacio absoluto podría ser conocido a través de la existencia de fuerzas centrífugas en movimiento rotacional (Newton, [1687] 1974: 10). Si el espacio es una realidad física, como Newton creía, y si el movimiento acelerado constituye un criterio para su identificación, el espacio posee una estructura dual: absoluto para el movimiento acelerado y relativo para el movimiento rectilíneo uniforme. Para ilustrar esta distinción, propuso el siguiente experimento mental: imaginemos que en el universo sólo existe un balde parcialmente lleno de agua. El balde es suspendido de una cuerda (no importa de qué esté sostenida) que es retorcida y luego liberada. A medida que la cuerda se desenrolla observamos distintos estados en el sistema formado por el balde y el agua. En el primer estado, no hay movimiento relativo entre el agua y el balde. La superficie del agua es plana. En el segundo

momento, inmediatamente después de que la cuerda es liberada y el balde comienza a girar, el agua del balde está en un movimiento relativo, es decir se mueve respecto de las paredes del balde y la superficie del agua permanece plana. En el tercer estado el agua y el balde están en reposo uno respecto del otro y la superficie del agua se vuelve cóncava (Figura 4). Newton explica la diferencia entre el primer y último estado de la siguiente manera: en el primer momento el agua y el balde están ambos en reposo absoluto respecto del espacio absoluto. En el tercer estado el agua y el balde están en movimiento absoluto, respecto del espacio absoluto. La diferencia entre movimiento relativo y absoluto explica la forma del agua dentro del balde. (Newton, [1687] 1974: 11-12). Este ejemplo fue discutido posteriormente por Mach, ([1893] 1960: 227-233). De acuerdo con su interpretación, la forma cóncava del agua no revela la aceleración respecto del espacio absoluto, sino la aceleración relativa a otras masas en el universo. En un universo efectivamente vacío, el fenómeno que Newton describe no ocurriría: el agua dentro del balde permanecería plana.

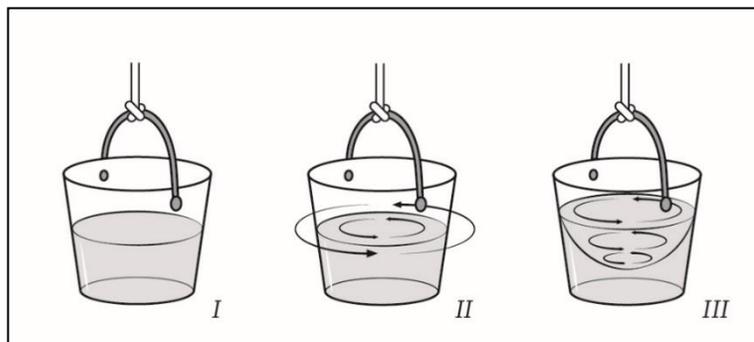


Figura 1.4

La ausencia de ejecución material ha resultado una ventaja para explorar sistemas que resultan inaccesibles a la experimentación real. Algunos experimentos mentales cumplieron importantes funciones en la investigación de hipótesis astrofísicas y cosmológicas. Este es el caso de los experimentos mentales que fueron empleados para formular y resolver la llamada Paradoja de Olbers. La misma plantea la contradicción entre un universo que se supone

espacialmente infinito y poblado por infinitas estrellas en una distribución homogénea y la oscuridad del cielo nocturno. Los diferentes intentos de solución proponen modelos que explican en virtud de la distribución de las estrellas, la distancia respecto del observador o la energía necesaria para que el cielo sea completamente brillante, el fenómeno de la luz faltante.

Otro rasgo sugestivo de los experimentos mentales es su carácter *contrafáctico*. Algunos experimentos que cumplieron un importante papel en el desarrollo de la física reciente se basan en suposiciones contrafácticas o en extrapolaciones que permiten evaluar hipótesis o principios físicos. Así por ejemplo, con el fin de ilustrar el carácter estadístico de la segunda ley de la termodinámica, Maxwell (1871: 308-309) propuso un experimento mental en el que conjetura un caso en el que ésta no se cumpliría. En términos cualitativos, la ley en cuestión establece que el calor fluye naturalmente de los cuerpos calientes a los cuerpos fríos. De acuerdo con Maxwell, es posible imaginar dos compartimientos separados, uno lleno con aire caliente y otro lleno con aire frío, unidos por una pequeña puerta controlada por un pequeño ser inteligente. La velocidad promedio de las moléculas en el compartimiento con aire frío es menor que la velocidad promedio de las moléculas en el compartimiento con aire caliente. El pequeño “demonio” deja pasar las moléculas más rápidas de aire frío dentro del compartimiento de aire caliente y las moléculas más lentas del compartimiento de aire caliente dentro del compartimiento lleno de aire frío. Como consecuencia, aumenta la energía cinética promedio de las moléculas del gas caliente y disminuye la energía cinética promedio de las moléculas del gas más frío, volviéndose más caliente el primero y más frío el segundo, lo que constituiría una violación de la segunda ley de la termodinámica. El experimento muestra un aspecto altamente contraintuitivo de la teoría haciéndolo más evidente (Maxwell, 1871: 308-309).

Con fines similares, Schrödinger (1935: 157) dirige un experimento mental contra la interpretación de Copenhague de la mecánica cuántica, en particular,

contra el principio de superposición de estados. En el experimento se realiza una extrapolación de las consecuencias este principio al mundo macroscópico. Según Schrödinger, es posible conjeturar las consecuencias bizarras de este principio si imaginamos a un gato encerrado en una caja con un dispositivo (asegurado de cualquier interferencia directa por parte del gato) que contiene una pequeña cantidad de una sustancia radioactiva. Si uno de los átomos de la sustancia se desintegra, se activa un mecanismo que libera un martillo que rompe un contenedor de ácido cianhídrico. Suponiendo que hay una probabilidad igual a 0.5 de que un átomo de la sustancia radioactiva se desintegre en el curso de una hora, transcurrido ese tiempo es posible suponer que la desintegración no ha ocurrido, o bien que ha ocurrido. Si no se observa dentro de la caja, el gato se encuentra en una superposición de estados de vida y muerte con la misma probabilidad. De manera que el escenario lleva a admitir que el gato se encuentra en una superposición de estados, es decir, está vivo y muerto al mismo tiempo.

Es posible afirmar provisionalmente que la manipulación mental de escenarios imaginados les ha permitido a los científicos derivar consecuencias acerca del comportamiento de fenómenos físicos. Establecer las fuentes de este conocimiento constituye uno de los puntos centrales de la controversia epistemológica acerca de los experimentos mentales. Identificar inconsistencias en las teorías, y derivar contradicciones de la aplicación de principios explicativos son algunas de las funciones que estos dispositivos han desempeñado. Por otra parte, los casos revisados hasta aquí exhiben diferentes estrategias que se articulan en la descripción del escenario imaginario con la presentación de ciertos principios teóricos. Fundamentalmente, las idealizaciones y las suposiciones contrafácticas han sido empleadas para construir un escenario imaginario. Algunas reflexiones acerca de las estrategias características de los experimentos mentales y de sus principales funciones aparecen diseminados en los textos pioneros de la filosofía de la ciencia, aunque la temática no ocupa un lugar de importancia hasta finales del siglo XX. Sin embargo, en estas elaboraciones

iniciales se encuentran esbozadas las principales perspectivas analíticas contemporáneas. A continuación, se revisan estas perspectivas pioneras.

1.3 El problema epistemológico de los experimentos mentales

El tratamiento filosófico de los experimentos mentales, es generalmente referido a las reflexiones que acompañan la madurez de la física moderna. Algunos filósofos han indicado que el origen del concepto se remite a la noción griega de δεικνυμι que designaba a las pruebas no formales empleados por los matemáticos pre-euclídeos (Lakatos, 1976: 10), pero existen escasas referencias a tratamientos metacientíficos de estas prácticas hasta el siglo XIX. Trabajos recientes realizaron una reconstrucción del desarrollo de la teorización sobre estas prácticas y situaron el origen de la filosofía de los experimentos mentales en diferentes elaboraciones de la filosofía alemana del siglo XVIII (Fehige y Stuart, 2014: 179-220). Si bien no es el propósito de esta sección historiar las conceptualizaciones acerca de estos dispositivos, se desarrollan brevemente los primeros tratamientos de esta cuestión y se esbozan algunas de las razones de la ausencia de esta práctica científica tan extendida como temática de la filosofía de la ciencia. Se consideran como punto de partida los trabajos que acuñaron o hicieron uso por primera vez del término y que ponen explícitamente en relación a los experimentos mentales con los experimentos ordinarios y el razonamiento hipotético.

Las primeras reflexiones acerca de las características epistémicas de los experimentos mentales coinciden con el desarrollo de la filosofía de la ciencia como disciplina independiente y se encuentran diseminadas en artículos y apéndices a los textos canónicos. La noción de experimento mental ha sido asociada al concepto de *Gedankenexperiment*, acuñado por Mach (1897 y 1905). Sin embargo, el término aparece con anterioridad en el trabajo de Ørsted (en danés

Tankeexperiment), ([1811] 1998). Estos tratamientos inaugurales ubicaron a los experimentos mentales entre otras formas de razonamiento científico y bosquejaron explicaciones acerca de su funcionamiento.

Aunque Ørsted no presenta una definición de los experimentos mentales ni les atribuye objetivos cognitivos específicos, dos elementos de su caracterización son centrales: los ubica dentro de la metodología científica y los asimila a una forma de razonamiento común a las matemáticas y a las ciencias fácticas. Desde su perspectiva, los experimentos mentales funcionan como hipótesis que permiten inferir deductivamente consecuencias acerca del comportamiento del mundo natural. Estas conjeturas deben ser confrontadas posteriormente con los resultados de experimentos reales (Ørsted, [1811] 1998: 298). Según Ørsted, se trata básicamente del mismo tipo de actividad que tiene lugar en la experimentación ordinaria con objetos físicos. Los experimentos mentales constituyen una forma de exploración de hipótesis en la que la mente no es meramente contemplativa. Al imaginar un punto que dibuja una trayectoria o al representarse una línea que gira que alrededor de uno de sus extremos, la mente es activa (Ørsted, [1911]1998: 296) en el mismo sentido en que lo es cuando se realiza un experimento real.

Mach, por su parte, establece una descripción más precisa de la naturaleza y funciones de los experimentos mentales apoyada en el examen de varios casos paradigmáticos.² Considera que los experimentos mentales son instancias del método de la variación concomitante de Mill, esencialmente similares a los experimentos reales. Su función es la de iniciar, a través de la descripción de un escenario específico, un proceso mental que permite movilizar el conocimiento instintivo con el fin de desarrollar un conocimiento nuevo y explícito. Asimismo,

² Mach dedica un considerable espacio a los experimentos mentales en su obra. En *La ciencia de la mecánica* discute en detalle el experimento de Stevin y el del balde de Newton. Luego, en 1876 les dedica un opúsculo que posteriormente reformula e incluye como capítulo en su obra *Conocimiento y error* de 1905.

Mach elabora una explicación en el marco de la epistemología empirista acerca del fundamento del conocimiento que pueden proporcionar. La experimentación consiste, desde esta perspectiva, en una habilidad innata que comparten los seres humanos y algunas otras especies animales. Se fundamenta en la capacidad de representarse ciertas circunstancias y de asociar con ellas determinadas conjeturas. El conglomerado de experiencias que se imprime en los sujetos a través de la observación constituye la base respecto de la cual se confrontan las conjeturas procedentes de nuestra capacidad innata de representación (Mach, [1905] 1948: 158). Este conjunto de impresiones de la naturaleza, tiene la doble ventaja de ser, por un lado, más accesible al sujeto que las experiencias sensibles y por otro lado, de no tener una naturaleza subjetiva. Los experimentos mentales permiten descubrir nuevas propiedades acerca de hechos físicos por la combinación de nuestra capacidad de representación y el conocimiento instintivo procedente de la experiencia. Ante una situación que involucra experiencias familiares, resultan fácilmente deducibles aquellas consecuencias que entrarían en flagrante contradicción con la masa de experiencias acumuladas. En algunos casos, a partir de este ejercicio es posible arribar a un resultado definitivo y en otros el experimento mental es una preparación para la realización de experimentos ordinarios. Asimismo, este método tiene para Mach un alto valor didáctico y un papel fundamental en el desarrollo cognitivo.

A veces una mirada echada retrospectivamente sobre experiencias anteriores y la ficción de nuevas combinaciones de circunstancias podrían instruirnos sobre la exactitud con que nuestros pensamientos representan las experiencias y sobre el grado de concordancias de los pensamientos entre sí. Se trata aquí de un procedimiento de depuración lógica y económica para estudiar el contenido de las experiencias puestas en forma de pensamientos: ¿Qué circunstancias determinan un cierto resultado? ¿Qué circunstancias son dependientes o independientes unas de otras? Una mirada de conjunto nos proporciona más luz sobre esas cuestiones de lo que podría hacer la experiencia detallada. (Mach, [1905] 1948: 161)

En *La Ciencia de la mecánica* (1893) Mach examina varios casos históricamente notables de experimentos mentales en física. Subraya allí la

importancia de la experimentación mental como una especie de la experimentación real. Distingue asimismo entre usos legítimos e ilegítimos de estas prácticas. El experimento de Stevin ([1983] 1919: 24-32) representa un empleo lícito de estas prácticas. La eficacia del experimento del plano inclinado descansa en el supuesto de que la cadena no puede moverse incesantemente. Dicho supuesto contiene conocimiento puramente instintivo, a saber, que no existen máquinas de movimiento perpetuo. No obstante, la conclusión a la que arriba Stevin parece contener información adicional. El resultado del experimento se apoya en la representación de cierto estado de cosas que entra en clara contradicción con la masa de experiencias acumulada. La abstracción de las condiciones que afectarían la experiencia real, como el rozamiento, facilitan la inferencia al principio de equilibrio de los pesos sobre el plano inclinado. La autoridad de una forma de conocimiento instintivo combinada con un gran poder de abstracción le confieren al experimento de Stevin su fuerza persuasiva. Por otro lado, el experimento del balde en rotación empleado por Newton para probar la existencia del espacio absoluto es para Mach un caso de uso ilegítimo de los experimentos mentales. Esto se debe a que el caso se refiere a una construcción puramente mental que no puede ser puesta en relación con la masa de conocimiento experiencial. Dado que no podemos extender nuestro conocimiento más allá de los límites de la experiencia, Mach considera que el concepto de espacio absoluto es espurio. Asimismo, sugiere que la experiencia impugna las conclusiones a las que Newton arribó. El fenómeno atribuido al efecto del espacio absoluto, es en realidad el efecto de la rotación respecto de otros cuerpos y no se produciría en ausencia de los mismos ([1983] 1919: 227-236). En síntesis, desde la perspectiva de Mach los experimentos mentales son esencialmente el mismo tipo de práctica que los experimentos reales y en sus usos legítimos emplean la masa de conocimiento empírico impresa en los sujetos para contrastar conjeturas e hipótesis.

A estas contribuciones iniciales a la conceptualización de los experimentos mentales, les fue seguido un largo silencio. Otros aportes importantes se encuentran recién hacia mediados del siglo XX. Una de las principales razones de la ausencia de los experimentos mentales en las reflexiones sobre el razonamiento científico puede encontrarse en el influjo de la filosofía empirista del Círculo de Viena. Su vehemente rechazo de la filosofía especulativa motivó, en parte, la idea de que estos dispositivos no pueden constituirse en herramientas metodológicas para el conocimiento del mundo natural. De manera que prevaleció entre los filósofos de la ciencia cierto escepticismo respecto de las funciones que los experimentos mentales son capaces de desempeñar en la justificación de hipótesis científicas. Mucho antes, Duhem ([1914] 1982: 304–311)³, había condenado los experimentos mentales en ciencia como argumentaciones falaces o peticiones de principio. Desde su perspectiva, estos son incapaces de incrementar nuestro conocimiento acerca del mundo natural y, por tanto, de sustituir a los experimentos reales en esta empresa.

Popper (1959) compartió la sospecha de Duhem respecto de los usos de los experimentos mentales en el contexto de justificación. Aunque reconoció su fecundidad en la interpretación de la teoría cuántica, la termodinámica y la teoría de la relatividad limitó sus usos legítimos a las funciones heurísticas y refutatorias. Desde esta perspectiva, si bien es posible reconocer la recurrencia de los experimentos mentales en la literatura científica, el empleo de los mismos como evidencia a favor de una hipótesis le parece inadmisibles. Es decir, estos no pueden desempeñar funciones probatorias o *apologéticas* en el razonamiento científico, debido a que son incapaces de proveer evidencia a favor de una

³ Kühne, (2009) le atribuye al triunfo del método experimental galileano el abandono de este tipo de prácticas en los años posteriores y a las críticas al apriorismo kantiano el rechazo de los experimentos mentales como prácticas capaces de proporcionar conocimiento acerca del mundo natural. Por su parte, Buzzoni (2017: 4) considera que las perspectivas de Mach y Duhem pueden ser puestas en continuidad ya que el primero considera que la experiencia es la piedra de toque de los experimentos mentales y los desarrollos de la filosofía de la ciencia posterior continúan en esta línea.

hipótesis, y mucho menos, de corroborarla. No obstante, Popper reconoce como lícitos otros empleos. El *uso crítico* de estos experimentos es legítimo en el contexto de justificación ya que consiste en mostrar las contradicciones o inconsistencias de una teoría. El representante más conspicuo de este uso es el caso esgrimido por Galileo contra la teoría aristotélica del movimiento. La *función heurística* de los experimentos mentales también es una función legítima; en estos casos la representación de un escenario imaginario sirve a la exposición o ilustración de las teorías. Ejemplos de este uso se encuentran en los argumentos empleados para representar la indivisibilidad del átomo en el atomismo antiguo. En el Apéndice XI de *The Logic of Scientific Discovery*, Popper esboza una regla para el uso eficaz de los experimentos mentales. En el marco de una controversia o disputa el uso de los experimentos mentales está sujeto a una norma de acuerdo con la cual su empleo solo es legítimo en los casos en los que el punto de vista del oponente es establecido con claridad y cualquier idealización del argumento defendido sea concedido o aceptado por el oponente. En general, Popper considera que los experimentos mentales son asimilables a métodos argumentativos.

Por otro lado, parte de la ausencia de los experimentos mentales como tema de la reflexión filosófica puede atribuirse a la centralidad que ocupó el estudio de las teorías en la filosofía de la ciencia hasta la década de 1960. La aparición de perspectivas historicistas, que pusieron de relevancia factores que habían sido hasta el momento ignorados en el estudio del razonamiento científico, propició nuevas consideraciones acerca de la naturaleza y funciones de la experimentación. Koyré puso especial atención en la Revolución Científica y examinó los experimentos empleados por Galileo para refutar la teoría aristotélica e introducir la propia. El detallado estudio histórico de la ley galileana de la caída de los cuerpos (Koyré, 1960) es uno de los primeros trabajos que señala la importancia de los experimentos mentales en el desarrollo de algunas nociones centrales de la ciencia moderna. Allí Koyré proporciona una interpretación

apriorista de los experimentos mentales en el contexto de justificación de las hipótesis físicas. Desde su punto de vista, este método es sumamente efectivo en proporcionar evidencia a favor de una teoría porque, a diferencia de los experimentos reales que siempre comportan cierto grado de imprecisión e incertidumbre, permite: “comprender lo real sensible como una desviación del modelo puro que nos ofrece” (Koyré, 1960: 250). Al operar con objetos teóricamente perfectos, el método de los experimentos mentales posibilita concretar la teoría y adelantarse a la experiencia. Desde esta perspectiva, los experimentos mentales son capaces de justificar hipótesis científicas sin apelar a nuevos datos empíricos.

¿Cómo realizar una caída en el vacío antes de la invención de las bombas neumáticas? Y en cuanto a los experimentos en el lleno, ¿cómo medir exactamente la ventaja o el rezago insignificantes de los cuerpos lanzados desde lo alto de una torre antes de la invención de los relojes de precisión? ¿Cómo hacer una medición exacta de la densidad del aire?” (Koyré, 1960: 242)

Para Koyré la aserción galileana de la caída simultánea de los cuerpos graves no descansa, más que en razonamientos *a priori* y en experimentos imaginarios. La estrategia de Galileo consiste en dar un paso al límite y probar el principio teórico a partir de la presentación de una situación idealizada. La experiencia probaría el principio en cuestión si la realización del experimento hubiese sido posible. En síntesis, de acuerdo con Koyré la confianza de Galileo en los experimentos mentales está justificada porque “la buena física se hace *a priori*” (Koyre: 1960: 257).

Kuhn (1964), por su parte, atribuyó a los experimentos mentales un importante papel en los procesos de reconceptualización que tienen lugar durante las revoluciones científicas. En su artículo dedicado a esta cuestión se encuentra la primera formulación explícita de la paradoja del conocimiento en los experimentos mentales:

Concediendo que todo experimento imaginario cuyos resultados han sido fructíferos incorpora en su diseño alguna información previa sobre el mundo, ésta no se halla en discusión dentro del experimento. [...] ¿Cómo es entonces que, apoyado exclusivamente en datos familiares, se puede llegar con un experimento imaginario a una nueva comprensión de la naturaleza? (Kuhn, [1964] 1977:19)

Según Kuhn, los filósofos han considerado que los experimentos mentales desempeñan un papel marginal en la justificación de las teorías, pero esta opinión se funda en una mala comprensión del proceso de constitución de los conceptos. Dado que dependen de información empírica que ya se encuentra disponible en el sistema de creencias del experimentador, les fue atribuida la única función de ayudar a reconocer y eliminar confusiones inherentes al aparato conceptual del científico. De acuerdo con Kuhn, no existe una alteridad entre conceptos y experiencia, sino que los primeros se encuentran unidos en una matriz de ley, teoría y expectativas empíricas de la cual no pueden ser extraídos.

Los rasgos perceptivos que un paradigma ilumina hasta el punto de mostrar sus regularidades, cambian cuando cambian los compromisos del científico con el paradigma. Por lo tanto, no pueden encontrarse datos brutos de la experiencia, sino más bien algo recogido con dificultad por el científico hasta que la experiencia inmediata esté determinada por el paradigma. (Kuhn, [1962] 2004: 232).

La ejecución de un experimento mental que le permite descubrir al científico que cierto concepto es, por ejemplo, inadecuado para dar cuenta de un fenómeno, le permite conocer simultáneamente algo acerca del mundo y acerca de su aparato conceptual. Así por, ejemplo, el experimento que Galileo emplea para poner de manifiesto las paradojas a las que conduce la aplicación del concepto aristotélico de velocidad e introducir su propia noción, logra su propósito porque muestra que la estructura del mundo no encaja en la matriz conceptual que se pretendía aplicar. La eficacia del experimento no depende de una mera operación lógica que muestra una contradicción conceptual, sino que se deriva de las dificultades del intento de “hacer encajar” una experiencia aún no asimilada al aparato conceptual.

Desde el punto de vista de Kuhn, los experimentos mentales cumplen funciones similares a las desempeñadas por la observación y los experimentos reales, aunque no presentan datos empíricos nuevos. Al mostrar inadecuaciones entre las expectativas teóricas y la experiencia le proporcionan al científico acceso a información que ya tiene y que, a pesar de ello, le resulta de alguna manera inaccesible. Como herramientas analíticas de este tipo, los experimentos mentales aparecen principalmente durante los períodos de crisis, ya que contribuyen a promover la reforma conceptual básica haciendo posible que el científico emplee, como parte integral de su conocimiento, lo que antes le resultaba inaccesible. En este sentido, y en oposición a la opinión extendida (de acuerdo con la cual estos no pueden cumplir funciones relevantes en los episodios de cambio teórico), los experimentos mentales desempeñan para Kuhn un papel similar, en algunos aspectos, al de las anomalías.

En suma, los primeros tratamientos de los experimentos mentales en la literatura filosófica tendieron a destacar un conjunto de rasgos esenciales de estos dispositivos: su parentesco con la experimentación científica ordinaria, su naturaleza mental o imaginaria y su importancia en el desarrollo de la ciencia moderna. La perspectiva empirista asumió que la única fuente de conocimiento acerca del mundo natural es la experiencia, por lo que los experimentos mentales capaces de proporcionar conocimiento significativo funcionan en base a la recuperación o reconstrucción del conocimiento experiencial. Mach desarrolló un argumento evolucionista de acuerdo con el cual, si se considera que la experiencia es la única fuente fiable de conocimiento y que los seres humanos comparten con los animales la capacidad innata de conjeturar, es posible dar sentido a la idea de que los experimentos mentales son el resultado del refinamiento de la habilidad de experimentar apoyada en el conglomerado de conocimiento empírico. El influjo de la filosofía positivista tuvo como consecuencia cierta desconfianza respecto de los poderes epistémicos de estas herramientas, que fueron vistas como tácticas especulativas. En la filosofía de Popper, el reconocimiento de su valor

heurístico y de la importancia de las funciones relativas al descubrimiento y derivación de inconsistencias, explican su relevancia en la argumentación científica, aunque sus usos legítimos están restringidos a la ilustración y refutación de hipótesis. No obstante, el poder refutatorio de los experimentos mentales legítimos, no está fundado para Popper en su carácter empírico sino en su carácter lógico: exhiben inconsistencias en una teoría y de esta manera prueban que no puede ser verdadera. El desarrollo de perspectivas historicistas dio lugar a nuevas consideraciones acerca de la naturaleza de los experimentos mentales que se alejaron de la línea empirista. Koyré concluyó del estudio histórico del principio de equivalencia de Galileo, que los experimentos mentales son perfectamente capaces de proporcionar conocimiento *a priori* de la naturaleza. Kuhn por su parte, les atribuyó una función específica en los procesos de reconceptualización, rol que pueden cumplir en virtud de que permiten evidenciar incompatibilidades o inadecuaciones entre los conceptos y las expectativas empíricas.

Estos desarrollos iniciales proporcionaron una primera formulación de la paradoja del *nuevo* conocimiento en los experimentos mentales y establecieron las bases de las líneas de investigación contemporáneas a partir de dos maneras de responder a esta cuestión. El problema epistemológico de los experimentos mentales en las perspectivas contemporáneas consiste en i) determinar qué funciones pueden desempeñar respecto de las hipótesis y teorías empíricas, ii) establecer cuáles son las fuentes de conocimiento acerca del mundo natural que son capaces de proveer y iii) sobre qué criterios se puede realizar una distinción entre experimentos exitosos y fallidos. Los experimentos mentales permitirían, en principio, ampliar nuestro conocimiento del mundo natural a través de la contemplación de escenarios imaginarios, es decir, prescindiendo de la intervención en el mundo natural. Una de las interpretaciones posibles del funcionamiento de los experimentos mentales, la más audaz y discutible, consiste en afirmar que son capaces de justificar teorías científicas, esto es, de

proporcionar evidencia a favor y en contra de hipótesis, en base a la intuición intelectual. Desde esta perspectiva, los experimentos mentales suministran modelos ideales para la interpretación de los fenómenos. Otra interpretación posible propone que, dado que solo puede haber conocimiento del mundo a partir de la experiencia, los experimentos mentales son incapaces de proporcionar evidencia a favor de hipótesis científicas (a lo sumo pueden ilustrar teorías o mostrar inconsistencias en una teoría) y su alcance debe ser restringido al de los argumentos teóricos y las hipótesis heurísticas.

1.4 Los experimentos mentales en la filosofía de la experimentación

Una parte significativa de la reflexión sobre el conocimiento científico se ha centrado, en los últimos años, en la actividad experimental. No obstante, la importancia que este tópico tuvo en el estudio histórico y social de la ciencia, el análisis sistemático de la experimentación por parte de la filosofía ocupó, hasta principios de la década de 1980, un lugar periférico. Hasta entonces prevaleció la imagen de la ciencia de acuerdo con la cual la experimentación es siempre precedida por la teoría y solo adquiere sentido a partir de ésta.

En contra de esta posición tradicional, la perspectiva de los filósofos de la experimentación destacó la independencia y riqueza de las prácticas experimentales e identificó un conjunto de problemas epistemológicos que tales prácticas plantean. La obra de Hacking (1983) constituye uno de los primeros trabajos que adopta este enfoque. De acuerdo con Hacking, la filosofía de la ciencia, desde el segundo decenio del siglo XX, consistió casi exclusivamente en las teorías científicas (Hacking, 1983:150). Dentro de esta disciplina, el examen de las observaciones preteóricas y los experimentos fue prácticamente nulo. El prejuicio a favor de la teoría condujo, en su versión más fuerte, a la creencia de que los experimentos sólo son significativos si tienen la finalidad de contrastar

alguna conjetura teórica, es decir, una hipótesis o teoría previamente formulada. Hacking se propuso imprimir un giro hacia el empirismo baconiano y poner de manifiesto la independencia de la experimentación respecto de la teoría y sus múltiples funciones en diferentes contextos de la investigación científica.

Con este mismo enfoque, las obras de Franklin (1984) y Galison (1987) desarrollaron estudios de casos con los que buscaron esclarecer, entre otras cosas, los rasgos más relevantes de la intervención experimental, una serie de criterios para evaluar la confiabilidad de los resultados, las fases de la experimentación, y las relaciones entre experimentación, teorías e instrumentos. Posteriormente, un segundo grupo de obras, entre ellas las de Collins (1985) y Pickering (1992), aparece a raíz de las controversias en torno a la autonomía experimental y a las suspicacias surgidas respecto de los criterios de confiabilidad de los resultados experimentales.

Aunque los enfoques desde los cuales exploran las prácticas experimentales son heterogéneos, puede afirmarse que los filósofos de la experimentación suscriben principalmente tres tesis: a) la experimentación es independiente de la teoría; b) la experimentación es capaz de cumplir múltiples funciones; y c) la observación proporciona una base de conocimiento objetiva. Asimismo, existe un acuerdo más o menos extendido acerca de las características que los experimentos reales comparten. En general, los filósofos convienen en que para llevar a cabo un experimento, los experimentadores tienen que intervenir activamente en el mundo material, y que, al hacerlo, producen nuevos objetos, fenómenos y procesos (Radder, 2003: 4). La realización material del sistema experimental incluye como componentes el objeto de estudio, el aparato y su interacción y la intervención activa en el medio ambiente de ese sistema. En este sentido, el experimento se diferencia de la teoría, incluso si el trabajo teórico es siempre acompañado de actos materiales como escribir una fórmula. También prevalece la opinión de que no cualquier tipo de intervención en el mundo cuenta como un experimento. En general, los experimentos requieren de cierta estabilidad

y reproducibilidad. Satisfacer estos requisitos presupone un control del sistema experimental y su medio ambiente, y cierta disciplina por parte de los experimentadores.

Los estudios filosóficos de la práctica experimental han conducido a esclarecer las diferentes funciones que los experimentos pueden cumplir respecto de la teoría. Además de las funciones refutatorias y confirmatorias, a las que tradicionalmente fueron asociados, los estudios de casos han puesto de manifiesto que los experimentos pueden cumplir funciones exploratorias, como producir fenómenos nuevos, ilustrar teorías, o incluso dar lugar a nuevas perspectivas teóricas. Por otra parte, al estudiar el papel de la tecnología en la experimentación, los filósofos han puesto atención en los aparatos y equipos empleados y han mostrado que la investigación de los instrumentos científicos es una rica fuente de ideas para una filosofía de la experimentación científica.

En esta línea de investigación los experimentos mentales son concebidos como tipos particulares de argumentos teóricos que pueden ser o no ejecutables como experimentos reales. Desde la perspectiva de Hacking los experimentos ordinarios tienen *vida propia* (Hacking, 1983:150) porque pueden madurar, evolucionar, adaptarse o reestructurarse. Son capaces de producir adelantos en la tecnología de los instrumentos con el fin proporcionar mejores resultados, además de desarrollar las habilidades de los experimentadores. Incluso pueden ser llevados a la luz de nuevas teorías. Pero los experimentos mentales son fijos y, en gran parte, inmutables (Hacking, 1992: 307). Esto significa que están limitados a un contexto y no pueden ser empleados en otros. Al depender de una serie de supuestos básicos tomados de un marco nomológico, los poderes epistémicos de los experimentos mentales se limitan a su capacidad de exhibir tensiones en el marco conceptual de los experimentadores. En síntesis, la dependencia teórica y la ausencia de intervención fueron las principales razones para excluir a los experimentos mentales del tratamiento sistemático de las prácticas experimentales.

Considerar a los experimentos mentales como meros argumentos parece no hacer justicia a su la importancia histórica. Por otra parte, aunque muchos experimentos mentales pueden reformularse como argumentos, esto no explica por qué los mismos resultados pueden obtenerse de manera más rápida y fácil a través de la ejecución de un experimento mental. Al reemplazar el experimento mental por un argumento éste pierde al menos parte de su “fuerza demostrativa”.⁴

Existen similitudes entre experimentos mentales y reales, al menos en dos aspectos importantes que no fueron tomados en cuenta por la filosofía de la experimentación. En primer lugar, puede mencionarse el rol activo que el experimentador tiene en la experimentación mental. Esta es una característica distintiva de los experimentos reales. El relato del experimento mental incluye una serie de indicaciones que el experimentador debe seguir.⁵ Se requiere que el experimentador lleve a cabo alguna operación, active algún mecanismo o manipule ciertas variables en la imaginación y visualice un fenómeno que resulta de estos procedimientos. En este aspecto, se asemejan a los experimentos reales y se distinguen de los argumentos. En segundo lugar, es posible identificar en los experimentos mentales una característica que hace posible argumentar su independencia de las interpretaciones teóricas. De acuerdo con una perspectiva ampliamente difundida en filosofía de la experimentación, en cada una de las fases de la experimentación real (preparación, interacción y detección), es posible distinguir la realización material de su descripción teórica o interpretación. Con la noción de “realización material” se pretende aludir al conjunto de operaciones que se ejecutan materialmente. Esta noción permite definir “la igualdad de acción a pesar de las diferencias de interpretación” (Radder, 1996: 12). De esta propiedad depende la reproducibilidad, un rasgo característico de la experimentación exitosa. Es posible argumentar que en los experimentos mentales, del mismo

⁴ Este es uno de los argumentos DeMey (2003: 62) para defender el carácter experimental de los experimentos mentales.

⁵ Brown (2011: 1 y 2003: 35) señala estas semejanzas.

modo que en los reales, se puede distinguir el proceso experimental (equivalente a la *realización material*, al menos en relación con las fases de preparación e interacción) de su interpretación teórica.⁶ Una justificación más amplia del carácter experimental de los experimentos mentales se desarrolla en el Capítulo 2.

Una razón adicional que ha motivado el examen del papel epistémico de los experimentos mentales es la analogía entre estos y las simulaciones computacionales. La filosofía de la experimentación ha admitido en los últimos años, que los modelos y las simulaciones son también fuentes de conocimiento experimental a pesar de que carecen de intervención en el mundo material. Se han elaborado argumentos a favor de considerar a los modelos como instrumentos de medición (Morrison, 2009: 34). Esto es lo que permite poner a cierta clase de simulaciones a la par de mediciones experimentales. Asimismo, en astrofísica y cosmología, donde la experimentación real no es posible, la simulación es una fuente aceptable de conocimiento experimental. Por su parte, las estrategias de modelización empleadas en las simulaciones son esenciales en la experimentación real. Si bien algunos autores consideran que las simulaciones funcionan como otros instrumentos, ampliando nuestra capacidad de observación (Humphreys, 2004), la discusión acerca de su estatus epistémico está muy lejos de resolverse. Las simulaciones computacionales podrían describirse como instrumentos que tienen la función de imitar o representar el comportamiento de un sistema real. Algunas posiciones les atribuyen un carácter similar al de la experimentación real y consideran que pueden proporcionar evidencia para la justificación de una teoría científica. La interpretación de los

⁶ Esta analogía entre experimentos mentales y reales ha sido empleada por Bishop (2012) para justificar que los experimentos mentales no son meramente argumentos.

experimentos mentales como modelos mentales provee de argumentos análogos para considerarlos como fuentes de conocimiento experimental.

Las reflexiones elaboradas en el marco de la filosofía de la experimentación no proporcionaron una interpretación satisfactoria de la naturaleza y funcionamiento de los experimentos mentales. En algunos de los trabajos más importantes dentro de esta corriente, los experimentos mentales son deliberadamente ignorados. Aunque en general se ha asumido que la falta de intervención en el mundo natural es una razón suficiente para asimilarlos a argumentos teóricos escasean en el marco de la filosofía de la experimentación, argumentos explícitos que apoyen esta conclusión.

1.5 El estado actual de la discusión

El examen sistemático de los experimentos mentales en el contexto de las ciencias fácticas es relativamente reciente y el análisis restringido al uso de esta clase de experimentos en el contexto de justificación de las teorías físicas es incipiente. Los primeros trabajos completamente dedicados al examen del papel epistémico de los experimentos mentales fueron las obras monográficas de Brown (1991) y Sorensen (1992). El trabajo de Brown ejerció una enorme influencia en los desarrollos posteriores. La interpretación de Norton (1991, 1996) acerca de los poderes epistémicos de los experimentos mentales rivalizó directamente con la de Brown desde su aparición y ambas posiciones dieron lugar incluso a trabajos críticos dedicados exclusivamente a evaluar sendas perspectivas en función de sus virtudes y limitaciones explicativas (por ejemplo, Bishop, 1999 y Clatterbuck, 2012).

Los trabajos publicados en los *Proceedings of the Philosophy of Science Association* de 1992, resultaron determinantes para la evolución de diferentes líneas de análisis, algunas emparentadas con las posiciones de Norton y de Brown

y otras completamente originales. A partir de estas publicaciones iniciales, el número de trabajos centrados en la temática ha crecido mucho durante la última década. Recientemente, varias compilaciones de artículos y números de reconocidas revistas del área fueron enteramente dedicados a los experimentos mentales (por ejemplo, Frappier, Meynell y Brown, 2013; *Perspectives of Science* N° 22, 2014; Stuart, Feige y Brown, 2018). El tópico también fue incluido en recopilaciones de temáticas actuales en filosofía de la ciencia (Hitckcock, 2004). Otros estudios completamente dedicados a esta materia, aunque no exclusivamente centrados en los experimentos mentales en ciencias fácticas, de reciente data son: Häggqvist (1996); Gendler (2000), Rescher (2005); Cohen (2005); Ierodikanou y Roux (2011). Dos extensas compilaciones se abocaron también a este tema (Frappier, Meynell y Brown, 2013 y Stuart, Feige y Brown, 2018).

La discusión actual en torno a los experimentos mentales en las ciencias fácticas se centra en el mencionado *problema epistemológico* de estos más o menos en los mismos términos en los que fuera formulado por Kuhn ([1964] 1977: 19). En general, las propuestas existentes intentan determinar cuáles son las funciones que esta clase de experimentos puede desempeñar respecto de hipótesis y teorías científicas y establecer cuáles son las fuentes del conocimiento acerca del mundo natural que son capaces de proveer.

El libro de Sorensen (1992) presentó una teoría general de los experimentos mentales, tanto en ciencia como en filosofía, partiendo de la tesis de que existe una diferencia de grado, aunque no de clase, entre experimentos mentales y reales. De acuerdo con esta perspectiva, los experimentos mentales proceden de los reales por un proceso de atenuación del elemento de ejecución. Su funcionamiento es sustancialmente similar al de los experimentos ordinarios: a partir de la variación de los valores de una variable independiente se intenta rastrear el comportamiento de la variable dependiente. Mientras que los experimentos reales plantean cuestiones sobre las relaciones específicas entre

variables independientes y dependientes a través de un proceso de manipulación de fenómenos, los experimentos mentales buscan lograr el mismo objetivo a partir de la mera reflexión sobre el aparato experimental. De acuerdo con Sorensen, la ejecución material no es el rasgo distintivo de las prácticas experimentales:

La mayoría de los experimentos son acciones planificadas. Sin embargo, el componente de ejecución es inesencial, así como la realización física no es esencial para las dietas y los ejercicios. Cada uno de éstos puede existir simplemente como un conjunto de instrucciones. Por supuesto, las instrucciones son usualmente diseñadas para ser seguidas. Pero a veces un procedimiento cumple funciones independientes de su ejecución. (Sorensen, 1992: 191)

Continuando la línea inaugurada por Mach, la exégesis de la naturaleza de los experimentos mentales que ofrece Sorensen se funda en una concepción evolutiva de la capacidad humana de experimentar. El fundamento del conocimiento acerca del mundo natural que estos experimentos pueden proporcionar se basa en el conocimiento empírico que se encuentra “almacenado” de manera tácita en el sistema de creencias del experimentador:

[...] tenemos un amplio stock de intuiciones sobre el espacio y el tiempo, sobre las relaciones biológicas y sociales y una gran aptitud para la psicología de la creencia-deseo y el razonamiento causal. Los seres humanos también fueron dotados de un excelente juicio visual y dominio del lenguaje. (Sorensen, 1991: 253)

Para Sorensen, una de las funciones más importantes que los experimentos mentales pueden desempeñar es la de exhibir las inconsistencias internas de una teoría. El tipo de inconsistencia local que los experimentos mentales científicos habitualmente presentan puede ser regimentada en la forma de un argumento en el que se derivan consecuencias modales de un principio teórico (esto quiere decir que de una afirmación que expresa un principio teórico se infiere una consecuencia necesaria o probable de él) que en conjunción con un grupo de condiciones iniciales provistas por un escenario contrafáctico (que es lógica o metafísicamente posible) da como resultado una consecuencia

metafísica o lógicamente imposible. Los diferentes tipos de experimentos mentales se distinguen por la manera en que se proporciona una interpretación que resuelva la paradoja planteada a partir de la presentación del escenario imaginario. En la perspectiva de Sorensen, los experimentos mentales son “expediciones a mundos posibles” (Sorensen, 1992: 135) cuya función principal es la de revelar consecuencias que son posibles o imposibles en el marco proporcionado por un principio teórico. Debido esta particular manera de reconfigurar nuestras creencias, Sorensen caracteriza a los experimentos mentales como *refutadores aléticos*.

El trabajo monográfico de Brown (1991) fue el primero completamente dedicado a los experimentos mentales en las ciencias fácticas y ofreció un minucioso análisis de algunos ejemplos canónicos. De acuerdo con Brown, las similitudes entre experimentos reales y mentales son suficientes para considerar que se trata del mismo tipo de prácticas. Su propuesta general, de base realista, establece que, al menos en algunos casos, los experimentos mentales nos proporcionan conocimiento *a priori* acerca del mundo natural. Si bien Brown no formula una definición precisa de la noción de experimento mental, propone una caracterización a partir de la presentación de ejemplos clásicos. En esos casos, identifica un conjunto de rasgos comunes: los experimentos mentales, según Brown, incluyen visualizaciones, involucran algún tipo de manipulación mental, no son meras consecuencias de una teoría y, en su mayoría, no es posible llevarlos a cabo como experimentos reales. Propone, además, una taxonomía que diferencia entre experimentos mentales con funciones constructivas y destructivas. Dentro de los experimentos mentales del primer tipo se encuentra una subclase cuya ejecución posibilita la intelección de propiedades de clase natural. Los denominados, *experimentos mentales platónicos* permiten captar leyes de la naturaleza de manera intuitiva. Esta clase de experimentos mentales proporciona las garantías necesarias para justificar una hipótesis científica. La posición de

Brown involucra la concepción de las leyes de la naturaleza como relaciones entre propiedades abstractas (universales) susceptibles de ser captadas por la mente.

La compilación de artículos de Horowitz y Massey (1991) fue el primer volumen colectivo dedicado especialmente a esta temática. En este trabajo se encuentra una variedad de contribuciones que amplían el espectro de elementos de análisis y el examen de casos de estudio. Entre otras, aparece una de las primeras explicaciones del funcionamiento de los experimentos mentales que incorpora algunos de los supuestos de la filosofía de la experimentación y se desarrolla desde la perspectiva empirista. Allí Norton (1991) sostiene una interpretación deflacionaria de los poderes epistémicos de los experimentos mentales de acuerdo con la cual el alcance epistemológico de un experimento mental coincide con el de un argumento. Norton sostiene que a pesar de su apariencia, que se asemeja a la narración de un experimento real, los experimentos mentales son argumentos teóricos que: i) postulan estados de cosas hipotéticos o contrafácticos, y ii) invocan particulares irrelevantes para la generalidad de la conclusión (Norton, 1991: 129).

De esta caracterización, se sigue una clasificación de los experimentos mentales de acuerdo con el mecanismo de reorganización de la información empírica empleado en el argumento. Los experimentos mentales del Tipo I son argumentos deductivos, usualmente por reducción al absurdo. Los experimentos del Tipo II son argumentos inductivos en los que a partir de la presentación de un caso se realiza una generalización en la que se eliminan las particularidades invocadas en la descripción del escenario imaginario (Norton, 1991: 131). La interpretación argumental proporciona un criterio para distinguir entre usos legítimos y espurios de los experimentos mentales. Dado que esta clase de experimentos debe su efectividad a la inferencia en virtud de la cual se obtienen sus conclusiones, la reconstrucción del experimento como un argumento permite identificar premisas falsas o inferencias inválidas (Norton, 2004: 52).

La principal estrategia que Norton emplea para defender esta concepción consiste en formular explícitamente los argumentos involucrados en casos históricamente relevantes de experimentos mentales en ciencia. Mediante la reconstrucción de ejemplos Norton pretende dar sustento a dos subtesis. Por un lado, se propone probar que a todos los experimentos mentales en física, subyace un argumento. Por otro lado, se propone mostrar que la fuerza heurística de estos depende de la eficacia del argumento en cuestión y que, por tanto, su apariencia experimental puede ser eliminada sin perjuicio de su eficacia persuasiva.

La perspectiva de Norton representa la posición dominante en filosofía de la ciencia de acuerdo con la cual los experimentos mentales, dado que carecen de intervención en el mundo natural, son reducibles a argumentos teóricos. Varios autores se alinean con esta posición. Rescher (1991), por ejemplo, considera que un experimento mental es básicamente una forma de razonamiento hipotético en la que se asume provisionalmente una hipótesis como verdadera para derivar una conclusión que se aplica a otros casos que pueden considerarse análogos. Irvine (1991) también comparte esta opinión y establece una serie de condiciones que los experimentos mentales científicos deben satisfacer y que los diferencian de otras instancias de razonamiento contrafáctico: deben estar en una relación privilegiada con las observaciones empíricas pasadas y con una teoría de fondo razonablemente bien desarrollada.

Las posiciones de Norton y Brown se encuentran enfrentadas ya que se sustentan en tesis ontológicas y epistémicas que tradicionalmente se han considerado opuestas. La discusión entre ambos acerca de si los experimentos mentales trascienden el empirismo y proporcionan algún tipo de conocimiento *a priori* acerca del mundo físico ha continuado hasta la actualidad (Brown, 2004a, 2004b, 2010; Norton, 2004a, 2004b) y no tiene todavía argumentos decisivos a favor de la posición platonista de Brown o de la posición empirista de Norton.

El trabajo monográfico de Häggqvist (1996) inspeccionó las principales perspectivas acerca de la naturaleza de los experimentos mentales y a partir de ese examen crítico proporcionó una interpretación novedosa del funcionamiento de los experimentos mentales en filosofía y ciencia. Häggqvist sostiene que la perspectiva argumental ignora una distinción fundamental entre entidades lingüísticas, como enunciados e inferencias, y entidades no lingüísticas, como eventos y procesos. A diferencia de los argumentos, los experimentos mentales no son compuestos de entidades con valor de verdad ni puede afirmarse significativamente que sean válidos o inválidos. El tipo de actividad hipotética o imaginativa relacionada con la ejecución de un experimento mental es similar a la sucesión de eventos físicos que tienen lugar en un laboratorio. En la realización de un experimento real, la manifestación de fenómenos causa que los observadores sostengan ciertos enunciados observacionales como verdaderos. Estos enunciados son subsecuentemente empleados como premisas de los argumentos cuyo objetivo es poner a prueba hipótesis o teorías. En síntesis, los experimentos mentales son, según Häggqvist procedimientos destinados a generar la aceptación o el rechazo de ciertos enunciados o proporcionar fundamentos para la creencia en ciertos argumentos.

Diversos autores, como Bishop (1999), McAllister (1996 y 2004) y Gendler (2000 y 2004), revisaron la discusión respecto de los poderes epistémicos de los experimentos mentales. Estas perspectivas introdujeron nuevos elementos para el análisis de esta práctica científica. Bishop examinó las relaciones entre instancias de replicación de un mismo experimento mental y argumentos diferentes desarrollados para interpretarlas. McAllister exploró la noción de fenómeno empleada en algunos casos paradigmáticos de experimentos mentales en la ciencia moderna, y Gendler inspeccionó el papel que la visualización en la imaginación de un escenario desempeña en la formación de creencias. Estas posiciones llamaron la atención sobre algunos aspectos de la experimentación mental que habían sido previamente pasadas por alto. Bishop introduce la

distinción entre el experimento mental propiamente dicho y el argumento que lo acompaña y señala, apoyándose en el estudio del experimento mental de la caja de fotones empleado por Einstein y Bohr, que un mismo experimento puede tener interpretaciones teóricas directamente opuestas. Este análisis recupera algunas categorías de la filosofía de la experimentación para conceptualizar el caso e instala la discusión acerca de los criterios de identidad de los experimentos mentales. El trabajo de McAllister por su parte, atiende a las características históricas y filosóficas involucradas en la noción de evidencia. Si bien el trabajo de Gendler está enfocado en los experimentos mentales en filosofía, su reflexión acerca del papel de las actividades de visualización introduce la cuestión acerca de cuál es el papel de la representación de un escenario imaginario en estos experimentos.

Miščević (1992) y Nerssesian (1992), por su parte, elaboraron una explicación del funcionamiento de los experimentos mentales basada en su similitud con los modelos mentales. De acuerdo con este punto de vista, los experimentos mentales consisten en la construcción de un modelo mental por parte de un científico que imagina una serie de eventos y procesos e infiere resultados. Este es el mecanismo en virtud del cual se obtiene nuevo conocimiento. Aunque el experimento se comunica a través de una narrativa, su ejecución consiste en la representación de un modelo mental que la narrativa permite construir. La facilidad con la que se derivan conclusiones de la ejecución de un experimento mental se explica porque los modelos mentales son un medio heurísticamente superior a las inferencias lógicas, debido a que en ellos se emplean habilidades usadas para ubicarse espacialmente, que habitualmente no son verbalizadas. Aunque el relato incluye abstracciones, la construcción de los modelos en la imaginación depende fundamentalmente de nuestro conocimiento perceptivo.

La interpretación de los experimentos mentales como modelos científicos, abre un marco interpretativo recientemente desarrollado en la filosofía de la

experimentación (por ejemplo, en Gelfert, 2016). Las simulaciones y los modelos computacionales y mecánicos han sido consideradas como fuentes de conocimiento experimental. La proliferación del uso de simulaciones computacionales en el contexto de justificación de teorías en las cuales la experimentación real no es posible, motivó su comparación con los experimentos mentales. Dentro de estos dominios, las funciones de las simulaciones son en apariencia similares a las que en principio pueden atribuirse a los experimentos reales. Algunas posiciones incluso consideran que su ejecución puede contribuir a la justificación de una teoría científica. Esta perspectiva se apoya en las similitudes entre las simulaciones y los modelos mecánicos y cuestiona el papel de la interacción causal con el fenómeno como justificadora del estatus epistemológico de los experimentos reales (Morrison, 2009). La tesis básica es que las simulaciones nos permiten crear el tipo de ambiente controlado en el que podemos variar las condiciones iniciales, los valores de las variables y parámetros, etc., por lo cual, en este sentido, funcionan como el equipamiento de laboratorio usado para medir magnitudes físicas.

Otros filósofos, en cambio, sostuvieron que la similitud señalada no da razones suficientes para extender a las simulaciones el estatus epistémico de los experimentos reales, Giere (2009) y Winsberg (2010), entre otros, consideran que la interacción causal con el fenómeno o sistema objetivo es fundamental para la comprensión de la noción de experimentación tal como tradicionalmente la entendemos. De acuerdo con esta perspectiva, las simulaciones están involucradas en el descubrimiento y la justificación de hipótesis científicas, aunque su uso más frecuente las liga al contexto de aplicación más que a la contrastación de teorías científicas (Winsberg, 2010). Dadas las similitudes entre experimentos mentales y simulaciones computacionales, las teorizaciones desarrolladas en torno a estas últimas proporcionan un marco para examinar analógicamente las funciones de los experimentos mentales que aún no ha sido suficientemente explorado.

Existen otras contribuciones que han agregado algunos elementos a la discusión de la función epistémica de los experimentos mentales, pero, en líneas generales, pueden ser asimiladas a alguna de las posiciones antes expuestas. Como fundamento del conocimiento que los experimentos mentales proporcionan, las diferentes teorías identifican a la intuición empírica (entendida como una forma de conocimiento instintivo producto de la experiencia sensible acumulada), la intuición racional o los esquemas conceptuales. Algunas de estas perspectivas explicitan además una serie de criterios que permitirían distinguir casos ilegítimos, es decir experimentos mentales que arrojan conclusiones falsas o resultados espurios. De las perspectivas mencionadas solo las de Norton, Brown, Sorensen y Häggqvist desarrollan un estudio sistemático de los experimentos mentales, es decir, solo estas perspectivas proveen una definición, una tipología y una fundamentación de sus funciones basada en el estudio de una variedad de casos. Solo la perspectiva argumental (representada fundamentalmente por Norton e Irvine) provee criterios para establecer su confiabilidad epistémica.

Prevalece entre los desarrollos teóricos contemporáneos una reticencia a definir explícitamente la noción de experimento mental. En su lugar se encuentran caracterizaciones, ejemplificaciones, descripciones de rasgos perennes, comparaciones y analogías con otras prácticas. La reconstrucción que aquí se realiza de las mismas tiende a simplificar los matices para acercar las caracterizaciones a definiciones. El punto de partida para un estudio metódico de los experimentos mentales parece demandar una delimitación adecuada de su objeto de estudio. A continuación se propone una definición preliminar que permite incluir un conjunto suficientemente amplio de casos y que señala las cuestiones relevantes para la indagación filosófica.

1.6 Definición provisoria y prospectiva

Una de las razones que pude identificarse como causante de la ausencia de una perspectiva unificada acerca de los experimentos mentales y de la escasez de atención que este tópico recibió dentro la filosofía tradicional de la ciencia, e incluso en el marco de la filosofía de la experimentación, es la diversidad de casos que se nuclean bajo este rótulo. Por otra parte, aunque existe una extendida coincidencia en la idea de que consisten en formas de razonamiento hipotético que tienen como objetivo principal examinar principios teóricos, no hay consenso acerca de si es conveniente abordar su estudio filosófico considerando a estas prácticas como experimentos o como argumentos. En general la adopción de una de estas dos posiciones se considera como una tesis acerca de la naturaleza de esta clase de experimentos que determina a su vez la inclusión o exclusión de casos de estudio.

Por otra parte, las perspectivas desarrolladas en el apartado anterior asumen que los experimentos mentales dependen de un ejercicio imaginativo que se realiza a partir de la lectura de una narrativa en la que se describe un escenario y una serie de instrucciones. No obstante, pocas de ellas explicitan este supuesto y casi ninguna describe las características de la narrativa ni elabora los aspectos epistémicos de esta actividad de representación. Asimismo, existen diferentes opiniones respecto del papel que los conceptos teóricos y la experiencia desempeñan en la evaluación de hipótesis a través de un experimento mental. Algunas perspectivas asumen que los experimentos mentales muestran las limitaciones de los conceptos al ser aplicados a la experiencia, otras suponen que permiten captar directamente los principios que gobiernan el comportamiento de los fenómenos, pero generalmente no se explicita el vínculo entre el tipo de razonamiento contrafáctico que presentan como típico de los experimentos mentales y los principios teóricos a los que se dirigen.

Una delimitación poco clara del objeto de estudio no posibilita un examen que se enfoque en los aspectos que son, desde el punto de vista de la filosofía de la ciencia, relevantes para un estudio comprensivo. Las caracterizaciones de los experimentos mentales previamente desarrolladas (quizás es conveniente denominarlas de este modo ya que, como se ha señalado, no constituyen propiamente definiciones) adolecen de una serie de limitaciones. Algunas son demasiado amplias en tanto no delimitan adecuadamente el tipo de práctica que constituye el objeto de estudio en cuestión y no son suficientemente explícitas respecto de las operaciones de visualización que la propia posición presenta como esencial a estos dispositivos. Este es el caso de la caracterización de Brown (1991 y 2011). Otras perspectivas son demasiado estrechas porque, aunque permiten establecer de manera más o menos unívoca criterios de confiabilidad epistémica, no dan cuenta de elementos que en principio son fundamentales en este tipo de experimentos. Este es el caso de la perspectiva argumental presentada por Norton (1991 y 1996). Otras posiciones tienen el defecto de la ambigüedad ya que no determinan de manera suficientemente clara la relación entre los supuestos contrafácticas y los principios teóricos. Esto resulta en ciertas arbitrariedades en las distinciones y en la selección de casos ejemplificadores. Las descripciones de Sorensen (1992), Mišćević (1992) y Nersessian (1992) son susceptibles de esta crítica.

La caracterización de Brown (1991 y 2011), por su parte, resulta ambigua en varios aspectos importantes. En primer lugar, no especifica suficientemente las condiciones que el escenario imaginario empleado para evaluar una hipótesis debe satisfacer. Señala que en todo experimento mental el experimentador debe realizar algún tipo de actividad y que la visualización es esencial para la obtención del resultado. No obstante, admite que en el caso de Galileo, la refutación del principio aristotélico de acuerdo con el cual la velocidad de la caída de los cuerpos es proporcional a su peso, se opera por una reducción al absurdo (Brown, 2011: 46). En este sentido, la visualización del escenario no cumpliría ninguna función

en la inferencia de las conclusiones ya que la contradicción se sigue de la realización de un argumento. En consecuencia, su caracterización es inadecuada para uno de los ejemplos que el mismo Brown considera paradigmáticos. Asimismo, no explica claramente qué quiere decir con manipulación mental. Esta característica es sumamente relevante para la defensa de su punto de vista debido a que la manipulación de una clase especial de fenómenos, los que ocurren en los experimentos mentales, son los que, según Brown, su carácter experimental. Brown se limita a señalar que los experimentadores no son meramente pasivos en la realización de un experimento mental en cuanto deben “activar” en su imaginación algún tipo de mecanismo y ver qué sucede. Por otra parte, señala que en ciertos experimentos mentales tiene lugar la presentación de fenómenos *a priori* que ejemplifican clases naturales o exhiben instancias “ideales” de las leyes.⁷ Incluso si se concediera que el experimentador es activo en cuanto imagina que realiza las acciones descritas en la narrativa del experimento, no todos los experimentos mentales presentan, según la caracterización de Brown, fenómenos *a priori*. Por lo que su definición a partir de rasgos comunes excluye muchos casos. Tampoco queda claro el alcance de la afirmación de que estos experimentos son irrealizables como experimentos reales. Por ejemplo, el experimento de Galileo, aunque incluye idealizaciones que lo tornaron técnicamente irrealizable en el momento de su formulación, podría ejecutarse (y de hecho ha sido realizado) como experimento real.

La definición de Sorensen (1992), de acuerdo con la cual los experimentos mentales en ciencia y filosofía pertenecen junto con los experimentos reales al mismo tipo de prácticas también presenta problemas. En primer lugar, existe una serie de diferencias sustanciales entre los casos de experimentos mentales en ciencia y los experimentos mentales en filosofía que resultan relevantes al momento de elaborar una caracterización de estas prácticas. En las ciencias

⁷ La cuestión de qué significado podría atribuírsele a la noción de fenómeno en los experimentos mentales se desarrolla en el capítulo 2. Una crítica a la interpretación que Brown hace de la noción de fenómeno se elabora en el capítulo 4.

fácticas, los experimentos mentales poseen un marco nomológico definido por el estado del conocimiento, en virtud del cual es posible evaluar si resultan epistemológicamente significativos.⁸ En los experimentos mentales filosóficos estas restricciones no se cumplen en algunos de los casos más representativos. Otra diferencia importante reside en el hecho de que la narrativa de los experimentos mentales científicos es similar a la descripción de un experimento real: se enuncian condiciones iniciales, se explicita el funcionamiento de algún aparato experimental y se describen las operaciones que el experimentador debería realizar. Los experimentos mentales en filosofía se asemejan más a una narrativa de ficción, en la que en ocasiones se describe una situación contrafáctica (o incluso contranomológica) de la que se derivan conclusiones.

La perspectiva de Norton, de acuerdo con la cual los experimentos mentales son reducibles a argumentos teóricos, también resulta limitada. En principio, esta descripción no hace referencia a la eficacia heurística de estos, ya que supone que los detalles involucrados en la descripción del escenario imaginario son irrelevantes para la realización del argumento. El punto de vista de Norton, según el cual los experimentos mentales son argumentos encubiertos, tampoco permite plantear la cuestión acerca de por qué en algunos casos los experimentos mentales son interpretados desde perspectivas teóricas diferentes.

La caracterización de Mišćević y Nerssesian, de acuerdo con la cual los experimentos mentales son modelos mentales, enfatiza su eficacia heurística, pero restringe la inclusión de algunos casos importantes en los que las suposiciones contrafácticas desempeñan un papel esencial. De acuerdo con Nerssesian, los modelos mentales permiten hacer inferencias de manera más directa que los argumentos porque involucran las mismas habilidades empleadas para la ubicación espaciotemporal. Lo que permite a un modelo ser un análogo

⁸ Wilkes (1988: 18) sugiere que la adecuación a lo que podría o no suceder dado nuestro conocimiento científico actual debe ser una restricción que los experimentos mentales filosóficos epistemológicamente significativos también deberían satisfacer.

estructural de un fenómeno es la representación espaciotemporal de las relaciones y de la estructura causal de los eventos y las entidades involucradas (Nerssesian, 1992: 294). No obstante, esta perspectiva no logra explicar cómo es posible que a partir de la narrativa los experimentadores construyan un análogo estructural de una situación que es de hecho imposible en el mundo real. Esto excluiría de los experimentos mentales los casos en los que los experimentos son irrealizables, como el ejemplo del balde giratorio de Newton, o en los que estas habilidades parecen no cumplir un papel importante, como el caso del demonio de Maxwell.

Con el fin de ampliar el espectro de casos que pueden ser incluidos en el análisis y proporcionar una definición inicial que incluya también los aspectos que las caracterizaciones elaboradas hasta el momento excluyen, resulta de utilidad recuperar los rasgos comunes que exhiben los casos presentados en la segunda sección de este capítulo. El más importante, que como se ha indicado es el que ha motorizado la reflexión filosófica, es que, en principio, dichos experimentos se proponen derivar conclusiones acerca del mundo natural sin introducir nuevos datos. Otro de los rasgos que se ha señalado se relaciona con el modo en que presentan sus resultados. Experimentos como el del plano inclinado elaborado por Stevin, hacen apremiante la aceptación de sus conclusiones. Algunos de los casos presentados fueron interpretados desde perspectivas teóricas diferentes, como ocurrió con el experimento de la caja de fotones de Einstein. Asimismo, los ejemplos emplean estrategias como la idealización y el uso de supuestos contrafácticos en la construcción del escenario imaginario. Además, los ejemplos presentados no fueron realizados como experimentos reales por diferentes razones: imposibilidad técnica, imposibilidad material o técnica o porque su ejecución como experimentos reales fue considerada trivial.

Es difícil decidir de antemano, si todos estos rasgos son esenciales a los experimentos mentales. La incursión en el escenario imaginario, por ejemplo, parece ser una condición esencial para derivar conclusiones en los experimentos

mentales, pero podría argüirse que una definición que no realice algunas distinciones más sutiles permitiría incluir episodios de razonamiento hipotético que, aunque contienen supuestos contrafácticos que motivan el análisis de una hipótesis, arriban a conclusiones a través de premisas y argumentos que son independientes de dichos supuestos. Lo mismo ocurre con el carácter de irrealizables de estos, ya que algunos experimentos mentales fueron posteriormente ejecutados como experimentos reales. Asimismo, la facilidad con que permiten derivar conclusiones también es discutible ya que, al menos en algunos casos, la eficacia del experimento depende del conocimiento que el experimentador tenga del marco teórico en el cual se presenta. No obstante, estos cuatro elementos (permiten derivar conclusiones acerca del mundo físico, carecen de realización material, poseen cierto grado de eficacia heurística, incluyen supuestos contrafácticos e idealizaciones) parecen estar presentes en un número suficientemente grande de casos y aparecen asociados a algunos de los análisis que nos interesa desarrollar. Es posible, entonces, elaborar sobre esta base una lista preliminar de rasgos comunes.

En primer lugar, es incuestionable que estos experimentos se comunican a través de una narrativa similar a la descripción de un experimento real. Un punto importante del análisis es la semejanza entre experimentos mentales y reales, por lo que esta hipótesis puede resultar útil para explorar este vínculo. Este análisis permitirá explicar asimismo el hecho de que algunos experimentos mentales hayan sido interpretados desde diferentes perspectivas teóricas, en ocasiones desde perspectivas rivales. En segundo lugar, es evidente que los experimentos mentales se realizan solo en la imaginación. Parece necesario explorar si este rasgo implica que son irrealizables en algún sentido interesante del término. También parece ser un hecho que los experimentos mentales contienen supuestos contrafácticos e idealizaciones. Evaluar si este es un rasgo esencial de los experimentos mentales y si determina algún parentesco con los modelos mecánicos y computacionales también es un punto importante para el análisis

prospectivo. Asimismo, todos los ejemplos presentados tienen, en cada caso, un objetivo cognitivo específico: ilustrar teorías, refutar hipótesis, proporcionar evidencia a favor, derivar consecuencias absurdas de un principio, mostrar su plausibilidad. Valorar de qué manera cumplen con estos objetivos permitiría ofrecer una explicación completa de sus funciones. Además, evaluar los poderes epistémicos de los experimentos mentales que parecen desempeñar algún papel en la selección de teorías permitiría extraer algunas conclusiones acerca de su función en el cambio conceptual.

La definición que se ofrece a continuación parece cumplir con los *desiderata* para los cuales fue formulada:

“La ejecución de un experimento mental en física consiste, en términos generales, en la representación de un escenario posible y en la manipulación imaginaria de ciertas variables. Este ejercicio tiene lugar dentro de un marco nomológico y procede a partir de la postulación de una situación hipotética o contrafáctica y la descripción de un conjunto de condiciones iniciales que se comunica a través de una narrativa. En ocasiones esta descripción también incluye un relato acerca del funcionamiento de un aparato o instrumento y es acompañada de una representación visual o un gráfico del dispositivo experimental. Su ejecución no requiere de instrumentos y se lleva a cabo enteramente en la imaginación. Su realización hace posible, en principio, extraer consecuencias acerca del comportamiento de los fenómenos físicos pero en cada caso la realización del experimento tiene un objetivo cognitivo determinado: ilustrar una teoría, refutar una hipótesis, proporcionar evidencia a favor de una afirmación, derivar consecuencias absurdas de un principio o mostrar su plausibilidad”.

1.7 Conclusiones

El uso de los experimentos mentales en la física moderna atestigua su empleo en el contexto de justificación de las teorías. Los mismos han sido ampliamente utilizados en la física teórica y la cosmología y desempeñaron un papel importante en la obra de Galileo, Newton y Einstein. El resultado de muchos de estos experimentos ha sido el abandono de una teoría o el establecimiento de una nueva. Debido a que su ejecución no involucra la realización material del sistema experimental, sus resultados no pueden atribuirse a la incorporación de nueva información empírica. No obstante, parecen proporcionar alguna forma de conocimiento acerca del mundo natural. Determinar las fuentes de este conocimiento constituye uno de los puntos centrales de la controversia epistemológica. Las interpretaciones iniciales dieron lugar a tres líneas de análisis: una perspectiva empirista representada por Norton (1991, 1996, 2004a, 2004b, 2013) de acuerdo con la cual la experiencia acumulada y la evolución de las habilidades especulativas son el fundamento del conocimiento del mundo natural y permiten justificar teorías y anticipar experimentos reales; una interpretación *apriorista* representada por Brown (1991, 2004a, 2004b, 2011, 2013, 2017) de acuerdo con la cual los experimentos mentales suministran modelos ideales para la interpretación de fenómenos y una tercera representada principalmente por Gendler (2000, 2004, 2013), Miščević (1992) y Neressian (1992 y 2007), también de base empirista, según la cual los experimentos mentales son incapaces de proporcionar evidencia a favor de hipótesis científicas (a lo sumo pueden ilustrar teorías o mostrar inconsistencias en una teoría) y que su alcance debe ser restringido al de los argumentos teóricos. Básicamente, las distintas perspectivas vigentes entienden que: a) los experimentos mentales son experimentos esencialmente similares a los reales, b) son argumentos teóricos c) son modelos mentales.

En el contexto de la filosofía de la experimentación los experimentos mentales fueron deliberadamente ignorados. Por lo general, se ha admitido que la falta de intervención en el mundo natural es una razón suficiente para asimilarlos a argumentos teóricos. No obstante, algunos han aducido que los experimentos mentales exhiben rasgos que demandan de una explicación ulterior, a saber: su valor heurístico, la posibilidad de ser interpretados a partir de marcos teóricos diferentes, el papel activo que en ellos desempeña el experimentador y su similitud con los modelos y simulaciones computacionales.

Las definiciones usuales de la noción de experimentos mentales fallan en incluir casos que son considerados paradigmáticos, excluyen algunas funciones importantes o no ponen de relevancia aspectos del funcionamiento de estos que resultan fundamentales para la explicación de sus poderes epistémicos. La definición que hemos propuesto se centra en los rasgos compartidos por los ejemplares presentados y en algunos problemas filosóficos que no han sido suficientemente tematizados. Algunas de las preguntas que se intentaran responder a partir de esta definición preliminar son las siguientes: ¿en qué se funda el aspecto experimental de estos dispositivos?, ¿qué funciones cumplen en estos las visualizaciones?, ¿qué tipo de fenómenos tienen lugar en los experimentos mentales?, ¿qué papel desempeñan las suposiciones contrafácticas?, ¿qué papel desempeñan las idealizaciones?, ¿qué rasgos comparten los experimentos mentales efectivos?, ¿qué tipo de razonamiento implica la derivación de conclusiones a partir de la realización de un experimento mental?, ¿qué funciones legítimas pueden desempeñar? . Este trabajo aspira proporcionar una explicación lo más completa posible del funcionamiento de los experimentos mentales en las ciencias físicas. Esto permitiría, por un lado, inscribir el examen de estas prácticas en el de teorizaciones más generales referentes a la experimentación y la argumentación científica, y, por otro, haría posible elucidar los casos históricamente interesantes a partir de su incorporación a modelos comprensivos que expliquen su funcionamiento y que a la vez permitan

explicitar una serie de supuestos metafísicos y epistémicos contextualmente definibles. La amplitud de la definición preliminar que hemos propuesto pone de manifiesto todos los aspectos que un estudio sistemático de los experimentos mentales debería incluir: los elementos heurísticos de su producción y comunicación, el objetivo cognitivo de su elaboración, el mecanismo cognitivo en virtud del cual producen representaciones, su relación con inferencias lógicas, su vínculo con hipótesis y teorías, los criterios de identidad y los criterios de confiabilidad que permitan establecer sus usos legítimos.

CAPÍTULO 2

Experimentos mentales y experimentos reales. Analogías positivas y negativas

2.1 Introducción

Antes de adoptar un punto de vista acerca de la naturaleza de los experimentos mentales es necesario aclarar en qué se funda su aspecto experimental. Algunos filósofos han argumentado que, aunque es posible establecer analogías entre los experimentos reales y los experimentos mentales, la ausencia de intervención en el mundo material priva a la experimentación mental de la fuente más importante de conocimiento objetivo acerca del mundo natural y, en consecuencia, no puede afirmarse que los experimentos mentales sean experimentos más que en un sentido metafórico. Otros, en cambio, han sostenido que los experimentos mentales son esencialmente el mismo tipo de procedimiento que los experimentos reales, puesto que ambos comparten muchas características.

Las elaboraciones filosóficas desarrolladas en los últimos años acerca de la epistemología de las prácticas experimentales han argumentado, entre otras cosas, que los resultados experimentales son autónomos, que cumplen múltiples funciones en relación con las teorías y que gozan de cierta estabilidad frente al cambio teórico. Desde este punto de vista, subsiste un acuerdo generalizado según el cual la objetividad de los resultados experimentales se funda en la corroboración independiente y en la posibilidad de establecer cadenas de interacción causal entre el observador y el objeto bajo investigación. Aunque se admite que los fenómenos que tienen lugar en el laboratorio son producto de una

actividad constructiva que involucra el conocimiento teórico y práctico del científico, la intervención material es el fundamento último de la objetividad científica. Debido a que en los experimentos mentales no es posible establecer cadenas de interacción causal ni aplicar las mismas estrategias de validación epistémica que emplean los experimentos reales, por lo general, las posiciones filosóficas ligadas a la epistemología de la experimentación han negado que los experimentos mentales sean de hecho una clase de experimentos.

El enfoque de acuerdo con el cual los experimentos mentales son una clase de experimentos fue propuesto inicialmente por Mach. Dentro de esta perspectiva, es posible distinguir dos tesis alternativas acerca de la naturaleza de los experimentos mentales: a) el punto de vista de acuerdo con el cual los experimentos mentales son análogos estructurales de los experimentos ordinarios que habilitan una vía privilegiada de conocimiento del mundo natural tal como afirma Brown (1991 y 2011); y b) la perspectiva que sostiene que los experimentos mentales constituyen casos límite de la experimentación real, como sostienen Sorensen (1992), Gooding (1992) y Häggqvist (1996). En general, son tres los aspectos que estos autores señalan que experimentos mentales y reales comparten: 1) su carácter visual y el papel activo del experimentador; 2) la idea de que la actividad fundamental de la experimentación es la introducción de variaciones en una situación controlada; y 3) la idea de que los experimentos pueden cumplir múltiples funciones y son relativamente independientes de la teoría.

El objetivo de este capítulo es establecer los alcances de la interpretación de acuerdo con la cual los experimentos mentales son propiamente experimentos. Se intenta establecer los límites de esta analogía con el fin de determinar en qué podría fundarse su valor evidencial y cuál podría ser su papel en el contexto de justificación de las teorías físicas. Partimos aquí del supuesto de que la interpretación filosófica de los experimentos reales puede resultar esclarecedora en la investigación de la naturaleza y las funciones de los experimentos mentales.

Las posiciones que interpretan a los experimentos mentales como siendo esencialmente experimentos asumen que, dado que es posible identificar en los experimentos mentales mecanismos de generación de creencias análogos a los que tienen lugar en los experimentos ordinarios, los primeros tienen la capacidad de justificar hipótesis de la misma manera que estos últimos. Se busca establecer si el tipo de variación que es posible realizar en un escenario imaginario, el carácter visual del razonamiento y el papel que el experimentador desempeña en los experimentos mentales son aspectos comunes relevantes para argumentar que existe una continuidad entre ambas prácticas. Con este fin, se recupera la conceptualización de la observación científica y la noción de fenómeno desarrolladas en el contexto de la filosofía de las prácticas experimentales. Por otra parte, diversos filósofos han defendido la robustez de los resultados experimentales frente al cambio teórico y la independencia de estas prácticas respecto de la teoría. Se reformula esta tesis y se examinan sus consecuencias para el caso de los experimentos mentales. Luego, se explicitan los argumentos empleados para sostener que los experimentos mentales son suficientemente similares a los experimentos reales. Se discute, entonces, si alguna de las versiones de la tesis de la autonomía experimental podría extenderse a los experimentos mentales.

2.2 Observación e intervención en el mundo natural

Una de las tesis principales que comparten los análisis filosóficos de las prácticas experimentales es la que sostiene que la observación proporciona una base objetiva para el conocimiento. Esta hipótesis está conectada con otras de carácter fundamental para la conceptualización de la experimentación que sostiene que los resultados experimentales son autónomos respecto de la teoría. Aunque en los experimentos mentales no se realizan observaciones, algunos autores sostienen que los procesos de representación visual que en éstos tienen

lugar pueden cumplir una función análoga a la desempeñada por la observación. Para decidir si en algún sentido es viable sostener que las representaciones visuales pueden desempeñar un papel análogo al de la observación y constituir una base objetiva de conocimiento en los experimentos mentales, resulta preciso inspeccionar, al menos de manera sinóptica, las teorías disponibles acerca de cómo los datos del mundo físico son registrados, organizados e interpretados para proporcionar información confiable acerca del mundo natural en los procesos de observación científica.

El tratamiento de la noción de observación en la filosofía de la ciencia tomó diferentes matices de acuerdo con las perspectivas dentro las cuales fue conceptualizada. Dentro de la concepción clásica de las teorías, la distinción entre lenguaje teórico y observacional resultó fundamental para esclarecer la cuestión de la significatividad de los enunciados. Las críticas que apuntaron las dificultades para establecer distinciones tajantes entre dos clases de vocabulario y las que pusieron en evidencia la carga teórica de toda observación, socavaron los fundamentos para realizar tal distinción. Los trabajos centrados en las prácticas experimentales recuperaron en parte estas críticas pero se interesaron en restituir el valor objetivo de la observación.

Algunas de las conceptualizaciones de las nociones de observación y observabilidad en ciencias naturales más extensamente aceptadas en este contexto interpretan a la observación en términos de interacción causal entre objetos externos y el aparato perceptual de los observadores. Estas perspectivas, por lo general, admiten que este proceso está mediado por el instrumental científico y asumen que la observación está teóricamente cargada, señalando que las percepciones epistémicamente significativas dependen de la identificación de los ítems relevantes en términos de un marco conceptual disponible. El carácter objetivo de la observación reside en que las propiedades intrínsecas de los objetos físicos limitan las características de la interacción.

Dudley Shapere (1982) fue el primero en vincular la noción de interacción física con la de observación científica. De acuerdo con este punto de vista, la observación es una subespecie de interacción entre una entidad "X"⁹ y un receptor, en la que cierta información de "X" es transferida al receptor. La observación es una función del estado de conocimiento prevaleciente y puede cambiar conforme cambia dicho conocimiento. El conocimiento físico prevaleciente especifica qué cuenta como "un receptor apropiado" y qué cuenta como "información". Asimismo clasifica los tipos de información que existen, las formas en las cuales se transmite y se recibe la información de diversos tipos, el carácter y los tipos de interferencia y las circunstancias y la frecuencia en la que dichas interacciones ocurren (Shapere, 1982: 488). Esta interpretación es la que prevalece con algunos matices en la filosofía de la ciencia contemporánea y se encuentra referida numerosas veces en la conceptualización de las prácticas experimentales.

Ian Hacking, por su parte, distinguió explícitamente a la experimentación de la observación (Hacking, 1983: 173). En la experimentación los científicos intervienen activamente en los objetos y sistemas naturales que están investigando. El carácter intervencionista de la experimentación engendra problemas epistemológicos particulares, diferentes de los que plantea la observación. No obstante, afirmó, estos dos aspectos de la práctica científica, mantienen relaciones cercanas. En particular, Hacking apoya la robustez de las observaciones a través de los cambios en las teorías de los aparatos y de los fenómenos, en la manipulación de los objetos bajo observación. Advertir los resultados que se siguen de la intervención sobre el objeto bajo observación incrementa nuestra confianza en el buen funcionamiento del dispositivo

⁹ Para Shapere la especificación del proceso de observación consta de tres partes: i) la teoría de la emisión de información desde una fuente o entidad "X", ii) la transmisión de la información desde la fuente al receptor y iii) la teoría del receptor. En su artículo examina la detección de neutrinos como un caso de observación directa. Las partes de este proceso incluyen: i) teoría acerca de la producción de energía en el interior de las estrellas, ii) la teoría acerca de cómo los neutrinos viajan del sol a la tierra, cómo interactúan con otras partículas y las diferencias entre lo que cuenta como interferencia y como interacción y iii) la teoría del aparato experimental para captar neutrinos.

experimental y en el resultado mismo de la observación. Asimismo, en tanto las teorías subyacentes al mecanismo de detección no estén directamente vinculadas con el objeto de observación, éste puede considerarse teóricamente independiente. Las teorías acerca del funcionamiento del microscopio, por ejemplo, se han modificado drásticamente y la capacidad de este tipo de instrumental se ha incrementado considerablemente, pero a pesar de todos estos cambios los resultados permanecen estables (Hacking, 1983: 191).

Entre los estudios sistemáticos sobre la noción de observación en la física, encontramos tres trabajos de relevancia. El primero es el de Harold Brown (1987), quien desarrolló una versión de la teoría causal representativa de la percepción; el segundo es el de Peter Kosso (1989), quien elaboró una interpretación interaccional- informacional de la observación; y el más reciente es el Hans Radder (2006), quien analizó la naturaleza de la contribución conceptual en el proceso de observación científica. Uno de los puntos fundamentales que estas perspectivas comparten es la idea de acuerdo con la cual los *ítems* que percibimos son efectos de la interacción causal entre algún tipo de receptor (que puede involucrar tanto al aparato perceptual humano como al instrumental científico) y el mundo físico.

El punto fundamental de la perspectiva de Brown (1987) es la defensa de la objetividad de la observación. De acuerdo con esta interpretación, la observación es un proceso en el que se obtiene información acerca de algún aspecto mundo examinando un *ítem*¹⁰ que es asequible a la percepción, para lo cual es preciso contar con el conocimiento de fondo pertinente para reconstruir la cadena causal que va desde el *ítem* perceptible hasta el objeto del mundo que lo produce (Brown, 1987: 93). La observación es capaz de justificar teorías porque, aunque está teóricamente cargada, siempre involucra una interacción con el mundo físico.

¹⁰ "*Ítem*" es utilizado por Brown como término ontológicamente neutral para designar entidades, procesos y en general, cualquier cosa del mundo externo (Ver Brown, 1987: 3)

Cualquier información que recibimos del mundo a nuestro alrededor pasa a través de nuestros sentidos -incluso si es a través de una impresión o una computadora- y todos nuestros instrumentos de observación eventualmente producen *outputs* que podemos percibir. Estos *outputs* en conjunción con las teorías disponibles y en conjunción con información acerca de las propiedades de nuestros instrumentos producen lo que en la ciencia moderna cuenta como observaciones. (Brown, 1987: 48-49)

Así, por ejemplo, las observaciones realizadas por intermedio de fotografías de Rayos X o a través de un telescopio permiten obtener resultados estables, que proporcionan una fuente superior de información a la que se puede obtener solo mediante los órganos perceptivos en tanto permiten, por ejemplo, agrandar o medir el *ítem* resultado de la interacción. Asimismo, la semejanza entre el *ítem* percibido y el *ítem* bajo observación no es para Brown una condición para legitimar información acerca del *ítem* observado. Lo que importa en casos con estas características, como el de la cristalografía de rayos X, es que los *ítems* perceptibles contengan información sobre el *ítem* bajo observación que pueda ser decodificada a partir del conocimiento teórico disponible. Por otra parte, Brown desestima uno de los rasgos que tradicionalmente se le atribuyó a las creencias observacionales. Las creencias formadas en los procesos de observación científica no son básicas (en el sentido de no inferenciales) ya que se requiere una cantidad considerable de teoría para decodificar la información relevante en el *ítem* observado:

La inferencia consciente es una de las maneras en las que aquello que el proceso de ilación ocurre y no hay razones epistemológicas sólidas para distinguir los casos en los cuales el proceso de ilación es intuitivo y rápido de aquellos en los que es explícito e insume tiempo. (Brown, 1987: 74)

El trabajo de Kosso (1989) también apela a un esquema causal-interaccional para explicar los procesos de observación científica y proporcionar una noción de observabilidad acorde con la práctica científica. De acuerdo con su perspectiva, el par ordenado <objeto x , propiedad P > es "observado" en tanto hay una interacción o una cadena de interacciones entre " x " y un aparato tal que permita transmitir la información " x es P " a un observador. Las teorías y las leyes

físicas especifican cuáles son las interacciones físicas posibles, cuándo pueden ocurrir y en qué consisten. Una cadena de interacción consiste en una secuencia temporal de interacciones en la cual cada antecesor comparte un elemento con su sucesor. Esta noción permite distinguir grados de observabilidad de acuerdo con diferentes criterios referidos a las características del objeto, a la cantidad de teoría necesaria y a la distancia entre el receptor y el objeto en cuestión. Asimismo, al recurrir a las leyes físicas para describir los procesos de observabilidad, la interpretación interaccional-informacional proporciona un criterio para evaluar la fiabilidad de las inferencias a una causa "x" como fuente de información. Según este criterio, una inferencia es concluyente si para explicar la existencia de "x" no emplea una teoría acerca de sus propiedades específicas o sobre la existencia de "x".

Recientemente, Radder (2006) introdujo otros elementos importantes para elucidar la noción de observación científica. De acuerdo con su perspectiva, cualquier proceso observacional es *materialmente realizado* y *conceptualmente interpretado* desde el inicio. La observación científica es una actividad que explota potencialidades naturales y recursos socioculturales disponibles. Que la observación es un proceso *conceptualmente interpretado* implica que la percepción no está completamente determinada por el estímulo visual sino que depende de características específicas de los observadores y del proceso observacional. Asimismo, la interpretación conceptual del proceso observacional no es un logro exclusivo del observador individual ya que tanto los procedimientos de observación como los conceptos empleados para interpretarla son aprendidos en el seno de una comunidad sociocultural. Esta comunidad proporciona la fuente de interpretación conceptual y el criterio de la observación futura. La noción de *realización material* se refiere a las actividades que los agentes humanos llevan a cabo para materializar el proceso de observación científica. Estas incluyen la ejecución de diferentes tipos de acciones, la puesta en práctica de habilidades y la aplicación de esquemas conceptuales internalizados.

Como es posible apreciar a partir de este breve recorrido por algunas conceptualizaciones recientes, las nociones de observación y observabilidad acordes con los desarrollos teóricos centrados en las prácticas experimentales se apoyan en el concepto de interacción física, amplían la noción de perceptor antes restringida al científico como sujeto de conocimiento, destacan que estos procesos conllevan un alto grado de interpretación teórica y señalan la inseparabilidad entre las características de la percepción y la intervención (culturalmente determinada) del científico que lleva a cabo este proceso. Aunque desde este punto de vista la percepción ordinaria pierde un lugar de privilegio y se reconoce la importancia de la carga teórica en la interpretación de las observaciones, el vínculo entre el perceptor y el ítem perceptible es fundamental para sustentar la objetividad, ya que se lo considera como indicio de un *origen* de la cadena de interacción causal en un objeto del mundo externo.

2.3 Fenómenos y datos

La noción de fenómeno resulta particularmente relevante en el desarrollo de una explicación de la naturaleza y funciones de los experimentos mentales porque, de acuerdo con algunos autores, en el tipo de fenómeno que estos experimentos son capaces de presentar reside su poder probatorio. Según algunos filósofos, en los experimentos mentales podemos representarnos fenómenos idealizados o abstractos. A través de estos fenómenos tenemos acceso directo a principios teóricos o leyes. Interesa revisar de qué manera son conceptualizados los fenómenos en el marco de la filosofía de las prácticas experimentales y evaluar si la noción de fenómeno a la que apelan algunas explicaciones de los poderes epistémicos de los experimentos mentales es plausible.

La crítica de la imagen tradicional de la ciencia y la proliferación de estudios filosóficos acerca de la experimentación derivaron en la revisión de algunas de las

nociones fundamentales empleadas para explicar las relaciones entre teoría y observación. Este es el caso de noción de fenómeno, que Bogen y Woodward (1988) elaboraron junto con una contundente crítica a la noción tradicional de evidencia y la explicitación de una distinción entre datos y fenómenos a partir del estudio de casos.

Es posible rastrear una primera revisión y reformulación de la noción de fenómeno dentro de la filosofía de la experimentación en la obra de Hacking (1983). De acuerdo con Hacking, un fenómeno es un evento o proceso de cierto tipo que ocurre regularmente bajo circunstancias definidas. Cuando conocemos la regularidad exhibida por el fenómeno, la expresamos mediante un enunciado legaliforme. No obstante, no todos los fenómenos consisten en conjunciones constantes preexistentes en la naturaleza. Esta es, según Hacking, una representación propia de la concepción filosófica dominada por la teoría de acuerdo con la cual:

Los fenómenos revelados en el laboratorio son producto de la obra de Dios, esperando a ser descubiertos. Tal actitud es natural en una filosofía dominada por la teoría, formulamos teorías acerca del mundo, conjeturamos leyes de la naturaleza. Los fenómenos son regularidades consecuencia de esas leyes. (Hacking, 1983: 222)

Algunos fenómenos no existen fuera de ciertos tipos de aparatos, independientemente de cierta disposición experimental, sino que son creados en el laboratorio, a partir de la generación de condiciones específicas. De acuerdo con Hacking, así como somos capaces de analizar la complejidad presente en la naturaleza y las diferentes leyes, presentando en el laboratorio el fenómeno puro, aislado, podemos concebir nuevos fenómenos, que no existen en la naturaleza.

De acuerdo con Bogen y Woodward (1983), los fenómenos son efectos o procesos que son potenciales objetos de predicción o explicación sistemática por parte de las teorías generales y que pueden servir de evidencia para éstas. Por ejemplo, la deflexión de la luz estelar al pasar cerca del Sol medida por Eddington

en el eclipse de 1919 es un fenómeno predicho y explicado por la relatividad general. Los datos, sin embargo, se diferencian de los fenómenos. Los datos son registros públicos (en el ejemplo, fotografías de las posiciones de las estrellas) producidos por mediciones y experimentos, que sirven como evidencia de la existencia del fenómeno o de que este posee ciertas características. Los datos guardan una relación causal con los fenómenos pero reflejan, además, aspectos idiosincráticos de las operaciones de medición y experimentación de las que son el resultado. En palabras de Bogen y Woodward:

Los factores que intervienen en la producción de datos a menudo serán tan dispares y numerosos, y los detalles de sus interacciones serán tan complejos, que no será posible construir una teoría que nos permita predecir su aparición o rastrear en detalle cómo se combinan para producir elementos de datos particulares. Los fenómenos, por el contrario, no son idiosincráticos a contextos experimentales específicos. Esperamos que los fenómenos tengan características estables y repetibles que sean detectables por medio de una variedad de procedimientos diferentes, que pueden producir tipos de datos bastante diferentes. (Bogen y Woodward, 1988: 371)

Según esta perspectiva, el examen filosófico de la observación y la experimentación científica revela que la relación de justificación entre teoría y observación no es deductiva, como se interpretó tradicionalmente. Debido a que el proceso de medición y experimentación es complejo e idiosincrático y, además, involucra muchos tipos diferentes de factores causales, es difícil explicar sus aspectos detallados a partir de una teoría general y ciertas condiciones iniciales, es decir, deductivamente. La dirección de la inferencia, en muchos casos, es inductiva, desde los datos a algún aspecto del fenómeno.

En síntesis, los análisis filosóficos centrados en las prácticas experimentales reformularon la noción tradicional de fenómeno. Según estos enfoques, los fenómenos no son entidades preexistentes en la naturaleza, sino que son el producto de la actividad científica. Esta concepción constructivista de los fenómenos, supone no obstante, que existe una relación causal entre los datos y los patrones que a partir de estos los científicos pueden reconocer.

2.4 La vida propia de los experimentos

El rasgo más conspicuo de las prácticas experimentales es la intervención en el mundo natural. Aunque este ha sido reconocido como el elemento definitorio de la experimentación, dentro del procedimiento experimental los filósofos han diferenciado etapas o elementos. Hacking (1992) identificó diferentes elementos en la experimentación, a los que clasificó en tres grupos: *ideas*, *cosas* y *marcas*. Llamó "ideas" a los componentes intelectuales de un experimento, como las preguntas, el conocimiento de fondo presupuesto, la teoría sistemática, el modelado del aparato experimental y las hipótesis tópicas. Denominó "materiales" a la substancia material de la investigación que los investigadores llaman "cosas"; entre éstas pueden encontrarse, por ejemplo, los instrumentos, las ratas de Noruega y los electrones polarizados. En general, esta categoría está constituida por objetos, fuentes de modificación, herramientas, detectores y generadores de datos. Finalmente, denominó "marcas" a los resultados de un experimento, entre estos: los datos, la reducción de datos, la evaluación de datos, el análisis de datos, y la interpretación (Hacking, 1992: 243-249).

Por su parte, para Radder (2003) el proceso experimental incluye el objeto de estudio, el aparato y su interacción. En cada una de estas fases es posible distinguir la realización material de la interpretación teórica. La experimentación implica la realización material del sistema experimental (es decir, el objeto de estudio, el aparato y su interacción) así como la intervención activa en el medio ambiente de ese sistema. En este sentido, el experimento contrasta con la teoría incluso si el trabajo teórico es siempre acompañado de actos materiales como escribir una fórmula (Radder, 1996: 11). Por otra parte, no cualquier tipo de intervención en el mundo cuenta como un experimento: generalmente, los experimentos requieren por lo menos de cierta estabilidad y reproducibilidad. Cumplir con este requisito presupone un control del sistema experimental y su

medio ambiente, así como cierta disciplina del experimentador y de las otras personas involucradas en la realización del experimento.

Una consecuencia obvia de esta posición es que los denominados “experimentos mentales” no reúnen las características propias de los experimentos. Por una parte, no involucran ninguno de los elementos que corresponden al aparato experimental, en la denominación de Hacking, o a la realización material, en términos de Radder. Por otra parte, los experimentos mentales están en principio, más determinados por la teoría que los experimentos reales. Para argumentar este punto Hacking se remite directamente a Bacon en su caracterización de los experimentos. Señala que la experimentación consiste en manipular el mundo para conocer sus secretos, no solo en observar sino en “torcerle la cola al león” (Hacking, 1983: 149). De acuerdo con este punto de vista, los experimentos mentales se diferencian de los experimentos ordinarios en un sentido importante: “no tienen vida propia” (Hacking, 1992: 302). Aunque son capaces de exponer tensiones entre marcos conceptuales, característica que explica su extendido uso, a diferencia de los experimentos reales, son incapaces de evolucionar y se encuentran fijados al marco teórico y al contexto dentro del cual son propuestos. Según Hacking:

Los experimentos mentales se parecen a los chistes o a las ilusiones ópticas. Son experiencias que son difíciles o imposibles de exorcizar. Son como las paradojas que siempre te atrapan no importa que tan preparado estés. (Hacking, 1992: 307)

La ausencia de intervención en el mundo natural tiene como consecuencia que los resultados de los experimentos mentales sean teórica y contextualmente dependientes. Los experimentos ordinarios tienen vida propia, para Hacking, porque evolucionan, son orgánicos, estables, maduran y pueden ser replicados, reciclados y reequipados o reorganizados a partir del empleo de mejores instrumentos. Pero, a diferencia de los reales, los experimentos mentales son fijos e inmutables. Son similares a las pruebas matemáticas, aunque se diferencian de éstas en que las mismas contienen ideas que pueden ser usadas de nuevo en

nuevos contextos. Los experimentos mentales permanecen fijos al contexto en el que aparecen. Una vez que han sido escritos y publicados, son como íconos, no tienen vida propia.

En el mismo espíritu pero de manera menos taxativa, Radder (2009) señala que los experimentos mentales se conciben mejor si no se los considera como experimentos sino como tipos particulares de argumentos teóricos, que pueden ser o no realizables en la práctica experimental.

2.5 Visualizaciones y manipulación de escenarios imaginarios

Las perspectivas que interpretan a los experimentos mentales como una especie de experimentos deben explicar de qué manera estos dispositivos proporcionan conocimiento objetivo del mundo natural en ausencia de observaciones que introduzcan nuevos datos y sin la ejecución del proceso experimental. Una de las maneras en que los representantes de esta interpretación han explicado el mecanismo de conocimiento en los experimentos mentales es apelando a las operaciones de visualización que tienen lugar en ellos. De acuerdo con este punto de vista, la representación visual desempeña un papel central en la formación de creencias acerca del mundo físico. Este papel es análogo al de la percepción en la observación de fenómenos físicos. Por otra parte, quienes comparan los experimentos mentales con los reales, argumentan que la manipulación del escenario imaginario, es decir, la variación de elementos en una situación imaginada, permite revelar rasgos del mundo natural de forma similar a como la manipulación de variables en una situación controlada lo hace en los experimentos reales.

La tesis de que los experimentos mentales y reales son especies de un mismo género de prácticas tiene dos variantes. De acuerdo con la primera, los experimentos mentales son análogos estructurales de los experimentos

ordinarios que habilitan una vía privilegiada de conocimiento del mundo natural. De acuerdo con la segunda, los experimentos mentales constituyen casos límite de la experimentación real.

Los argumentos de la primera variante de la perspectiva “experimental”¹¹ enfatizan la idea de que en los experimentos mentales tiene lugar alguna clase de manipulación del escenario representado que posibilita la captación de fenómenos puros. En los experimentos mentales tiene lugar la visualización de fenómenos abstractos o ideales. De acuerdo con Brown:

Los experimentos mentales tratan con fenómenos. Obviamente no manipulan datos empíricos -esto es verdadero por definición. Pero el hecho de que involucren un proceso pictórico sugiere que es necesario mantener algo de la observación como involucrado centralmente cuando tratamos de entenderlos. (Brown, 1994: 131)

Lo que se preserva de la observación en los experimentos mentales está vinculado, según Brown, con los aspectos heurísticos del dispositivo y con el mecanismo de conocimiento en cuestión. La naturaleza visual de la representación facilita la comprensión del fenómeno en cuestión. Las representaciones gráficas que acompañan en muchas ocasiones a la narración del experimento, proporcionan una imagen del fenómeno ideal que describen. El aspecto visualizable del fenómeno representado contribuye a la inmediatez con la que los experimentadores derivan un resultado en esta clase de experimentos. Así, por ejemplo en el experimento del balde giratorio de Newton ([1687] 1974) el fenómeno es visualizable en el siguiente sentido: en el primer estadio la superficie del agua permanece plana porque se encuentra en reposo respecto del espacio absoluto, en el tercero la superficie del agua se vuelve cóncava porque está rotando respecto del espacio absoluto.

¹¹ Emplearemos la denominación “perspectiva experimental” para referirnos a la posición que asimila los experimentos mentales a los reales.

En cuanto al tipo de conocimiento que los experimentos mentales proporcionan, el fenómeno es una abstracción, que de acuerdo con Brown, es de una naturaleza diferente de la de los datos empíricos y desempeña un papel específico en el razonamiento científico, ya que permite intuir el fenómeno en cuestión. Esta interpretación está conectada con su perspectiva acerca de la función de los diagramas y los procesos de visualización en matemáticas. Tradicionalmente, se ha interpretado a esta clase de estrategias como ayudas psicológicas o didácticas para comprender una prueba, diferentes de las pruebas mismas que consisten en derivaciones expresadas en lenguaje verbal o simbólico. Brown sostiene que los esquemas y diagramas pueden funcionar ellos mismos como pruebas. En el contexto de las matemáticas, donde admite la posibilidad de conocer verdades intuitivamente, las representaciones visuales de este tipo funcionan como intermediarios que nos permiten vislumbrar certezas:

Algunas 'imágenes' no son realmente imágenes, son más bien ventanas al cielo platónico (...) como los telescopios asisten al ojo desnudo, algunos diagramas son instrumentos que asisten al ojo de la mente. (Brown, 2008: 40)

Las visualizaciones de los fenómenos en los experimentos mentales en física, cumplen la misma función que los diagramas en matemática, esto es, permiten confirmar teorías directamente. Las clases de fenómenos con las que tratan algunos experimentos mentales, permiten acceder directamente a leyes naturales sin requerir de la mediación de una inferencia ni de la presentación de datos empíricos. Esto se justifica porque, de acuerdo con Brown, hay una facultad que nos permite formarnos creencias sobre la naturaleza independientemente de la inspección de casos de la experiencia; esta facultad es análoga a la percepción sensible que proporciona evidencia en los experimentos ordinarios. Los experimentos mentales son una vía de acceso a esta facultad extraordinaria.

Brown señala que las representaciones visuales en los experimentos mentales son la contraparte de la observación en los experimentos reales. El argumento de Brown, es en líneas generales, el siguiente: 1) los datos son objeto

de la observación o la experiencia; 2) los fenómenos son entidades abstractas, que conservan un carácter fuertemente visual (Brown, 1994: 140), son clases naturales o patrones que pueden ser visualizados porque guardan una relación de semejanza con los datos; 3) al igual que las observaciones, las visualizaciones pueden desempeñar un papel evidencial. Brown concluye que podemos justificar teorías en virtud de situaciones visualizables o de imágenes de los fenómenos.

El otro rasgo que Brown vincula con el carácter experimental de los experimentos mentales es que los escenarios imaginados involucran, además, una cierta “manipulación” de objetos. Así, por ejemplo, unir en la imaginación objetos de distinto peso, como en el experimento de los cuerpos en caída libre de Galileo, es una forma de manipulación. Para Brown el experimentador lejos de ser un observador pasivo, es doblemente activo en la ejecución de un experimento mental: por un lado, porque su realización implica representarse o imaginar una situación, y, por otro lado, porque implica imaginarse a sí mismo interviniendo activamente en tal escenario (Brown, 2011: 18).

Aunque Brown apela a un sofisticado esquema para dar cuenta del proceso de adquisición de conocimiento para el caso de los experimentos mentales, consideramos que esta explicación no es clara. En primer lugar, al menos en el marco de una epistemología platonista, que es en la que Brown enmarca el funcionamiento de los experimentos mentales, la visualización no cumple ningún papel en la captación de entidades abstractas. El tipo de operación referida por Brown es equivalente, en cierto sentido, a la *diánoia* platónica, la intelección directa de entidades abstractas. En la epistemología platónica la *diánoia* posibilita el salto desde el mundo sensible al reino de las entidades abstractas. En la epistemología de Brown, el tipo de visualización que tiene lugar en los experimentos mentales permite al experimentador *ver* una clase de fenómeno que ejemplifica una ley. Este fenómeno constituye una instancia prototípica de la ley natural en cuestión (Brown, 1994: 98). De manera que, en la perspectiva de Brown, la visualización es un paso previo al conocimiento de las leyes naturales.

Nuestra apreciación es que aquí Brown combina arbitrariamente elementos de la tradición platónica, que resultan útiles en este marco para explicar el conocimiento en matemáticas, con su propia concepción del conocimiento *a priori* en matemáticas. El mecanismo de conocimiento invocado, la intelección directa o intuición racional no requiere de la visualización como intermediario. Incluso si pudiera defenderse la utilidad heurística de la visualización para el tipo de conocimiento *a priori*, éste no se funda en la relación del sujeto que conoce con esta representación sino con el objeto abstracto que contempla.

Otra afirmación problemática de Brown es que la visualización es evidencia del principio en cuestión, de una manera parecida a cómo los diagramas pueden ocupar el lugar de las pruebas formales. Podría suponerse que, como representación idealizada, la visualización guarda algún tipo de relación con todas las instancias de la ley o principio en cuestión y en esto reside su capacidad de servir como evidencia de la ley. Pero Brown no dice prácticamente nada acerca de la naturaleza de esta evidencia y, nuevamente, podría decirse que no hay necesidad de recurrir a este tipo de representaciones visuales si las leyes se captan directamente.

Una última observación sobre las analogías positivas que Brown señala entre experimentación mental y real, es que no elabora criterios que permitan detectar errores en la experimentación mental. Para este filósofo, es posible intuir principios abstractos incorrectamente dado que la intuición es análoga a la percepción visual, y, como ella, también es falible. Pero a diferencia de la percepción ordinaria, en la cual se puede recurrir, por ejemplo, a leyes físicas para identificar errores, Brown no explica qué criterios permitirían detectar errores en la captación de entidades abstractas. Por lo demás, el carácter dinámico de las representaciones visuales y la perspectiva de primera persona no aportan elementos que fundamenten el carácter objetivo de esta vía de conocimiento, dado que el resultado de esta clase de experimentos no está constreñido por las características del mundo, sino por nuestra capacidad de representación.

En el segundo grupo de perspectivas según las cuales los experimentos mentales constituyen casos límite de la experimentación real, también se establece una analogía entre visualización y observación. Aunque estas posiciones tienen en común una tesis sobre la naturaleza de los experimentos mentales, le atribuyen a las operaciones de visualización diferentes papeles. Häggqvist (2009: 73), por ejemplo, sostiene que los experimentos mentales no pueden ser reducidos a argumentos, fundamentalmente porque no están compuestos por entidades con valor de verdad, sino por eventos o procesos que nos permiten formar creencias del mismo modo que en los experimentos ordinarios:

Su ubicación es diferente, por supuesto: los procesos que constituyen (un caso de) un experimento físico tienen lugar en el laboratorio (o algo similar); los procesos que constituyen (un caso de) un experimento mental son psicológicos y tienen lugar en gran medida dentro del cráneo del experimentador mental, aunque también pueden involucrar dispositivos protésicos de este tipo. (Häggqvist, 2009: 62)

La contemplación de una secuencia de eventos motiva, en los experimentos ordinarios, la formulación de enunciados observacionales que son puestos en relación con enunciados teóricos a través de un argumento. Para ello, los experimentadores emplean un conjunto de hipótesis y teorías presupuestas que les permiten realizar inferencias. Aunque en los experimentos mentales la relación entre enunciados observacionales e hipótesis teóricas es más problemática, guarda una semejanza con la relación entre enunciados en el caso de los experimentos reales. En la analogía propuesta por Häggqvist, la representación del escenario contrafáctico permite poner a prueba los resultados observacionales inferidos de los enunciados teóricos junto con ciertas condiciones iniciales.

Sorensen (1992), por su parte, afirma que los experimentos mentales proceden de los experimentos ordinarios y se fundan filogenéticamente en nuestra capacidad como especie de experimentar. La habilidad de representarnos

situaciones hipotéticas es esencialmente el mismo tipo de habilidad que gobierna el funcionamiento de los experimentos ordinarios. La única diferencia entre estos y los experimentos mentales es que los últimos prescinden de la ejecución. Sorensen no desarrolla las analogías entre experimentos mentales y reales en relación con el aspecto observacional, pero argumenta que la variación hipotética de los elementos de una situación es el mismo tipo de estrategia que se emplea en el diseño de un experimento real y afirma que en esto se funda la continuidad.

Gooding (1992), a su vez, sostiene que la participación personal y la incorporación del conocimiento científico aceptado son aspectos fundamentales del carácter experimental de los experimentos mentales. Los experimentos, por lo general, ponen a prueba la viabilidad de una teoría a través de prácticas que la vinculan con los aspectos del mundo acerca de los cuales versa y de esta manera posibilitan la contrastación empírica (Gooding, 1992: 281). Los experimentos mentales que hacen posible esta clase de crítica recuperan lo suficiente del conocimiento contextual para que un proceso experimental funcione en el mundo físico tal como es representado por la teoría. Del mismo modo que las narrativas de los experimentos ordinarios deben transmitir el *saber cómo* del experimentador para posibilitar su replicación, la narrativa de los experimentos mentales debe transmitir lo suficiente del *saber cómo* experimental para hacer que los lectores puedan seguir las transformaciones indicadas por la narrativa y participar indirectamente del proceso. Desde este punto de vista, la participación personal es fundamental ya que de ésta depende, en gran parte, el poder heurístico de los experimentos mentales.

Gooding rechaza la idea según la cual la principal diferencia entre experimentos mentales y reales reside en que los primeros recurren a representaciones mientras que los segundos involucran recursos materiales y habilidades prácticas. Esta diferencia se desdibuja, según Gooding, si se considera que tanto los experimentos mentales como los experimentos reales efectivos solamente son posibles cuando los científicos disponen del

conocimiento corporizado necesario. La familiaridad con el mundo representado por la teoría es fundamental en los experimentos mentales ya que su poder persuasivo reside en que, por un lado, describen un mundo con el que los experimentadores están lo suficientemente familiarizados y, por otro lado, introducen algún elemento que produce extrañeza, por ejemplo, mostrando que algunas acciones son impracticables en el mundo tal como es descrito por la teoría o que conducen a paradojas (Gooding, 1992: 281). Gooding señala que

[...] los experimentos mentales descansan en visualizaciones y modos de aprehensión que dependen de la *cognición corporizada*. Nos permiten localizarnos con los constituyentes de un mundo simulado a través nuestras capacidades perceptivas y de juicio ordinarias y compartidas. Para apelar al ojo de la mente vuelven sobre algo muy parecido al sentido común de la percepción. (Gooding, 1992: 287)

Detrás de esta forma de cognición se encuentra, según Gooding, una gran cantidad de conocimiento corporizado del mundo real. El experimentador puede “ver”, por ejemplo, el efecto de la gravedad sobre los cuerpos o imaginarse a sí mismo sin peso en virtud de su familiaridad con el mundo tal como es representado por el estado actual del conocimiento científico. A diferencia de los procesos de observación científica ligados a la experimentación que se han desplazado más allá de la percepción ordinaria, los experimentos mentales se basan en la visualización y en los modos de aprehender que dependen del conocimiento corporizado y que se asemejan más al sentido común que a la observación científica. La impresión de que en los experimentos mentales es posible captar rápidamente cómo funciona alguna parte del mundo físico es una ilusión. La ilusión de comprensión inmediata se debe a que los escenarios presentados en esta clase de experimentos son sumamente estables y los experimentadores son sumamente hábiles, pero esta situación es producto del estado del conocimiento científico y del conocimiento incorporado por los experimentadores:

[...] en estos casos, una compleja alianza de habilidades, tecnología y educación hace confiable a la percepción. Esto hace al proceso observacional tan transparente que parece que aprendemos algo de la realidad directamente, pero esta apariencia es una ilusión. (Gooding, 1992: 288)

Gooding, Haggqvist y Sorensen argumentan que el mismo tipo de operaciones (u operaciones muy similares) se llevan a cabo en un experimento mental y uno real y en este aspecto se funda tanto la explicación de su funcionamiento como la garantía de su confiabilidad epistémica. Desde nuestro punto de vista, la idea de que la contemplación de un escenario imaginario causa enunciados que tienen el mismo estatus que los enunciados observacionales es sumamente discutible. Haggqvist atiende a la relación entre los enunciados observacionales y teóricos y asume que la relación lógica entre enunciados teóricos y observacionales es idéntica en los argumentos que acompañan a los experimentos mentales y reales. Pero no examina la relación entre la visualización de un escenario imaginario y el valor de verdad de los enunciados que motiva. En este caso, como en la interpretación de Brown, el problema reside en determinar qué relación existe entre estas representaciones y estados de cosas en el mundo natural. Por otra parte, los argumentos de Sorensen para defender la idea de que el rasgo esencial de los experimentos es la modificación de una situación para establecer relaciones entre variables no son concluyentes. Apelan a metáforas y casos poco típicos de experimentos. Pero, fundamentalmente, contradice una de las tesis centrales del experimentalismo de acuerdo con la cual lo característico de todo experimento es la manipulación de un fenómeno. Esta tesis merece ser tratada con mayor atención. Baste por ahora señalar que Sorensen no presenta apropiadamente un argumento para dar cuenta de cómo la manipulación de una situación imaginaria podría suplir la intervención en el mundo natural. Simplemente, señala que es posible descartar como típicos todos los rasgos propios de los experimentos reales (ejecución, uso de instrumental, carácter público) excepto la manipulación controlada de variables y que este es un rasgo compartido por los experimentos mentales.

Tampoco la noción de visualización como producto de la incorporación de cognición corporizada, a la que apela Gooding, permite establecer una analogía clara con el tipo de observación que ocurre en un experimento ordinario. Aunque la observación de los científicos esté configurada por el conocimiento experto y esto les permita dirigir la atención sobre ciertos aspectos del mundo en lugar de otros, o les permita reconocer patrones donde los observadores neófitos no los ven, o incluso aunque la familiaridad con la teoría determine maneras de ver el mundo, estas habilidades no garantizan que los científicos puedan anticipar el mundo de manera correcta. Es posible que las representaciones vinculadas a este conocimiento anticipen correctamente un fenómeno. Por otra parte, aunque la noción de visualización incluye algunos rasgos atribuidos a la observación científica, como la incorporación del conocimiento en los observadores, el propio Gooding señala que el tipo de visualización invocada en los experimentos mentales es más parecida al sentido común que a la observación científica debido a que se basa en la familiaridad que el científico tiene con el mundo. Las visualizaciones en los experimentos mentales son posibles porque la narrativa del experimento permite imaginar una escena con tales o cuales características o un aparato que funcione de cierta manera. Pero el vínculo entre estas imágenes y los fenómenos que son objeto del experimento no es aclarado. La capacidad de representación producto del estado de desarrollo de la ciencia y la incorporación del conocimiento científico son las que posibilitan “percibir” los modos estables en los que se articula la experiencia. Es difícil ver como este modo de percibir, que, según Gooding, es más cercano al sentido común que a los procesos de observación científica, puede garantizar la obtención de algún conocimiento científico. En el marco de esta teorización el vínculo entre esta forma de conocimiento y las visualizaciones también es opaco.

En síntesis, la afirmación de que existe una semejanza no trivial entre observación científica y visualización de un fenómeno se funda en una débil analogía. La perspectiva de Brown representa la posición más radicalizada al

postular que los experimentos mentales son análogos estructurales de los reales que habilitan una forma de intuición intelectual. El valor heurístico de esta clase de experimentos se relaciona con instantaneidad con la que se obtienen los resultados y su valor evidencial se debe al tipo de operación (la visualización de un fenómeno) que es la contraparte de la observación y que genera creencias a partir de la contemplación de fenómenos que ejemplifican leyes. Pero este punto de vista demanda un compromiso muy costoso con principios metafísicos y epistémicos cuya suscripción implica una serie de problemas. La dificultad más importante es la de explicar cómo es posible la interacción con entidades abstractas. En el marco de la epistemología de la experimentación, este es el elemento que garantiza la objetividad del conocimiento del mundo natural. La segunda gran dificultad es la ausencia de criterios de corrección para la intuición intelectual. Dentro del segundo grupo de perspectivas, la dificultad para establecer criterios de confiabilidad epistémica es también una razón poderosa para socavar la analogía entre visualización y observación. A diferencia del relato de un experimento real, las operaciones descritas en la narrativa de un experimento mental no se ejecutan y por lo tanto, no es la naturaleza la que determina el resultado, sino nuestras representaciones.

2.6 Fenómenos y experimentos mentales

La única posición que admite que los experimentos mentales son capaces de proporcionar evidencia que puede confirmar hipótesis científicas es la de Brown (1991, 2004a, 2004b y 2011). Brown sostiene que, si se acepta la noción constructiva de fenómeno propuesta por Bogen y Woodard (1988), entonces, tiene sentido considerar a los fenómenos como entidades abstractas y con ello, es plausible que en algunos casos puedan ser captadas directamente, es decir, sin inferirlos a partir de los datos.

La idea de que los fenómenos son los objetos de las teorías y de que, en cierto sentido, son construidos a partir de los datos, es central en la defensa que Brown realiza de la posibilidad del conocimiento *a priori* del mundo natural. A partir de la alteridad entre fenómenos y datos, postula la idea de que los primeros son entidades abstractas que corresponden a clases naturales (Brown, 1993: 125). De la teoría citada, concluye que los fenómenos no son observables en ningún sentido interesante del término y que las relaciones de justificación se establecen entre teorías y fenómenos y no entre teorías y experiencias u observaciones (véase Brown, 1994: 119).

Con el objetivo de defender esta interpretación, reconstruye casos que él considera que son ejemplos de construcción de fenómenos. Un caso paradigmático de este proceso es la clasificación de los elementos químicos en función de su peso y de sus propiedades químicas elaborada por Mendeléyev y la taxonomía de los elementos en función de su número atómico, llevada a cabo por Moseley. Ambas clasificaciones coinciden, excepto por algunos casos. No obstante, en la tabla periódica de Moseley existe un acuerdo completo entre la recurrencia en las propiedades químicas y el incremento del número atómico, que se identifica con el número de protones que contiene el núcleo del átomo de cada elemento. Según el criterio de clasificación de Moseley, los diferentes isótopos de un mismo elemento (como el deuterio y el tritio respecto del hidrógeno) pertenecen a una misma clase natural, aunque tengan propiedades físicas muy diferentes unos de otros. Para Brown, el de Moseley es un claramente un caso de construcción del fenómeno ya que muestra cómo el descubrimiento de propiedades esenciales es relevante para delimitar clases naturales. Según Brown (1994: 125), “el mundo está lleno de datos pero hay relativamente pocos fenómenos”. Este es el tipo de trabajo que Moseley llevo a cabo a partir de la identificación de propiedades de clase natural.

La construcción de fenómenos es un proceso que Brown compara con la dialéctica ascendente a la que Platón se refiere en la metáfora empleada en *Fedro*

265e: "el carnicero es capaz de seccionar el cuerpo según sus articulaciones". Debido a que los fenómenos son entidades abstractas, sería posible en principio acceder a ellos sin los datos a partir de los cuales pueden ser construidos. Esta posibilidad estaría dada por la actividad de visualización que tiene lugar en los experimentos mentales. Asimismo, Brown ofrece una interpretación del valor evidencial de los fenómenos abstractos, paralela a la de Bogen y Woodward. El aspecto visualizable del fenómeno no entra en contradicción, desde este punto de vista, con su alteridad respecto de los datos empíricos y no implica que el fenómeno pueda ser percibido. Ocurre que los fenómenos que tienen lugar en los experimentos mentales son clases naturales o patrones que podemos visualizar. Tienen relación con la observación dado su carácter visual, pero deben distinguirse de los datos de la percepción sensible porque son objetos abstractos. Es posible intuir leyes de la naturaleza en estas situaciones de visualización precisamente porque los fenómenos ejemplifican clases naturales o patrones. Las clases naturales corresponden a un tipo de propiedad esencial que implica que si un miembro de la clase tiene esa propiedad, entonces, todos los miembros la tienen. El tipo de inferencia que hace posibles los experimentos mentales es la relevante para justificar una teoría o hipótesis científica: los fenómenos son evidencia para las teorías o hipótesis.

El problema con este punto de vista es que supone precisamente lo que la crítica a la concepción tradicional intenta desechar. La idea de que podemos acceder a los fenómenos con independencia de los datos supone que estos preexisten a su construcción o al menos que no dependen ontológicamente de los datos. Esto parece estar directamente en desacuerdo con la noción constructiva de fenómeno en la que Brown dice apoyar su interpretación. Para la posición de Bogen y Woodward, los datos son evidencia del fenómeno y es a través de un complejo proceso de construcción que implica la eliminación del error, la singularización de una gran masa de datos, el análisis estadístico, etc. que los fenómenos son producidos. Dado que existe una conexión causal entre datos y

fenómenos, no se postulan procesos de conocimiento de los fenómenos que prescindan de los datos.

Por otra parte, la perspectiva de Brown parece estar comprometida con una ontología robusta en lo referente a la naturaleza de los fenómenos. Desde su punto de vista, la realidad está articulada en patrones universales y estables que subyacen a las manifestaciones empíricas (Brown, 1994: 135). Es por esto que los experimentos mentales son capaces de justificar teorías al exhibir fenómenos puros que solo son conocidos de manera imperfecta en la experimentación ordinaria. Los resultados de experimentos mentales como el de Galileo proporcionan evidencia conclusiva en contra de una hipótesis y a favor de otra porque el fenómeno en cuestión trasciende al empirismo. No obstante, es una posición ampliamente difundida entre la filosofía de las prácticas experimentales que la significación evidencial es conferida a los experimentos en momentos particulares de la historia y en áreas específicas de la ciencia como consecuencia del esfuerzo persuasivo de los científicos, pero que éstos no son autoevidentes. Los experimentos mentales poseen una significación evidencial intrínseca para Brown (McAllister, 1996: 235) porque facilitan un acceso privilegiado a una forma pura de fenómenos.

En suma, creemos que de la noción de fenómeno no se sigue en ningún sentido que puedan conocerse fenómenos independientemente de la interpretación o el procesamiento de los datos a partir de los cuales son inferidos. Aunque como afirma Hacking, ciertos fenómenos solo puedan crearse en el laboratorio, esto se debe a que es posible manipular la naturaleza y no solamente revelar conexiones regulares preexistentes. Al menos desde esta clave de análisis, no parece viable acceder a los fenómenos sin mediación de los datos. Aun si esto fuera posible, no podría ofrecerse en el contexto de estos presupuestos una explicación razonable de cómo una intuición de este tipo puede justificar un principio teórico sin asumir una serie de compromisos metafísicos que de base son descartados por las teorizaciones sobre la práctica científica.

2.7 La autonomía de los experimentos mentales

De acuerdo con el punto de vista que argumenta que los experimentos mentales pertenecen al conjunto de las prácticas experimentales, éstos comparten una serie de rasgos con los experimentos reales. Hasta el momento se ha analizado la idea según la cual dado que los experimentos mentales poseen un carácter visual, en ellos el experimentador desempeña un rol activo, permiten visualizar fenómenos de cierto tipo y por lo tanto son esencialmente similares a los experimentos ordinarios. El examen reveló que las analogías positivas entre la observación y la visualización en los experimentos mentales son insuficientes para considerar que los experimentos mentales y los reales son especies del mismo género ya que no pueden establecerse para la visualización criterios de fiabilidad epistémica ni una explicación plausible del proceso de interacción en virtud del cual es posible obtener información acerca del mundo físico sin alguna forma de intervención sobre él. El carácter activo del experimentador y la perspectiva de la primera persona no parecen agregar razones para argumentar que esta forma de experimentación es capaz de proporcionarnos conocimiento objetivo.

No obstante, existen otras similitudes importantes entre los experimentos mentales y reales que permitirían respaldar la idea de que se trata de un mismo tipo de prácticas. En general, quienes defienden esta hipótesis afirman que la actividad fundamental de la experimentación es la introducción de variaciones en una situación controlada, que los experimentos mentales pueden cumplir múltiples funciones y que son relativamente independientes de la teoría. Esta es una caracterización usual de la experimentación y los aspectos mencionados se vinculan con la defensa de la tesis de la *autonomía* de la experimentación. Esta hipótesis ha sido apoyada, por un lado, por la afirmación empírica de que los experimentos cumplen múltiples funciones y no se encuentran subordinados a la teoría en tanto no son siempre precedidos por esta. Por otro lado, la estabilidad

de la evidencia experimental y su independencia respecto de las teorías son otras interpretaciones de esta tesis que han encontrado apoyo en el análisis de casos. A continuación se intenta determinar si es posible extender alguna interpretación de la tesis de la autonomía de la experimentación al caso de los experimentos mentales y si es posible defender sobre esta base que deben estos considerarse como experimentos.

Las posiciones que argumentan que los experimentos mentales son análogos estructurales de los reales ofrecen una explicación de cómo esta analogía puede sostenerse con independencia de la intervención en el mundo natural. En la perspectiva de Brown, la diferencia entre experimentos mentales y reales reside en una diferencia respecto del lugar en el que se llevan a cabo. Los experimentos mentales tienen lugar en el laboratorio de la mente. En estos casos ocurre una manipulación sobre la situación imaginada porque la narrativa mediante la que estos experimentos son comunicados compele al lector a introducir cambios en la situación y considerar cursos posibles de acción. Para Brown, los límites entre experimentación teorización son difusos, aunque podría decirse que, en un sentido amplio, un experimento incluye una teoría, cierto conocimiento de fondo, un fenómeno visualizable y un resultado. En un sentido más estrecho, el experimento consiste en aquello que se observa, esto es, el fenómeno. Esto vale tanto para los experimentos mentales como para los reales (Brown, 2011: 51). Por esta razón incluso, los experimentos mentales son capaces de presentar fenómenos que no se encuentran en la naturaleza. Tal es el caso del experimento del balde giratorio de Newton, en el que la superficie cóncava nos proporciona evidencia de la existencia del espacio absoluto, que es por sí mismo inobservable (Brown, 2011: 38).

Otra familia de posiciones que admite que los experimentos mentales son casos límite de los reales, funda la similitud entre ambas prácticas en la noción de experimento formulada por Mach ([1905] 1948:161), de acuerdo con la cual los experimentos mentales son instancias del método de la variación concomitante y

su función es la de disparar, a través de la descripción de un escenario específico, un proceso mental que permite movilizar el conocimiento instintivo con el fin de desarrollar un conocimiento nuevo y explícito. En esta línea interpretativa, Sorensen (1992: 186) afirma que los experimentos mentales son casos límites de los experimentos reales. En su enfoque, un experimento es esencialmente un procedimiento para plantear o responder una pregunta acerca de relaciones entre variables. La manipulación y el uso de instrumental son rasgos estereotípicos de los experimentos aunque su posesión o carencia no determina la pertenencia a esta clase de prácticas¹², lo que define a los experimentos es su objetivo cognitivo. En consecuencia, la ausencia de intervención en el mundo natural no determina la exclusión de los experimentos mentales de este conjunto de prácticas. Entre experimentos mentales y reales existe una diferencia de grado: los experimentos mentales evolucionaron de los experimentos ordinarios por un proceso de atenuación.¹³

En una posición similar se ubica Gooding (1992), para quien los experimentos mentales no requieren instrumentos ni experimentadores de carne y hueso, pero presentan alguna propiedad del mundo, anteriormente desconocida, con una fuerza que ningún experimento real puede equiparar. Su principal ventaja es que son sumamente fáciles de replicar y que su fuerza heurística es tal que no

¹² Sorensen (1992: 188) se limita a enumerar ejemplos de experimentos que no involucran el uso de instrumental o implican manipulación causal, pero no explica por qué no es esencial que los experimentos tengan estos rasgos. Dado que algunos de estos ejemplos son de experimentos mentales y otros son de experimentos fallidos, la referencia a ejemplos como prueba de la tesis que se pretende demostrar es circular, ya que supone que se acepte que la ausencia de ejecución material no es privativa de que el caso examinado sea un experimento.

¹³ Para Sorensen en el estado del conocimiento de la ciencia actual, la ejecución de algunos experimentos es inesencial. Los experimentos mentales pueden producir el mismo rédito epistémico a través de la mera suposición. Este hecho es el resultado de un proceso en el cual se fueron omitiendo paulatinamente partes del proceso experimental. El avance del conocimiento hizo posible que algunos ensayos o partes del proceso experimental fueran discrecionales porque los experimentadores fueron paulatinamente capaces de responder cuál habría sido el resultado de los ensayos faltantes (Sorensen, 1992: 195). En este sentido los experimentos mentales son experimentos atenuados, carecen del elemento de la ejecución. o son diseñados para ser ejecutados por diferentes razones aunque esto es posible solo porque los experimentadores pueden responder “¿qué sucedería?” en las situaciones descritas en el experimento.

se requiere de la realización de experimento real que confirme sus resultados. Los experimentos mentales no son para Gooding una práctica diferente de los experimentos reales, sino que constituyen otra forma de razonamiento experimental. Dado que los escenarios descritos en los experimentos mentales recuperan lo suficiente del mundo empírico la crítica informada empíricamente es posible porque la narrativa experimental invoca el conocimiento contextual que los experimentadores necesitan para hacer funcionar el proceso experimental.

En síntesis, la defensa de la perspectiva experimental en sus dos versiones implica resignificar la noción de intervención, dándole el sentido de manipulación mental de variables.

Como se ha señalado, los experimentos reales son independientes de la teoría en un sentido más profundo: tienen vida propia. En cambio, los experimentos mentales no son capaces de evolucionar ni de ser empleados en contextos teóricos diferentes de aquellos en los cuales fueron originalmente presentados (Hacking, 1992). Se han formulado diversos argumentos para impugnar esta tesis. Gooding (1984) responde directamente al punto de Hacking y señala que los experimentos mentales son producto del desarrollo y la evolución de la observación científica. Buzzoni (2009) señala que la observación de Hacking no muestra una deficiencia de los experimentos mentales sino que por el contrario expresa una importante debilidad de la perspectiva del *nuevo experimentalismo*¹⁴. Esta posición queda atrapada en la misma dicotomía que denunció en sus orígenes. Para no ser absorbidos por la dependencia teórica, los principales exponentes del nuevo experimentalismo supusieron que las prácticas experimentales son independientes de las prácticas teórico-lingüísticas y postularon que, a diferencia de las teorías, la experimentación "tiene una vida propia." Haciendo tal afirmación, Hacking terminó respaldando la misma

¹⁴ Con esta denominación Buzzoni se refiere a la corriente representada por Hacking, Franklin, Galison y otros filósofos de la ciencia, que ha promovido el giro del estudio de las teorías científicas hacia el estudio más amplio de la práctica científica.

dicotomía de teoría y experimento que criticó en Popper y en las filosofías relativistas de la ciencia (Buzzoni, 2009: 4). En el mismo sentido en que los principios de individuación de un experimento no pueden ser separados de los aspectos teóricos, sería imposible concebir la reproductibilidad sin una mediación teórica. Desde este punto de vista, los experimentos mentales poseen una naturaleza teórico dialógica similar a la de los experimentos ordinarios. Se fundan en el método de la variación concomitante y describen un aparato experimental cuyo diseño es intersubjetivamente contrastable.

Otra respuesta directa a la objeción de Hacking señala que los experimentos mentales tienen vida propia (Shinod, 2016). De acuerdo con esta posición, aunque son contruidos en base a una teoría y son dirigidos por los supuestos teóricos, este también sucede con los experimentos reales. Los experimentos mentales poseen autonomía en varios de los sentidos en las que esta cualidad se predica de los experimentos ordinarios. En primer lugar, poseen autonomía funcional ya que pueden cumplir legítimamente muchas funciones y sus resultados persisten frente al cambio teórico. Asimismo, es demostrable a partir del análisis de casos que los experimentos mentales poseen autonomía respecto de los contextos teóricos en los cuales fueron formulados. Esto implica que maduran, se adaptan y se reequipan. Según Shinod, el experimento de Einstein, Podolski y Rosen (EPR), por ejemplo, evolucionó desde su formulación inicial. Bohr redefinió el experimento y señaló que la noción de realidad propuesta por Einstein era ambigua. El experimento fue además reequipado por Bohm, quien introdujo el *spin* en lugar de la posición y el momento, y por Bell quien formuló un teorema experimentalmente contrastable. Asimismo, los experimentos mentales se relacionan con experimentos reales: el experimento de Aspect, por ejemplo, llevó a cabo la versión de Bohm del experimento EPR (Shinod, 2016: 16). Por otra parte, los resultados de los experimentos mentales permanecen estables. La robustez de los resultados de experimentos como el de los cuerpos en caída libre de Galileo se debe a que recuperan interacciones causales que son hasta cierto

punto independientes de las interpretaciones teóricas. El experimento de Galileo rechaza la caracterización aristotélica de la relación entre velocidad y peso de un cuerpo en caída libre y propone la interpretación correcta (Shinod, 2016: 18). Dado que el resultado del experimento mental de Galileo establece una relación fenoménica, una afirmación acerca de lo que la naturaleza es, el resultado permanece el mismo a pesar de los cambios en el dominio teórico.

La tesis de que los experimentos mentales no son siempre precedidos por una teoría específica, se vincula con la interpretación de los vínculos posibles entre experimentos y teorías. Según Hacking aunque es preciso tener algunas ideas previas acerca de la naturaleza y del aparato experimental antes de realizar un experimento, es falso que un experimento sea significativo solo cuando está poniendo a prueba una teoría acerca del fenómeno bajo escrutinio (Hacking, 1983: 154). Muchos de los primeros desarrollos de la teoría óptica, por ejemplo, dependieron de que se advirtieran algunos fenómenos interesantes como el de la doble refracción del espato de Islandia. En otros casos, la teoría y la experimentación proceden de direcciones diferentes y se vinculan a partir de un encuentro fortuito, como en el descubrimiento de la radiación cósmica de fondo (ver Hacking, 1983: 156-165). Steinle (1997), por su parte, argumentó que la experimentación exploratoria, es típicamente practicada en periodos en los que no hay teorías o incluso ningún marco conceptual disponible. No obstante su independencia de teorías específicas, esta forma de experimentación puede ser altamente sistemática y dirigida por pautas típicas (Steinle, 1997: 70). En síntesis, la función exploratoria es reconocida como un uso legítimo de los experimentos y es una razón para argumentar la independencia de la experimentación respecto de la teoría.

Es fácil mostrar, y algunos de los defensores de la perspectiva experimental lo han hecho, que los experimentos mentales no son siempre precedidos por alguna teoría. Brown (2011: 38) denomina *experimentos conjeturales* a aquellos que se proponen establecer un fenómeno e hipotetizar una teoría capaz de

explicarlo. Este es el caso de los experimentos mentales que sirvieron como herramientas útiles en campos en los cuales la experimentación real es sumamente restringida o es imposible, como en el caso de la cosmología. Las funciones más usuales de esta clase de experimentos es mostrar que ciertos conceptos involucran contradicciones, ilustrar una posición muy abstracta, detectar vaguedades, identificar casos límite de un concepto y ayudar a precisarlo. Al menos en este sentido, podría sostenerse que los experimentos mentales son independientes de la teoría en algún grado. No obstante, sigue resultando problemática la idea de que un experimento mental sea capaz de exhibir un fenómeno nuevo, nunca antes experimentado.

En síntesis, las perspectivas que sostienen que los experimentos mentales son una especie de experimentos responden solo parcialmente a los problemas que podrían plantearse en tal empresa. Fundamentalmente, sobre la base de los desarrollos precedentes en el marco de la filosofía de la experimentación, podría señalarse que las dificultades fundamentales son dos: a) los experimentos mentales no intervienen en el mundo material; y b) sus resultados no son independientes de toda teoría. Las posiciones que mantienen que los experimentos mentales son experimentos en el sentido descrito por Mach proporcionan razones, basadas en casos tomados de la historia de la ciencia, para sostener que los experimentos mentales cumplen múltiples funciones y, en algunos casos, que evolucionan en el sentido señalado por Hacking.

El carácter hipotético del escenario descrito en los experimentos mentales ha sido la razón principal por la cual fueron considerados como argumentos teóricos. Que el aparato experimental sea irrealizable o extremadamente difícil de materializar es una de las razones cuales se ha argüido que la descripción de un escenario contrafáctico es una premisa que se emplea para mostrar una contradicción o ambigüedad en algún principio teórico. No obstante, la posibilidad de replicar un experimento mental como experimento real obteniendo los mismos resultados puede considerarse un elemento para legitimar los resultados de esta

clase de experimentos y afirmar que, al menos algunos experimentos mentales, gozan, en un sentido restringido de “vida propia”.

2.8 ¿Son experimentos los experimentos mentales?

Un experimento que permite examinar la relación entre experimentos mentales y reales es el de la doble rendija tal como fuera propuesto por Feynman a mediados de la década de 1960. El caso es excepcional porque reúne varias de las características que han generado controversia entre los filósofos que se han ocupado de la epistemología de los experimentos mentales. Como experimento mental exhibió un fenómeno que solo mucho tiempo después pudo ser realizado en el laboratorio e ilustró uno de los principios fundamentales de la mecánica cuántica. En su formulación, Feynman incluyó imágenes que ayudan a visualizar el fenómeno en cuestión. La descripción del escenario involucra un sofisticado aparato experimental y relata una serie de operaciones que el experimentador debe realizar para obtener los resultados.

El experimento de la doble rendija tiene su antecedente más remoto en el experimento realizado por Young en 1801. Este constituyó la primera evidencia convincente de la naturaleza ondulatoria de la luz. A diferencia de la interpretación ampliamente aceptada en el momento, de acuerdo con la cual la luz consiste en corpúsculos que se propagan en línea recta, el resultado del experimento mostró que un haz de luz al ser difractado proyecta sobre una pantalla un patrón de interferencia consistente con el comportamiento de ondas que al pasar por una rendija forman nuevos frentes de onda que se dispersan en todas las direcciones generando zonas brillantes cuando las amplitudes de onda están en fase, y producen una interferencia constructiva y zonas oscuras cuando las ondas están fuera de fase y la interferencia entre éstas es destructiva. Young lo expresó en estos términos:

[...] estamos autorizados a inferir que luz homogénea a cierta distancia en la dirección de su movimiento, posee cualidades opuestas, capaces de neutralizarse o destruirse mutuamente, y de extinguir la luz cuando están unidas; que estas cualidades se suceden alternativamente en sucesivas superficies concéntricas a distancias que son constantes para la misma luz que pasa a través del mismo medio. (Young, 1804: 99)

El experimento de Young cobró especial relevancia en el contexto de la mecánica cuántica, que incorpora la llamada dualidad onda-partícula, según la cual, todas las entidades subatómicas exhiben en ciertas circunstancias propiedades características de las partículas, como la localización y el momento lineal, y, en otras circunstancias, propiedades características de las ondas, como la difracción y la interferencia. Esto se aplica igualmente a las partículas materiales, como los electrones, protones y neutrones que componen los átomos, como a la radiación electromagnética (la luz), que desde el punto de vista cuántico se concibe compuesta de fotones. Feynman (1965) diseñó un experimento mental técnicamente irrealizable en el momento de su formulación, que exhibe un fenómeno que no es explicable en los términos de la mecánica clásica de partículas. El diseño experimental supone un aparato capaz de emitir electrones, uno a la vez, frente al cual se encuentra delgada lámina metálica con dos ranuras muy pequeñas en ella. Más allá de esta pared hay otra lámina que sirve como detector de las partículas que atraviesan la primera pantalla. En esta segunda pantalla se encuentra instalado un detector móvil. El experimento mental se realiza en dos partes. La primera involucra calcular las distribuciones de probabilidad de que los electrones impacten en la pantalla detectora en tres escenarios diferentes: electrones que viajan a través de la ranura 1 cuando la ranura 2 está cerrada; electrones que viajan a través de la ranura 2 cuando la ranura 1 está cerrada; y electrones viajando a través de ambas ranuras (Figura 1a). Estos escenarios ilustran el principio de interferencia cuántica, es decir, el aspecto ondulatorio de las partículas. La segunda parte del experimento mental consiste en la observación de electrones individuales asociados con 'clics' de detección, lo cual muestra que en el tercer escenario se produce un patrón de interferencia, que

no concuerda con la distribución de probabilidad calculada suponiendo que son partículas que pasan por una sola de las dos rendijas.

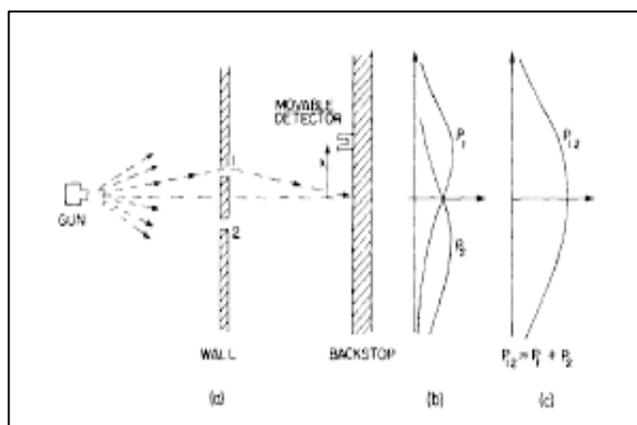


Figura 2.1

En la primera parte del experimento, si se supone que los electrones son similares a corpúsculos que llegan a la detección atravesando alguna de las rendijas de la lámina situada frente al cañón de electrones, podría calcularse la probabilidad relativa de que un electrón llegue a la pantalla de contención a una distancia del centro contando el número de proyectiles que llegan al detector en cierto tiempo y el número total que golpea la contención durante ese tiempo. Si se bloqueara primero una de las dos rendijas a través de las cuales pasan los proyectiles y luego la otra, la suma de la distribución de probabilidad que se obtiene en cada caso sería equivalente a la distribución de probabilidad que se obtiene cuando ambas rendijas están abiertas. No obstante, para el diseño experimental descrito arriba, el tipo de curva de distribución observada cuando ambas rendijas se encuentran abiertas (Figura 1 c) no se parece a la que podría preverse si los electrones se comportaran como partículas (Figura 1b). Feynman señala que cuando ambas rendijas están abiertas el patrón observado es compatible con un patrón de interferencia. Esto implica que no es posible interpretar este resultado como el que se obtendría si los electrones que llegan a la contención lo hicieran como gránulos que pasan individualmente por una u otra

de las rendijas. La interpretación matemática de este fenómeno es la misma que se emplea para dar sentido a un experimento similar que se realizara con ondas en el agua.

En algunos lugares las ondas están en fase [...] y los máximos de las ondas se suman para dar una gran amplitud y por lo tanto una gran intensidad. Decimos que las ondas están interfiriendo constructivamente en estos lugares. Habrá tal diferencia constructiva siempre que la distancia desde el detector sea mayor (o menor) en un número entero de longitudes de onda, que la distancia desde el detector al otro agujero. (Feynman, 1965: 37-5)

En la segunda parte del experimento, Feynman agrega una fuente de luz entre la pantalla con las rendijas y la pantalla de contención. Dado que las cargas eléctricas dispersan la luz, cada vez que un electrón pasa camino a la contención dispersa algo de luz, permitiendo al experimentador determinar dónde está el electrón. En este escenario la probabilidad total es equivalente a la suma de la distribución de probabilidad de cada una de las rendijas, es decir, no se produce un patrón de interferencia. Al disminuir la intensidad de la fuente luminosa empleada para detectar los electrones, se disminuye la frecuencia con la que se emiten los fotones. Esto produce que algunos electrones pasen frente a la fuente luminosa sin ser “vistos” y se genera nuevamente el patrón de interferencia. Si se disminuye lo suficiente la intensidad de la luz (empleando luz infrarroja u ondas de radio) para que la amplitud de onda sea mayor que el tamaño de la separación de las rendijas, se produce por el paso de los electrones, un destello borroso, que no permite establecer a través de cuál de las rendijas pasó el electrón. De manera que luego de esta modificación se genera nuevamente el patrón de interferencia. Feynman afirma que es imposible disponer el detector luminoso de tal manera que se pueda saber por qué rendija pasó el electrón sin destruir el patrón de interferencia.

Una realización material relativamente reciente del experimento de Feynman tuvo lugar en 2013. Hasta entonces no se había construido un aparato

experimental con las características del descrito por Feynman (Figura 2).¹⁵ Los resultados del experimento son descritos por los investigadores en la siguiente conclusión:

En este documento, mostramos una realización completa del experimento mental de Feynman e ilustramos una característica clave de la mecánica cuántica: la interferencia y la dualidad onda-partícula de la materia. Controlando la transmisión a través de las particiones individuales los patrones de difracción de la ranura 1 (P1), la ranura 2 (P2) y ambos (P12), observando así las propiedades ondulatorias de los electrones. Además, al registrar los eventos de detección de un solo electrón que se difractan a través de la doble rendija pudimos construir un patrón de difracción, observando las propiedades corpusculares de electrones. (Bach, 2013: 8)

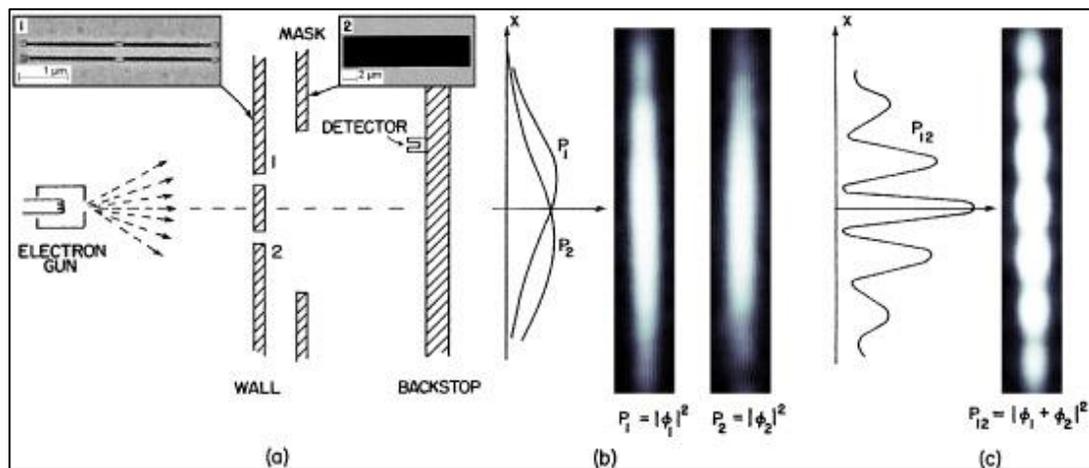


Figura 2.2

¹⁵ Frecuentemente se afirma que la primera realización del experimento fue llevada a cabo por Tonomura y sus colaboradores en 1989. No obstante, las realizaciones previas a la que aquí recuperamos, no se cumplieron todas las condiciones del experimento de Feynman. Tonomura por ejemplo utilizó un cable delgado y cargado para dividir electrones y volver a unirlos, en lugar de las dos rendijas en una pared que descritas por Feynman. Por su parte experimentos de Pozzi (2012) fueron los primeros en usar ranuras nano-fabricadas en una pared; sin embargo, las ranuras se cubrieron rellenándolas con material para que no pudieran abrirse y cerrarse automáticamente. En el experimento de Bach y sus colaboradores (2013) ambas rendijas se pueden abrir y cerrar mecánicamente a voluntad y, lo que es más importante, combina esto con la capacidad de detectar un electrón a la vez.

Los resultados del experimento real fueron los predichos por el experimento mental. El experimento fue formulado por Feynman como un caso que permite ver claramente el comportamiento cuántico de los electrones, a pesar de que en su momento era irrealizable como experimento real. Para explicar el fenómeno se emplean comparaciones con el comportamiento de sistemas con los cuales los físicos están familiarizados (aunque los ejemplos referidos por Feynman involucran cierto grado de idealización). El primero de ellos sugiere considerar el tipo de patrón de distribución que podría obtenerse en un escenario como el descrito si se lanzaran proyectiles a través de una pared que tiene dos rendijas (se supone además que los proyectiles son indestructibles, por lo que llegan siempre completos a la pared de contención y que el arma dispara aleatoriamente los proyectiles en todas las direcciones por lo que solo algunos logran atravesar la pared). Feynman supone que, si se instalara un detector en la pared de contención, sería fácil determinar la probabilidad de que un proyectil impacte a una distancia x del centro. El examen del experimento con proyectiles permite arribar a una conclusión más o menos obvia: la distribución de los impactos en la pared de contención es la suma de la distribución de los proyectiles que impactaron en la contención al atravesar la primera rendija y los proyectiles que impactaron al atravesar la segunda rendija.

A continuación propone realizar un experimento análogo pero con ondas en el agua. Este caso muestra una importante diferencia con el anterior y supone que el experimentador conoce el comportamiento de las ondas y algunas de las leyes a las que obedece. En este dispositivo, el detector mide la intensidad de las ondas. Cuando ambas rendijas están abiertas, la intensidad de las ondas no es la suma de las intensidades de las ondas provenientes de cada una de las rendijas, sino que en algunos lugares, donde las ondas están "en fase", los picos de las ondas se suman para producir una gran amplitud y, por lo tanto, una gran intensidad. Las dos ondas "interfieren constructivamente" en tales lugares. En los lugares donde las ondas llegan al detector con un desfase el movimiento ondulatorio en el

detector será la diferencia de las dos amplitudes. Allí donde las ondas interfieren destructivamente el valor de la intensidad de la onda es bajo. Cuando las ondas están en perfecta oposición de fase, la interferencia destructiva es completa y el resultado es una plana, de amplitud nula.

Los dos experimentos que anteceden a la formulación del experimento mental tienen un alto valor pedagógico ya que facilitan las conclusiones. El diseño del aparato experimental es similar al del primer experimento (con proyectiles), aunque Feynman agrega que tal diseño experimental no podría disponerse porque el aparato tendría que hacerse en una escala increíblemente pequeña para mostrar los efectos en los que estamos interesados. No obstante, indica que el resultado del experimento se apoya en la masa de conocimiento experimental que lo antecede. Por un lado, el detector (que podría ser un contador Geiger) permite identificar cada "clic" con la presencia de un electrón y, por lo tanto, permite concluir que éstos llegan a la pantalla de contención como gránulos. Dado este supuesto, es natural imaginar que los electrones pasan a través de una u otra rendija, ya que se los considera como partículas elementales que no pueden dividirse. Si se bloquea una de una de las rendijas, la distribución de probabilidad que se obtiene es similar a la que se obtuvo en el experimento de los proyectiles. Pero cuando ambas rendijas están abiertas la explicación de la distribución de probabilidad se obtiene empleando la matemática que permite calcular la interferencia entre las ondas de agua (Feynman 37-8). Esto le permite derivar la conclusión de que los electrones llegan a la pantalla de contención como partículas pero su distribución es similar a la distribución de la intensidad de una onda.

¿Podría decirse en este caso que el experimento mental imita lo que debería haber sucedido si el experimento real se realizara? En cierto sentido, el caso depende del conocimiento experimental previamente acumulado. Asimismo, muestra una fuerte dependencia de los principios teóricos y de la interpretación matemática de los fenómenos con los que el caso está relacionado. Hay un

sentido interesante en que el fenómeno en cuestión es visualizable en el experimento mental. La curva que representa la distribución de probabilidad “exhibe” el patrón de interferencia que posteriormente se observó en el experimento real. El experimento de Feynman presenta de este modo un fenómeno que no pudo ser producido en el laboratorio hasta mucho tiempo después. Pero esta presentación es en cierto sentido “abstracta”, derivada de la interpretación matemática del mismo. No obstante, el experimento es sumamente informativo acerca del comportamiento de las partículas. Muestra al mismo tiempo un aspecto del tipo de conocimiento en cuestión: el límite de precisión con que el fenómeno puede ser conocido y las consecuencias de la intervención en la realización de experimentos a esa escala.

El caso presenta importantes similitudes con los experimentos ordinarios: describe un aparato experimental y una serie de operaciones que es necesario realizar para obtener los resultados. Ilustra un fenómeno lo que depende de una serie de principios teóricos. Aunque el grado de dependencia teórica es alto, esto no lo diferencia del experimento real que es su contraparte.

Por otra parte el experimento tiene un alto valor pedagógico. Se apoya en análogos macroscópicos, como proyectiles y ondas de agua, respecto de los cuales los físicos conocen de antemano los resultados que se obtendrían porque hay muchos experimentos que se han realizado. Por esa razón, después de Feynman, y hasta la actualidad, se lo ha utilizado en prácticamente todos los libros de texto de mecánica cuántica, y en casi todas las presentaciones divulgativas de esta teoría, como un ejemplo que muestra la extrañeza cuántica de manera ejemplar. También se lo ha empleado para exhibir el poder explicativo de teorías cuánticas alternativas. Así, por ejemplo, Bricmont (2017) lo usa como un argumento a favor de la mecánica Bohmiana, la cual, a diferencia de la teoría cuántica ortodoxa, permitiría explicar la producción del patrón de interferencia.

El objetivo de la presentación del caso es, de acuerdo con Feynman,

[...] examinar un fenómeno que es imposible de explicar de manera clásica, y que tiene en sí el corazón de la mecánica cuántica. En realidad, contiene el único misterio. No podemos hacer que el misterio desaparezca "explicando" cómo funciona. Solo te diremos cómo funciona. Al contarte cómo funciona, te habremos hablado de las peculiaridades básicas de toda la mecánica cuántica. (Feynman, 1965: 1-1)

Feynman afirma que el resultado del experimento de las dos rendijas con electrones, a diferencia de los experimentos macroscópicos, como el de Young, no puede explicarse, sino solo describirse. Con ello quiere decir que en el contexto de la teoría cuántica ortodoxa no es posible explicar por qué se produce un patrón de interferencia cuando los electrones atraviesan de a uno las rendijas e impactan en la pantallas detectora como corpúsculos puntuales. De alguna manera que la intuición tampoco puede captar, se comportan como si fueran partículas cuando uno de las rendijas está cerrada y se comportan como si fueran ondas cuando las dos rendijas están abiertas.

El experimento presenta con relativa simplicidad un fenómeno extraño para la experiencia ordinaria, que siempre se refiere a objetos macroscópicos. De este modo, pretende introducir cierta familiaridad con el comportamiento del mundo cuántico o entrenar la intuición. De acuerdo con Feynman, como no podemos aprender directamente sobre esta clase de fenómenos tenemos que aprender sobre ellos de una manera abstracta o imaginativa y no por conexión con nuestra experiencia directa.

¿Qué conclusión podría extraerse de la coincidencia entre los resultados del experimento mental y los del experimento real correspondiente? En principio, dado que los segundos confirman las predicciones de Feynman, podemos señalar que es posible replicar exitosamente el experimento mental como experimento real. No obstante, la replicabilidad es un rasgo que solo algunos experimentos mentales tienen. La narrativa detallada del procedimiento experimental, la relación privilegiada con conocimientos de fondo, la existencia de modelos matemáticos del fenómeno, son características que no todos los experimentos mentales

comparten. La dificultad en la realización técnica no constituye un obstáculo para la realizabilidad en principio. Por esa razón, es dudoso que la realización material pueda cumplir un rol decisivo en la evaluación de los experimentos mentales exitosos.

La mecánica cuántica se considera de manera unánime la teoría mejor confirmada por la evidencia experimental disponible en toda la historia de la física. No cabe duda de que el experimento real de las dos rendijas verificó una de las predicciones fundamentales de la mecánica cuántica y, por consiguiente, proporcionó nueva evidencia confirmatoria para esa teoría. ¿Puede decirse que el experimento mental de Feynman también la había confirmado? ¿O solo ilustró lo que la teoría predecía en el escenario descrito por el experimento? Esa es una cuestión sobre la que los filósofos de la ciencia no han alcanzado consenso, pero que deberemos analizar con mayor detalle en capítulos posteriores.

2.9 Conclusión

El objetivo de este capítulo ha sido analizar los alcances de la interpretación de acuerdo con la cual los experimentos mentales son una especie de experimentos. Recuperando la contribución de Hesse (2001: 299) se intentaron establecer los límites de la comparación entre experimentos mentales y reales atendiendo a las analogías positivas y negativas entre ambos. Como punto de partida de esta indagación, se asumió la hipótesis de que la interpretación filosófica de los experimentos reales puede resultar esclarecedora en la investigación del funcionamiento de los experimentos mentales. Esta comparación se centró en tres elementos: la justificación de la objetividad del conocimiento en los experimentos mentales en base a la noción de visualización, la posibilidad de que los experimentos mentales presenten fenómenos y la autonomía de los experimentos mentales.

Las posiciones que interpretan a los experimentos mentales como experimentos genuinos asumen que, dado que es posible identificar en los experimentos mentales mecanismos de generación de creencias análogos a los que tienen lugar en los experimentos reales, pueden justificar hipótesis de la misma manera que estos últimos. La revisión de las principales concepciones acerca de la naturaleza y funciones de la observación científica y de la experimentación, permite arribar a una serie de conclusiones acerca de la naturaleza de los experimentos mentales y de los papeles que las operaciones de visualización y la manipulación desempeñan en estos casos. Como se ha señalado, la percepción sensible ordinaria pierde, en el contexto de la observación científica, su papel epistémico privilegiado. La noción de observación elaborada por las conceptualizaciones de las prácticas científicas asume como fundamento la noción de interacción física y establece criterios de corrección sobre esta base. Asimismo, argumenta la objetividad del conocimiento observacional, fundamentado en la relación causal entre aquellos aspectos observables de los fenómenos y los eventos u objetos en el mundo que son causa de la interacción. No es posible, sin embargo, formular argumentos análogos para justificar la objetividad en el caso de las visualizaciones o esquemas mentales presentes en los experimentos mentales. Tampoco es posible determinar, de manera más o menos consensuada, criterios de corrección para las operaciones de visualización involucradas en los experimentos mentales. Éstas se encuentran supeditadas a las representaciones de los experimentadores y estas, a su vez, a las teorías de las que dichas representaciones dependen.

Por otra parte, la idea de que los fenómenos son abstracciones que se construyen a partir de los datos, es compatible con la perspectiva según la cual algunos fenómenos se pueden crear en el laboratorio. Pero las relaciones entre datos y fenómenos son de naturaleza causal, por lo que no parece plausible aceptar la posibilidad de concebir fenómenos *a priori*. El conjunto de supuestos epistémicos y metafísicos que se relacionan con la posición *apriorista* acerca de

la forma de conocimiento involucrada en los experimentos mentales se analiza con detalle en el capítulo 4. No obstante, es claro que esta alternativa es difícil de compatibilizar con las nociones desarrolladas en este capítulo, ya que en general adoptan epistemologías empiristas.

Las perspectivas que abogan por una interpretación experimental de los experimentos mentales sostienen que estos consisten en la representación de una situación en la imaginación y en la variación de algún aspecto de la misma. Esta actividad es lo suficientemente similar a la experimentación real porque: a) implica que el experimentador intervenga activamente, modificando algún aspecto de la situación imaginada; b) no se encuentra necesariamente supeditada a una interpretación teórica, es decir, no necesita ser precedida por una teoría, aunque se requiere de una carga teórica mínima para su realización. Se ha mostrado que la visualización o representación en la imaginación de una situación no implica algo análogo a la observación científica y que el aspecto dinámico de estas representaciones no es equivalente a la intervención en la naturaleza. La multiplicidad de funciones de los experimentos mentales no implica que se trate necesariamente de experimentos, ya que los argumentos también pueden apoyar de diferentes maneras una hipótesis, mostrar inconsistencias, identificar problemas conceptuales en una teoría y realizar otras funciones que se han considerado características de los experimentos mentales.

El argumento en virtud del cual se establece una continuidad entre ambos tipos de experimentos es analógico y, como tal, tiene limitaciones. Una de sus principales falencias es que no consigue justificar que los experimentos mentales tengan “vida propia” (la posibilidad de reinterpretaciones, reequipamiento, etc.). Las posiciones que sostienen que estas reinterpretaciones pueden conducir a experimentos reales, fallan en tanto que si se asume que los experimentos mentales evolucionan en experimentos ejecutados, la hipótesis acerca de su valor evidencial es puesta en jaque ya que en estos casos los resultados se alcanzan a través de la contrastación empírica, y no en ausencia de intervención o realización

material. Por lo demás, no podemos emplear las mismas estrategias que en los experimentos reales para determinar la robustez de los resultados de los experimentos mentales.

Quizás la tesis de acuerdo con la cual los experimentos mentales son casos límite, pueda ser reinterpretada como la tesis de acuerdo con la cual son experimentos idealizados o modelos. De cualquier modo, las relaciones de parentesco entre experimentos mentales y reales son mucho más lejanas de lo que posiciones como las de Brown o Sorensen intentan sostener.

CAPÍTULO 3

Experimentos mentales y razonamiento científico

3.1 Introducción

La razón principal para considerar que los experimentos mentales no son esencialmente experimentos es que carecen de uno de los elementos determinantes de la experimentación: la intervención activa en el mundo natural. El carácter “mental” de estos dispositivos implica que por definición no emplean instrumental, no son ejecutados físicamente, ni proporcionan nuevos datos empíricos. En el contexto de la filosofía de la experimentación y en la filosofía general de la ciencia hasta finales del siglo XX, fueron tratados como argumentos teóricos que emplean hipótesis. Dado que estos razonamientos emplean idealizaciones e involucran abstracciones y que en muchos casos es imposible replicarlos en el laboratorio, fueron asimilados a inferencias y su papel en el contexto de justificación fue restringido al que los argumentos puedan cumplir.

Muchos experimentos mentales que desempeñaron alguna función importante en la historia de la ciencia, han sido reconstruidos como argumentos. En estos casos, es posible identificar un conjunto de hipótesis teóricas y supuestos auxiliares que operan como premisas de un razonamiento en el que se derivan ciertas consecuencias. Esta es una buena razón para considerar que los poderes epistémicos de los experimentos mentales son reducibles a los de las inferencias lógicas o incluso que son meros recursos retóricos. Esta idea ha motivado la tesis defendida principalmente por Norton (1991, 1996, 2004), de acuerdo con la cual su apariencia experimental puede ser eliminada sin pérdida

de su eficacia heurística. Algunas versiones más débiles de este punto de vista han sido elaboradas por Rescher (1991) e Irvine (1991).

La interpretación *argumental* de los poderes epistémicos de los experimentos mentales goza de una amplia aceptación entre los filósofos de la ciencia, debida fundamentalmente a sus ventajas explicativas. Este enfoque incorpora el funcionamiento de esta clase de experimentos a mecanismos ordinarios de adquisición de nuevo conocimiento y proporciona criterios de fiabilidad. Además, se apoya en la epistemología empirista a la que suscribe la interpretación filosófica de las prácticas experimentales. De acuerdo con esta perspectiva, el único fundamento incontrovertible del conocimiento en el dominio de las ciencias fácticas procede de la interacción del mundo físico con nuestros sentidos o con algún detector capaz de transmitirnos información. Por ello, la única explicación razonable del nuevo conocimiento que proporcionan los experimentos mentales exitosos es que sus premisas contienen información verdadera procedente de la experiencia, de cuya relación con algún principio teórico se derivada la conclusión.

Aunque la perspectiva argumental ofrece una explicación plausible del funcionamiento de los experimentos mentales, no permite dar sentido a la intuición de acuerdo con la cual los experimentos mentales son semejantes a la experimentación ordinaria en al menos tres aspectos importantes: se comunican a través de una narrativa experimental, dependen para su ejecución de un marco teórico relevante, y en su diseño se identifican variables entre las cuales es posible establecer relaciones de dependencia. Si bien la analogía entre experimentos reales y mentales tiene sus limitaciones, los escenarios imaginarios invocados en estos últimos desempeñan una función importante en la generación de creencias a partir de los cuales se evalúan las hipótesis. Asimismo, algunos casos históricos han desempeñado un papel importante en la elección de teorías, y el examen de algunos ejemplos sugiere que es posible diferenciar el experimento mental del argumento que constituye su interpretación teórica.

Estas cuestiones han motivado el desarrollo de perspectivas que, aunque otorgan a los mecanismos de inferencia un papel en la explicación de los poderes epistémicos de los experimentos mentales, han señalado que las actividades de representación involucradas son esenciales tanto para justificar las conclusiones alcanzadas como para persuadir a los experimentadores. De acuerdo con estas perspectivas, que denominaremos *inferencialistas*, la simulación de una actividad experimental en la imaginación cumple un papel decisivo en la evaluación de los principios teóricos, facilitando de esa manera una forma particular de inferencia (Claterbuck, 2013), proveyendo el contenido empírico de las premisas de un razonamiento, exhibiendo una paradoja (Sorensen, 1992) o proporcionando un contraejemplo (Haggqvist, 2009).

El propósito de este capítulo es presentar y evaluar un conjunto de tesis sobre la naturaleza de los experimentos mentales con las que las posiciones *argumentales e inferenciales* se comprometen. Señalaremos que los enfoques que asumen una interpretación *argumental* son reduccionistas y no ofrecen una explicación satisfactoria del tipo de evidencia que estos experimentos pueden proporcionar, ni de su eficacia heurística. Asimismo, mostraremos que la defensa de la tesis eliminativista del carácter experimental de los experimentos mentales limita sus funciones a la refutación de hipótesis y la ilustración de principios teóricos, lo que no permite dar sentido a las múltiples funciones que estos experimentos han desempeñado en la historia de la ciencia. Señalaremos también las limitaciones de la perspectiva *inferencialista*. Argumentaremos que, aunque esta explicación atribuye a la representación del escenario y la ejecución del experimento un papel en la generación de creencias, no proporciona una explicación suficiente de la manera en que éstos justifican las conclusiones.

A pesar de las deficiencias de la posición argumental, los esfuerzos que desde este punto de vista se han hecho para elaborar reconstrucciones de muchos experimentos mentales importantes resultan de suma utilidad para el estudio de estas prácticas. La identificación de las premisas y del mecanismo de

inferencia vinculados a un experimento mental es indispensable para el desarrollo de un estudio sistemático de estos, ya que, como se indicó en el Capítulo 1, presentan una mayor dependencia teórica que los experimentos ordinarios. Por ello, en este capítulo argumentaremos que la reformulación del razonamiento asociado a un experimento mental favorece la explicitación de supuestos auxiliares y principios teóricos que constituyen el trasfondo de cada uno de estos. La clarificación de estos supuestos y la reconstrucción de las interpretaciones teóricas en el contexto de las cuales se formularon podrían contribuir a evaluar su efectividad y elaborar criterios objetivos de confiabilidad.

3.2 Reconstrucción y esquemas de inferencia

El examen filosófico de las prácticas experimentales muestra que la interpretación teórica de un experimento se realiza mediante un argumento en el que se explicitan los principios teóricos, las hipótesis auxiliares y su relación con las observaciones. Desde este punto de vista, la interpretación de un experimento es siempre diferenciable de su realización material (Radder, 2006: 4). Algunas de las perspectivas filosóficas sobre el funcionamiento de los experimentos mentales consideran que, dado que en estos el elemento de la ejecución está ausente, no se realiza materialmente ningún procedimiento y no tiene lugar ninguna observación ni se introducen nuevos datos, la componente argumental es la única constitutiva de esta clase de experimentos. A pesar de su apariencia experimental, las descripciones de escenarios, instrumental, aparatos y acciones que el experimentador debe realizar, son premisas hipotéticas de un razonamiento que puede ser claramente identificado. En líneas generales, esta es la suposición básica de la denominada *perspectiva argumental*. Los filósofos que suscriben este enfoque entienden que los experimentos mentales en ciencias fácticas se sirven del conocimiento empírico almacenado (en ocasiones implícito) en el sistema de creencias del experimentador. La información empírica disponible es

reorganizada a través de un argumento que justifica la ampliación de las conclusiones obtenidas a otros casos o exhibe una inconsistencia que compele al abandono de un principio teórico o a su modificación sustancial.

Dentro de la línea argumental, algunos filósofos sostienen que el desarrollo de la experimentación mental en filosofía es previo al desarrollo de los experimentos mentales científicos y que su evolución recorrió carriles paralelos a los del desarrollo de los mecanismos de inferencia lógica. Rescher (1992), por ejemplo, argumenta que estas son herramientas metodológicas características del razonamiento especulativo y que su principal función es ilustrar hipótesis. El antecedente más antiguo de esta estrategia se encuentra en los presocráticos. Es posible reconocer en los escritos de estos filósofos algunas formas típicas de inferencia identificables con esquemas de razonamiento conocidos. Por ejemplo, para justificar la hipótesis de que la tierra ocupa un lugar central en el sistema planetario, equidistante a cualquier punto exterior, Anaximandro conjetura un escenario imaginario en el cual la negación de la hipótesis que se pretende demostrar es verdadera. (Rescher, 1992: 34). La forma más célebre quizás de este tipo de argumentos refutatorios, ampliamente usado en la experimentación mental típica de las ciencias fácticas desde la revolución científica, es la reducción al absurdo. El desarrollo de esta forma de inferencia por parte de los Pitagóricos fue empleado posteriormente como método formal en la prueba matemática (Rescher, 1992: 36). Para Rescher los experimentos mentales empleados en la ciencia actual responden a esquemas de razonamiento similares.

Otra interpretación dentro del conjunto de posiciones argumentales sostiene que aunque los experimentos mentales científicos son esencialmente argumentos, estos están en una relación privilegiada con el acervo de conocimiento empírico en las ciencias fácticas. Irvine (1991) considera que, aunque el empleo de escenarios imaginarios introducido por los presocráticos fue una importante contribución a la metodología científica, los experimentos mentales constituyen una sofisticada herramienta de la ciencia moderna que se

diferencia de estas formas iniciales de razonamiento contrafáctico. El tipo de razonamiento conjetural que emplearon los presocráticos, aunque guarda algunas semejanzas con el razonamiento abductivo (particularmente cuando se trata de una conjetura que se postula como explicativa de un fenómeno conocido), no puede ser equiparada a las estrategias que en la ciencia moderna constituyen los experimentos mentales. La principal diferencia entre este caso y los experimentos mentales en ciencias fácticas, es la dependencia, que estos últimos presentan, respecto de un trasfondo teórico y del acervo de observaciones y resultados experimentales que proporcionan una base de contrastación en la disciplina en cuestión. Por otra parte, en los experimentos mentales filosóficos está ausente el tipo de variación de una situación controlada que caracteriza a la experimentación real en la que es posible identificar un conjunto de variables correlacionadas. Este rasgo ha motivado la comparación entre experimento reales y mentales. Incluso cuando los experimentos mentales científicos emplean supuestos contrafácticos (como la ausencia de gravedad), estas variaciones se introducen para determinar su relación con principios teóricos bien establecidos. Irvine concluye que, a pesar de las ventajas que se derivan del tipo de exploración involucrada en los experimentos mentales (es decir, que no precisan de instrumental ni realización física), el alcance y la función de los experimentos mentales están restringidos por un lado por la disciplina dentro de la cual tienen lugar, que proporciona un cuerpo de conocimiento dentro del cual los supuestos adquieren sentido para evaluar una hipótesis, y por otro por la validez de la inferencia empleada. Asimismo, sus conclusiones están limitadas al examen de las consecuencias alcanzadas en un contexto teórico y observacional particular (Irvine, 1991: 162).

La posibilidad de identificar esquemas argumentales en los experimentos mentales implica, desde esta perspectiva, que estos pertenecen a un tipo de práctica de razonamiento científico diferente de los experimentos ordinarios. Norton (1996: 333) afirma que la única explicación razonable de cómo es posible que los experimentos mentales proporcionen conocimiento sobre el mundo

natural sin introducir datos nuevos y, en general, sin ningún tipo de intervención en él, es asimilando sus poderes epistémicos a los de los argumentos teóricos. Asumiendo, como otros representantes de esta posición, una interpretación empirista del fundamento del conocimiento en las ciencias fácticas, Norton entiende que los experimentos mentales reorganizan, por medio de un argumento, información que se encuentra disponible en el sistema de creencias del experimentador. Esta perspectiva supone que estos no son epistémicamente irrelevantes, aunque considera que sus principales funciones son negativas (refutatorias) o están restringidas a la exposición o ilustración de hipótesis. En síntesis los experimentos mentales en ciencias fácticas tienen, desde la perspectiva de Norton, tres características fundamentales:¹⁶

1. Proveen información acerca del mundo físico. Esta información no proviene de la presentación de datos empíricos nuevos, sino que es derivada, a través de inferencias, de información empírica previamente disponible.

2. Postulan estados de cosas hipotéticos o contrafácticos. Esta es una condición ligada al carácter mental de estos experimentos, pues, si no incluyeran escenarios contrafácticos, no serían experimentos mentales, sino meras descripciones de experimentos reales o estados de cosas reales.

3. Invocan particulares irrelevantes para la generalidad de la conclusión. Este aspecto es el que los hace similares a los experimentos reales.

La versión de la perspectiva argumental defendida por Norton se basa en el análisis detallado de una serie de experimentos mentales tomados la historia de la ciencia y en una serie de argumentos apoyados por el estudio de casos. Para simplificar la presentación de esta posición, que ha sido desarrollada

¹⁶ Norton (1991: 336) caracteriza a los experimentos mentales en función de las propiedades 2 y 3.

extensamente en una serie de artículos publicados a lo largo de más de veinte años, se presentan de manera sinóptica sus tesis y argumentos.

Es posible diferenciar en la versión de la perspectiva argumental de Norton, dos clases de tesis: por un lado, un conjunto de tesis referidas a naturaleza de los experimentos mentales, que denominamos tesis ontológicas, y por otro lado un grupo de afirmaciones acerca del funcionamiento de los experimentos mentales que denominamos tesis epistemológicas.¹⁷

En relación con las primeras, Norton (1991: 131) sostiene que: a) todos los experimentos mentales en ciencias fácticas pueden ser reconstruidos como argumentos cuyos esquemas de razonamiento son asimilables a formas de inferencia conocidas. Dado que los experimentos mentales pueden ser reconstruidos como argumentos, Norton concluye que b) son esencialmente argumentos encubiertos. De esta afirmación se sigue que c) su apariencia experimental es un disfraz de su naturaleza argumental y que como tal puede ser eliminada sin pérdida de su eficacia heurística. La primera de las tesis ontológicas puede ser identificada como la “tesis de la reconstrucción” (Norton, 1996: 339); la segunda como “tesis de la identidad” (Norton, 2004a: 49) y la tercera como “tesis de la dispensabilidad” (Norton, 2004b: 1139). Esta última afirmación está emparentada con las tesis epistémicas acerca de las funciones que los experimentos mentales pueden desempeñar y acerca del fundamento de su poder evidencial. Aunque Norton no diferencia estas tres tesis, su explicitación permite aclarar algunos de sus argumentos.

¹⁷ Gendler (1998: 398) identifica en el trabajo de Norton la tesis de la eliminación, de acuerdo con la cual cualquier conclusión alcanzada a través de la ejecución de un experimento mental puede ser obtenida a través de un argumento. Esta tesis tiene dos lecturas: una débil, que Gendler denomina tesis de dispensabilidad, de acuerdo con la cual un experimento mental (exitoso) puede ser reemplazado sin perder fuerza demostrativa por un argumento que no involucre la narrativa experimental; y una fuerte, que interpreta como la tesis de la derivabilidad, de acuerdo con la cual la fuerza demostrativa de un experimento mental solo puede ser explicada por el hecho de ser reemplazada por un argumento (Gendler, 1998: 401-402).

Respecto de la función de los experimentos mentales, Norton afirma que *no llevan a cabo ninguna magia epistémica* en tanto nos informan acerca del mundo. Simplemente, aprovechan lo que ya sabemos y transforman este conocimiento por medio de una argumentación encubierta. Epistemológicamente, los experimentos mentales no pueden hacer más de lo que hace una argumentación:

A pesar de su exótica reputación los experimentos mentales nos convencen a través de métodos bastante prosaicos. Llegan a nosotros en palabras en un papel. Las leemos y seguimos los pasos para completar el experimento. Nos convencen sin experiencias exóticas, momentos bíblicos o estados mentales extáticos. A este nivel de descripción los experimentos mentales no difieren de la literatura en general. Una larga tradición de lógica informal mantiene que esta actividad es mera argumentación y que la mayoría de nosotros tiene alguna facilidad natural para ella. (Norton 2004a: 51)

La tesis de Norton acerca de la estructura y funciones de los experimentos mentales se aplica tanto al contexto de descubrimiento como al de justificación de las hipótesis y teorías científicas y establece que si los experimentos mentales pueden ser reconstruidos como argumentos, entonces: i) la ejecución de un experimento mental es la presentación de un argumento y ii) su alcance epistemológico coincide con el de un argumento (Norton 2004a: 50). Esta tesis puede ser interpretada como la “tesis de la reductibilidad” y se apoya en el supuesto empirista, que para Norton constituye la epistemología aceptada por default en la filosofía de la ciencia contemporánea, de acuerdo con el cual todo conocimiento del mundo físico tiene su origen en nuestra experiencia (Norton, 1991: 129; 1996: 335). Esta concepción empirista del conocimiento está presupuesta en todos los trabajos de Norton. Con esta se vincula una “tesis sobre la confiabilidad epistémica”, de la que se deriva un criterio para establecer la confiabilidad de los experimentos mentales (Norton, 2004: 52). Desde esta perspectiva, los experimentos mentales no proporcionan conocimiento nuevo en sentido estricto porque no presentan nuevos datos, pero permiten extraer nuevas conclusiones del conocimiento empírico previamente almacenado por medio de

un argumento identificable. Aunque dicho argumento suele estar implícito en la narrativa del experimento mental, su reconstrucción permite evaluar la confiabilidad del caso. Si las premisas proporcionan evidencia suficiente para apoyar la conclusión y si la inferencia empleada es válida, el experimento es confiable. En los casos en cuales que la inferencia empleada es inductiva, Norton se expresa de manera ambigua respecto de la noción de validez pero en general considera como válidos a los razonamientos que apelan a algunos principios generales de inducción.

Norton proporciona argumentos independientes para cada una de estas tesis. Para apoyar la tesis de la reconstrucción, examina una serie de casos históricos e identifica formas de inferencia típicas en cada uno de ellos. Del hecho de que puedan ser reconstruidos como argumentos, se sigue para Norton que los experimentos mentales son esencialmente argumentos. Para el experimento de los cuerpos en caída libre de Galileo, proporciona la siguiente reconstrucción:

1. Supuesto para la prueba por reducción al absurdo: la velocidad de caída de los cuerpos en un medio dado depende de su peso.
2. De 1 se sigue que si una piedra grande se mueve a ocho grados, una piedra pequeña de la mitad de su peso, se moverá a una velocidad de cuatro grados.
3. Supuesto: si una piedra grande en caída libre es conectada a una pequeña, la pequeña retrasará a la más grande y la grande acelerará a la pequeña.
4. De 3: si dos piedras en caída libre son conectadas, el compuesto caerá más lentamente que ocho grados de velocidad.
5. Supuesto: el compuesto de dos pesos tiene un peso mayor que el cuerpo de mayor peso aisladamente.
6. De 1 y 5: El compuesto caerá a una velocidad mayor que 8 grados.

7. Las conclusiones 4 y 6 se contradicen
8. Rechazamos el supuesto 1, porque implica una contradicción
9. Conclusión: todos los cuerpos caen con la misma velocidad independientemente de su peso.

Se trata de un típico argumento por reducción al absurdo. Muchos experimentos mentales son instancias de esta forma de argumento. La termodinámica presenta algunos de los casos más conocidos. Esto se debe a que sus leyes pueden formularse como aserciones acerca de imposibilidades y una forma simple de derivar consecuencias de estas aserciones es a través de argumentos por reducción al absurdo.

En apoyo a la tesis de la dispensabilidad, Norton argumenta que los particulares involucrados en la narrativa de un experimento mental son eliminados en la conclusión. En los argumentos por reducción al absurdo como el de Galileo, la descripción de la situación experimental ilustra un contraejemplo a una afirmación universalmente cuantificada, de cuya contradicción se sigue la conclusión (Norton 1991: 132). El experimento de Galileo emplea un tipo de inferencia deductiva y pertenece a la clase de experimentos mentales de tipo 1. Los argumentos en los cuales la generalidad de la conclusión depende de algún tipo de inferencia inductiva, en virtud de la cual los particulares invocados en la descripción del experimento son eliminados, son experimentos mentales tipo 2.¹⁸

Aunque la tesis empirista sobre el origen del conocimiento en los experimentos mentales es independiente de la interpretación argumental, Norton no formula explícitamente un argumento para apoyarla. Sin embargo, sostiene que junto con la tesis de la reconstrucción es la única que da sentido al nuevo conocimiento que los experimentos mentales pueden proporcionar. Su posición

¹⁸La clasificación de Brown de experimentos destructivos y constructivos (Brown 2011: 32) se corresponde con la distinción de Popper (1959: 516) entre experimentos críticos y heurísticos, y la de Norton (1991: 131) descritas en el capítulo 1 de este trabajo entre experimentos de Tipo 1 y de Tipo 2.

en este punto es que solo la experiencia es fuente de conocimiento del mundo físico, por lo tanto, si los experimentos mentales producen conocimiento sobre el mundo físico entonces este procede de nuestro contacto con él. Este conocimiento puede estar implícito, almacenado en la memoria o tener la forma de intuiciones desarrolladas evolutivamente y partir de este acervo es posible obtener nuevo conocimiento mediante la reformulación o reorganización o eliminación de un principio teórico a través de una inferencia lógica.

En relación con la tesis de la reductibilidad, Norton afirma que la apelación a patrones de inferencia lógica como determinantes de la estructura de los experimentos mentales permite proporcionar una explicación satisfactoria de sus poderes epistémicos. Los esquemas de inferencia deductivos permiten preservar la verdad y las inferencias inductivas la probabilidad (Norton 2004b: 1143) y con ello garantizar la verdad de las conclusiones¹⁹. La apelación a medios exóticos de conocimiento, según Norton, no permite proporcionar criterios para distinguir a los experimentos exitosos de los fallidos. Así por ejemplo, en el experimento mental de Galileo, Brown (2004a: 40) atribuye a la intuición racional la transición desde la identificación de la contradicción que se sigue de la presentación del ejemplo hasta el principio de equivalencia. Para Norton, en cambio, el análisis del experimento depende de la identificación de un supuesto tácito: la velocidad de caída de los cuerpos es directamente proporcional a su peso. Si se admite esta hipótesis, entonces, el paso de la refutación de la teoría aristotélica a la teoría de

¹⁹Norton (2004a: 53 y 2004b: 1143) afirma que los experimentos mentales explotan patrones de razonamiento que pueden ser incluidos en los esquemas de inferencia lógica conocidos. De ese modo, los argumentos empleados en los experimentos mentales pueden ser reconstruidos como deductivos o inductivos. Este marco permite evaluar las inferencias empleadas en los experimentos mentales según las prescripciones de la lógica. En primer lugar, las inferencias pueden ser válidas o inválidas. En el caso de los argumentos inductivos, Norton se refiere vagamente a principios vinculados a inferencias inductivas, como la "preservación de la probabilidad". Entendemos que se refiere al tipo de probabilidad inductiva con la cual dadas ciertas premisas que se suponen verdaderas, un razonamiento inductivo apoya la conclusión. De manera independiente a su teoría sobre el funcionamiento de los experimentos mentales Norton elabora una teoría material de la inducción. Sostiene (Norton, 2003: 650) que no hay principios generales de inducción y que cada caso debe ser evaluado particularmente ya que las inferencias inductivas derivan su legitimidad de los hechos pertinentes a la inducción en cuestión. Norton no aplica esta teoría de la inducción en la reconstrucción de experimentos mentales.

Galileo es una simple inferencia. Desde su punto de vista la hipótesis más viable es que la velocidad de la caída es una función monótona creciente de su peso (Norton, 1993: 342). Cualquier otra opción permitiría que los cuerpos más pesados cayeran más rápido que los más livianos. La teoría de Galileo admite la posibilidad de que el incremento en la velocidad de caída de los cuerpos sea una función constante. Cualquier hipótesis diferente de la de Galileo, implica una contradicción. Por otra parte, las inferencias lógicas involucradas en la formulación de los experimentos mentales exitosos se encuentran dentro de los esquemas lógicos conocidos. No hay ningún mecanismo extraordinario involucrado en la obtención de conclusiones a través de la formulación de experimentos mentales. Y en el caso de que así fuera, afirma Norton: “se puede confiar en que la lógica evoluciona para abrazar nuevas formas de argumentación que puedan emerger de los esfuerzos creativos de los experimentadores” (Norton 2004a: 55).

Finalmente, en relación con la tesis de la confiabilidad epistémica, Norton, señala que si los experimentos mentales pueden emplearse de manera exitosa para apoyar una hipótesis, esto se debe a que su funcionamiento coincide con el de un argumento. Para respaldar esta idea propone un *test* de confiabilidad que consiste en presentar pares de experimentos y contra-experimentos mentales en los que el escenario es el mismo pero la conclusión que se deriva de ellos es opuesta (Norton 2004a: 51). Para Norton, esto se debe a que en uno de los casos el resultado es espurio. Al reconstruirlos como argumentos es posible identificar la premisa falsa o la inferencia inválida responsable de la conclusión o el resultado equivocado. Si los experimentos mentales pueden utilizarse de manera epistemológicamente confiable, entonces deben ser argumentos.

En suma, las posiciones agrupadas bajo el rótulo de *perspectivas argumentales* comparten dos supuestos fundamentales. En primer lugar, asumen que la única fuente de conocimiento del mundo natural es la experiencia sensible. En segundo lugar, suponen que la posibilidad de reformular experimentos

mentales como argumentos permite concluir que los primeros son asimilables a los segundos y que sus poderes epistémicos se restringen a los de éstos últimos. De estos dos supuestos Norton infiere que el escenario experimental es un elemento heurístico accesorio que puede reemplazarse o eliminarse en la reconstrucción argumental de un experimento mental sin que este pierda su fuerza persuasiva. Esta última afirmación es el principal objeto de las críticas. Las posiciones de Irvine y Rescher no defienden explícitamente las tesis ontológicas sobre la naturaleza de los experimentos mentales formuladas por Norton, pero de su examen de las características epistémicas de esta forma de argumentación parece sugerir que adscriben en algún grado a esta idea. Rescher afirma que los experimentos mentales consisten en procedimientos de razonamiento hipotético en los que se asume como verdadera una premisa que podría ser falsa con el objeto de derivar una conclusión que resulte explicativa de un fenómeno o hecho intrigante (Rescher 1991: 31). Para Irvine la posibilidad de identificar premisas y mecanismos de inferencia es un signo del carácter argumental de los experimentos mentales. No obstante, su confiabilidad epistémica no está limitada por el tipo de argumento y la verdad de las premisas, sino por el contexto disciplinar dentro del cual tienen lugar (Irvine, 1991: 150). A continuación se intenta mostrar que, aunque Norton elaboró un arsenal argumentativo para defender la plausibilidad de su posición, su principal defecto consiste en desestimar el elemento *representacional* relegándolo al lugar de la heurística.

3.3 Dispensabilidad y reductivismo

La caracterización de los experimentos mentales proporcionada por la perspectiva argumental conduce, al menos en su versión más tenaz (la de Norton), a una concepción *deflacionaria* de sus poderes epistémicos y a una noción parsimoniosa de su naturaleza. Aunque estas pueden considerarse ventajas de una teoría del funcionamiento de los experimentos mentales, la simplicidad en

este caso puede verse como pobreza explicativa. Esta elucidación carece de la explicitación de la diferencia específica, lo que propiamente define a los experimentos mentales. Si bien la posición argumental arguye que no existe tal elemento o, más precisamente, que dicho elemento es un artificio que el examen filosófico puede desenmascarar, el reduccionismo de esta explicación elimina dos características de los experimentos mentales. Por un lado, esta interpretación no explica por qué la ejecución de un experimento mental es un proceso cuya eficacia heurística es mayor que la de un argumento. Por otro lado, la perspectiva argumental no da sentido al hecho de que el papel que los experimentos mentales han desempeñado en la elección de teorías es similar al de los experimentos ordinarios.

Si bien la semejanza de los experimentos mentales con los argumentos que emplean supuestos hipotéticos es notable, de ello no se sigue que el razonamiento en cuestión sea el único mecanismo involucrado en la obtención de nuevo conocimiento. La posibilidad de detectar una inferencia involucrada en la interpretación teórica es algo que los experimentos mentales tienen en común con los reales. La descripción de los procedimientos experimentales puede ser incorporada como parte de las premisas de un argumento pero esto no es indicio de que el experimento sea en sí mismo un argumento, es decir, puede argüirse con las mismas razones que todos los experimentos son argumentos. Sorensen (1992: 214) llama la atención sobre este hecho y señala que la carga de la prueba del argumento reductivista podría invertirse si se reformulase de este modo: si los experimentos mentales son entimemas, entonces, todos los experimentos son entimemas.

La tesis de la reductibilidad de acuerdo con la cual cualquier conclusión alcanzada por un experimento mental exitoso, puede alcanzarse por un argumento, involucra, como se ha señalado, dos afirmaciones subsidiarias: una sobre el contexto de descubrimiento y otra sobre el contexto de justificación. En relación con esta última, Norton afirma que el tipo de conocimiento proporcionado

por los experimentos mentales no constituye una vía de acceso privilegiada ni es superior a las formas de inferencia ordinarias. De acuerdo con su punto de vista, solo la perspectiva de Brown rivaliza con esta interpretación. Esta es la única alternativa que explícitamente apela a un proceso de formación de creencias extraordinario, la intuición racional. Pero esta explicación resulta injustificada en virtud de que, desde el punto de vista argumental, los mismos resultados pueden obtenerse a través de argumentaciones, por lo que Norton asegura que su postulación es innecesaria. Todas las demás posiciones que explican el funcionamiento de los experimentos mentales se circunscriben a la epistemología empirista y pueden ser asimiladas a la perspectiva argumental, fundamentalmente porque, para Norton, la reconstrucción argumental ofrece la única interpretación clara de cómo se obtiene nuevo conocimiento y es la única capaz de proporcionar criterios de fiabilidad.

Aunque los argumentos de Norton son poderosos no son suficientes para negar que existan mecanismos de reestructuración de creencias que no sean reducibles a las inferencias lógicas. La imputación de limitaciones a las explicaciones alternativas existentes depende de la aceptación de la tesis empirista tal como la formula Norton y del supuesto de que solo los argumentos pueden proporcionar un esquema para explicar el nuevo conocimiento en cuestión. Consideremos otra vez el experimento de Galileo de los cuerpos en caída libre. Si se retoma la reconstrucción de Norton, puede argumentarse que la consideración de la situación contrafáctica está involucrada de manera esencial en el razonamiento. En el argumento, la representación del escenario hipotético cumple un papel esencial en la consecución de la demostración. Aunque desde el punto de vista de Norton este rasgo podría explicarse como una contribución heurística, i.e. la narrativa experimental facilita la realización de una inferencia, es el ejemplo de dos cuerpos de diferente peso unidos en caída libre el que permite representarse el contraejemplo. A pesar de que estas observaciones no constituyen por sí mismas un argumento contra la tesis de la dispensabilidad,

motivan la sospecha sobre la idea de que las conclusiones a las que se arriba mediante experimentos mentales no están racionalmente justificadas solo en virtud de una inferencia.

Gendler (1998) defiende, contra la perspectiva argumental, la indispensabilidad de las actividades de visualización y modelización propias de los experimentos mentales para la obtención de conclusiones justificadas. Argumenta que las conclusiones del experimento de los cuerpos en caída libre de Galileo no están garantizadas por una inferencia identificable. De acuerdo con su punto de vista, la reconstrucción argumental no logra captar la fuerza demostrativa del experimento porque los supuestos atribuidos a la perspectiva aristotélica, tal como son formulados en la reconstrucción del argumento, podrían ser reinterpretados de manera que no se produzca la contradicción en virtud de la cual Norton supone que se realiza la reducción al absurdo de la tesis aristotélica. La interpretación argumental del experimento supone que el aristotélico está comprometido con las siguientes premisas:

- (1) La velocidad natural es mediativa.
- (2) El peso es aditivo.

De acuerdo con Gendler, la reconstrucción de Norton depende del supuesto de que, en los casos en los cuales el cuerpo en consideración está compuesto por dos objetos, el peso del compuesto es la suma de los pesos de los cuerpos que lo componen. Pero la velocidad natural del cuerpo compuesto no se obtiene de adicionar las velocidades de los cuerpos sino que es un promedio de las velocidades naturales de los cuerpos que componen este cuerpo combinado.

De las premisas 1 y 2 se sigue que:

- (3) La velocidad natural no es directamente proporcional al peso.

Dado que una propiedad mediativa no puede ser directamente proporcional a una que es aditiva, para Gendler esta reconstrucción sugiere un número de hipótesis que permitirían al aristotélico “escapar de la contradicción” (Gendler, 1998: 404). Una manera de bloquear la inferencia es suponer que las premisas 1 y 2 no se aplican al caso al que se hace referencia en el experimento (un compuesto de dos cuerpos del mismo material pero diferente peso), lo que podría expresarse de la siguiente manera:

(4) La velocidad natural no está físicamente determinada para un compuesto (de dos cuerpos que caen con diferentes velocidades).

(5) El peso no está físicamente determinado para un compuesto (de dos cuerpos de distinto peso).

Estas alternativas permiten rechazar las hipótesis 1 y 2 porque estas presuponen que la velocidad y el peso son propiedades que se aplican universalmente. Otra manera de evadir la conclusión absurda es introducir una hipótesis acerca de la naturaleza de los cuerpos compuestos, que determine la velocidad de acuerdo a cómo se identifique este compuesto:

(6) La velocidad es mediativa para compuestos en los que los cuerpos están unidos pero permanecen como dos cuerpos independientes y aditiva para compuestos donde los cuerpos están unificados (se comportan como un solo cuerpo).

De acuerdo con esta reconstrucción, lo único que el aristotélico puede ser forzado a aceptar es que de su posición se sigue que hay discontinuidades radicales en la naturaleza. Pero el aristotélico puede incluso escapar a esta imputación formulando un principio que determine una manera de calcular la velocidad de un compuesto en virtud del grado de conexión entre los cuerpos que

lo componen. Esta hipótesis haría posible mantener el principio según el cual la caída es proporcional al peso.²⁰

Para Gendler, lo que bloquea las hipótesis alternativas en la presentación experimental del caso de Galileo, es la activación de conocimiento empírico disponible. Los experimentadores “saben”, por su experiencia del mundo, que el peso y la velocidad son propiedades físicamente determinadas, pero la entificación no es una propiedad que pueda ser físicamente establecida. Al evocar conocimiento tácito sobre como caen los cuerpos, el experimento captura un conjunto de suposiciones empíricamente revisables que se presentan como principios inamovibles. De esto depende su fuerza persuasiva.

Según Gendler, la reconstrucción argumental no permite capturar estos supuestos. Incluso en este caso, si las hipótesis implícitas acerca de los fenómenos del mundo fueran especificadas de la siguiente forma:

(8) La velocidad natural y el peso están físicamente determinados.

(9) La entificación no está físicamente determinada.

Se obtendría una formulación aproximada de supuestos empíricamente revisables. Pero, tal como están formuladas, sin embargo, son afirmaciones demasiado fuertes. Lo que el experimento logra es exhibir el compromiso con ciertas tesis que no se presentan como revisables para el caso en cuestión (Gendler, 1998: 407).

El experimento mental posee una ventaja epistémica sobre el argumento porque las alternativas aristotélicas, aunque lógicamente posibles, están “bloqueadas” por el acervo de conocimiento empírico que la representación

²⁰ Para los cuerpos completamente unidos vale el principio aditivo (2) y la velocidad está determinada por la suma de los pesos, para los cuerpos conectados vale el principio mediativo (1) y la velocidad es una función del grado de conexión entre los dos cuerpos.

mental pone en acción (Gendler, 1998: 404). Esta conclusión acerca del experimento de Galileo puede extenderse a todos los experimentos mentales en ciencia. De acuerdo con Gendler, el caso de estudio “presenta un fenómeno familiar de manera novedosa” produciendo simultáneamente una reforma en el sistema conceptual del experimentador (Gendler, 1998: 412). Las conclusiones del experimento de Galileo están racionalmente justificadas, algo que la reconstrucción argumental es incapaz de hacer. La representación de la situación experimental muestra que la velocidad de la caída de los cuerpos no es una función del peso. Lo que el experimento logra hacer es reconfigurar el espacio conceptual. Previo al experimento no había, en el esquema conceptual aristotélico, espacio para considerar que la velocidad de la caída podría ser independiente del peso de los cuerpos. En palabras de Gendler:

Si la entificación es arbitraria y la velocidad y el peso naturales son fijados por el mundo, entonces una noción de velocidad natural dependiente de las características es simplemente incoherente. Así que una forma de pensar sobre cómo funciona el experimento mental es ésta: hace que el aristotélico reconozca la insuficiencia de su marco conceptual para tratar fenómenos que, a través de la contemplación de este caso imaginario, llega a reconocer como siempre parte de su mundo. (Gendler, 1998: 412)

Por otra parte la tesis sobre el contexto de descubrimiento es una tesis psicológica y como tal, su defensa no compete, según Norton, a la filosofía sino a la psicología. La tesis de la reductibilidad establece que la ejecución de un experimento mental es idéntica a la realización de un argumento. No obstante, es posible argüir que la representación de un estado de cosas hipotético es una actividad relevante para la generación de una creencia en base a la cual se realiza la evaluación de una hipótesis. Para defender la perspectiva argumental, sería necesario explicar de qué manera se recupera la información necesaria para arribar a la conclusión y, además, sostener que la inferencia en cuestión se realiza de manera inconsciente.

El argumento de Gendler depende de una interpretación filosófica que supone que el papel de los experimentos mentales debe ser evaluado dentro de “cuerpos vivos de conocimiento” (1998: 400) y que su función decisiva es hacer accesible información que no se encontraba proposicionalmente disponible en el esquema conceptual de los experimentadores. Esto supone una concepción filosófica de raíz khuniana de la racionalidad científica, del desarrollo científico y del cambio teórico, que Norton no estaría probablemente dispuesto a aceptar. A pesar de los diferentes puntos de partida de ambas interpretaciones, que las harían en cierto sentido inconmensurables, Gendler acierta al señalar un punto débil en la concepción argumental: la identificación de (todas) las premisas relevantes para la inferencia involucra en algún sentido reconstruir el contexto de supuestos en el cual el experimento es presentado como evidencia a favor o en contra de una tesis. Esta conclusión apunta contra uno de los elementos más importantes de la perspectiva argumental. Si los poderes justificatorios de los experimentos mentales residen fundamentalmente en sus argumentos, los experimentos mentales exitosos son exitosos independientemente de los contextos en los que aparecen (Norton, 1996: 348) ya que su eficacia depende del contenido empírico de las premisas y de la inferencia en cuestión. Esto parece necesario para sostener la afirmación según la cual la reconstrucción argumental tiene la misma fuerza persuasiva que el experimento.

Por otra parte, la reconstrucción de Gendler llama la atención sobre una característica importante de todo experimento mental: su efectividad depende de la presentación en este experimento de Galileo, de un caso anómalo. Desde la perspectiva de Gendler, la contemplación guiada de un escenario es la que mina los supuestos de fondo en la perspectiva aristotélica y exhibe la dificultad de aplicar el marco conceptual presupuesto al caso dado. Este rasgo resulta crucial para caracterizar la metodología empleada en los experimentos mentales. Estos se caracterizan por el empleo de un caso para derivar, de su relación con una conjetura, alguna conclusión. Para Norton, no obstante, los particulares aludidos

no son esenciales para extraer la conclusión de un experimento mental. Las peculiaridades del escenario solo son relevantes en cuanto ilustran algún principio general.

3.4 Inferencialismo

Una de las principales ventajas de adoptar una perspectiva argumental es que la identificación de un conjunto de premisas y de un esquema de inferencia proporciona un criterio de confiabilidad. En palabras de Norton: “los resultados (de los experimentos mentales) son confiables en tanto los supuestos son verdaderos y la inferencia es válida” (Norton, 1996: 336). Desde la perspectiva argumental, la lógica necesaria para reconstruir el razonamiento implícito en experimentos mentales es relativamente simple. Las inferencias deductivas ocultas en experimentos mentales exitosos como el de Galileo preservan la verdad de las premisas, las inferencias inductivas presentes en experimentos preservan la probabilidad.

Norton señala que la marca que permite identificar a los experimentos mentales exitosos debe ser un rasgo interno, en tanto su corrección no puede atribuirse ni a las intenciones de quien diseñó el experimento ni al contexto en que aparecen. Dado que el contenido de las premisas lo proporcionan los hechos del mundo externo, es la inferencia lógica lo que permite distinguir los experimentos exitosos. Aunque no circunscribe los experimentos exitosos a formas inductivas y deductivas, por ejemplo, el experimento del balde giratorio de Newton puede ser reconstruido como una inferencia a la mejor explicación que introduce una premisa falsa (Norton, 1996: 348), Norton considera que el repertorio de experimentos mentales científicos disponibles no hace uso de ninguna lógica exótica. Esto fundamentalmente por razones históricas: los experimentos discutidos actualmente en ciencia se parecen a los realizados por Mach y apelan a una lógica relativamente simple.

A pesar de su simplicidad, la norma inferencial propuesta por Norton no captura un rasgo esencial de los experimentos mentales. La contemplación de un escenario imaginario desempeña, sin duda, algún papel en el examen de la hipótesis bajo escrutinio, al menos en la aceptación de ciertas premisas fundamentales para inferencia. Por otra parte, el criterio de confiabilidad, tal como fue formulado más arriba, resulta limitado para la explicación del funcionamiento de una forma típica de los experimentos mentales en física, en los cuales se emplea un contraejemplo para mostrar, simultáneamente, la inadecuación de un principio teórico y la plausibilidad de una perspectiva alternativa. Las posiciones que denominamos *inferencialistas* intentan proporcionar un esquema para explicar el funcionamiento de los experimentos mentales que le otorgue un lugar a la representación del escenario imaginario y a las actividades que incluyen experimentar hipotéticamente. Las propuestas que, dentro de esta línea, formalizan los razonamientos en cuestión apelando a la lógica modal intentan dar sentido al empleo de conjeturas falsas en la justificación de principios teóricos.

Goffi y Roux (2011) exploran una interpretación de los experimentos mentales en términos de la noción de paradigma de Kuhn. Desde este punto de vista, el principal rasgo que los experimentos mentales exitosos presentan es que logran convencer a un interlocutor. Debido a que este es su objetivo, se parecen lo suficiente a las prácticas argumentales como para asumir que son esencialmente argumentos (aunque esto no implica suscribir las tesis de la identidad y dispensabilidad elaboradas por Norton, ni la tesis epistémica de la reductibilidad). Esta posición explora la idea de acuerdo con la cual los experimentos mentales son un método eficaz para exhibir inconsistencias en un esquema conceptual. Esta tesis es reinterpretada de manera que no gravita sobre las hipótesis acerca del desarrollo del conocimiento científico ni los procesos de adquisición de conceptos invocados en el contexto de la filosofía de Kuhn. La perspectiva de Goffi y Roux asume que los experimentos mentales ponen a prueba creencias

introduciendo una hipótesis que produce un conflicto en el sistema de creencias del interlocutor o lector del experimento mental.

Desde esta perspectiva, los experimentos mentales exitosos funcionan de manera análoga a los test de consistencia²¹, revelando una contradicción entre un conjunto especificado de creencias jerarquizadas y orientando al experimentador hacia el rechazo del elemento más débil con el fin de eliminar la contradicción (Goffi y Roux, 2011: 170). Aunque desde este punto de vista el resultado del experimento mental es el producto de una inferencia, la representación de una situación hipotética no es irrelevante en este proceso. El objetivo de los experimentos mentales es reconfigurar una región específica de las creencias del experimentador, por lo que, en este aspecto, se parecen a los experimentos reales, aunque la contrastación mediante la experiencia es sustituida por la reflexión sobre un escenario contrafáctico. La diferencia entre los experimentos mentales y reales reside, entonces, en la manera en que una creencia es contrastada por un experimento. Los experimentos mentales lidian con nuestras creencias, con un conjunto de supuestos compartidos que son el trasfondo del argumento en cuestión. Contra la interpretación machiana de acuerdo con la cual, la base de contrastación de los experimentos mentales es el conocimiento empírico almacenado pero proposicionalmente no articulado, Goffi y Roux (2011: 177) consideran que este fundamento se comprende mejor como un conjunto de creencias jerárquicamente organizadas, con el fin de evitar el misticismo y proporcionar una interpretación más elaborada de los argumentos en cuestión. Estos supuestos más o menos sofisticados proceden del intercambio con el mundo y su interpretación desde algún trasfondo teórico.

De esta caracterización general se derivan tres condiciones que conjuntamente conforman un criterio para identificar los experimentos mentales

²¹ Goffi y Roux no se refieren a pruebas de consistencia en el sentido formal, sino, más bien, a un análisis de la consistencia de un sistema de creencias con el fin de detectar posibles inconsistencias.

efectivos: i) ponen a prueba creencias (no tratan acerca de objetos, relaciones de objetos o intuiciones), ii) requieren que los interlocutores estén de acuerdo acerca del sistema de creencias (o la porción del sistema de creencias relevante en cada caso) acerca del cual trata el experimento; y iii) este sistema de creencias debe poseer una jerarquía claramente definida en función de la cual el experimento sugiere que uno de los elementos sea eliminado para preservar la consistencia.

El experimento de Galileo es un caso que ejemplifica claramente esta concepción. En su parte destructiva, se dirige a rechazar la suposición aristotélica de acuerdo con la cual:

(a) Dos cuerpos en caída libre en un mismo medio material caen con velocidades proporcionales a sus pesos.

Y a reemplazar este principio por la suposición galileana, según la cual:

(a') Dos cuerpos en caída libre en un mismo medio material, *ceteris paribus*, caen con la misma velocidad, cualquiera sea su peso.

En un contexto donde (a) es una ley de la naturaleza, se sigue que un cuerpo más pesado cae a una velocidad mayor que un cuerpo más liviano. De acuerdo con el análisis de Goffi y Roux en el diálogo de Galileo se supone que un aristotélico aceptaría como consecuencia de este principio las siguientes premisas:

(A₁) Si un cuerpo pesado es conectado a un cuerpo más liviano la velocidad natural es mediativa (es decir, en este caso la velocidad es mediativa).

(A₂) El peso es aditivo.

De lo que se sigue que la velocidad natural del compuesto es menor que la velocidad natural del cuerpo más pesado y el peso de los cuerpos combinados es

mayor que la del primero tomado aisladamente. De esta contradicción se infiere la negación de (a): dos cuerpos no caen con velocidades proporcionales a sus pesos. Luego, si se asume:

(A3) el peso es el único parámetro a tener en cuenta cuando se examinan las propiedades de los cuerpos en caída libre.

Es posible trascender a (a').

El argumento en cuestión es una *reductio*. En la reconstrucción del experimento ofrecida, aristotélicos y galileanos comparten los supuestos (A1), (A2) y (A3). (A1) se deriva de (a) y dado que los dos juntos conducen a una contradicción renunciar a (a) es la manera menos "costosa", es decir, la que menos ajustes de creencias requiere, de resolver la contradicción.

Aunque los defensores de esta perspectiva declaran no avalar el reductivismo de la tesis eliminativista de Norton (Goffi y Roux, 2018: 505), el papel que la representación de un escenario y de la ejecución del experimento mental desempeña en la detección de una inconsistencia no es completamente claro. La que denominan una "interpretación pragmática" del funcionamiento de los experimentos mentales puede interpretarse como una explicación del papel psicológico que el escenario hipotético desempeña en la derivación de una contradicción.

Los experimentos mentales invitan al experimentador a buscar el eslabón más débil de su cuerpo de creencias. Es en el orden de las creencias, no necesariamente percibidas antes de que se llevara a cabo el experimento mental, que el experimento mental hará que el eslabón más débil sea obvio. (Goffi y Roux, 2011: 170)

La representación en cuestión puede facilitar la aceptación de la premisa (A1), que se deriva de (a), pero el escenario invocado no justifica el abandono de esta creencia. Dado que se supone una relación jerárquica entre las creencias, es

en virtud de este vínculo²² que los experimentadores deciden cuál de los principios debe ser abandonado. Asimismo es la versión argumental del experimento la que permite vislumbrar tal relación. Por lo que esta posición no parece integrar en la explicación el elemento de visualización o manipulación de un escenario imaginario como elemento indispensable en la obtención de nuevo conocimiento ni de la eficacia heurística de los experimentos mentales.

Una propuesta explicativa que hace de la representación un elemento esencial de la ejecución del experimento, apunta al contenido empírico de las premisas del argumento. Para Brendel (2004) aunque los experimentos mentales pueden ser reconstruidos como argumentos, el fundamento del nuevo conocimiento y de la eficacia heurística de estos son las intuiciones que conforman el conocimiento de fondo del experimento. Según Brendel:

Si de hecho siempre hay un tipo complejo de argumentación en nuestras mentes cuando llevamos a cabo un experimento mental, las conclusiones sin esfuerzo, rápidas y muy a menudo compartidas de los experimentos mentales serían difíciles de explicar. Creo que es más plausible suponer que algunas premisas ocultas funcionan como conocimiento de fondo, y que no es necesario que se activen explícitamente cuando realizamos un experimento mental. Parte de este conocimiento de fondo consiste en intuiciones. (Brendel, 2004: 96)

Las intuiciones son “actitudes proposicionales mentales que van acompañadas por un sentimiento de certeza” (Brendel, 2004: 96). Se trata de supuestos adquiridos *a posteriori* que pueden ser identificados en la reconstrucción argumental del experimento. Dado que son modeladas en el contacto con la experiencia, algunas intuiciones relativas a los conceptos comunes aplicados en situaciones familiares son relativamente estables mientras que divergen o fallan cuando los conceptos se aplican en un entorno nuevo o

²² Hay muchas razones para abandonar una hipótesis, pero ninguna es decisiva. Como ya señaló Duhem ([1914] (1982)), siempre se puede mantener una hipótesis dada si se está dispuesto a modificar o sacrificar algunas otras, incluso introduciendo hipótesis *ad hoc* implausibles. El hecho de que una hipótesis esté más alta en la jerarquía, implica mayores costos, pero nada nos fuerza a abandonar las hipótesis de más bajo nivel, a menos que se asuma una norma de conservadorismo epistemológico, como la máxima de la mutilación mínima de Quine (1990).

desconocido. En los experimentos mentales las intuiciones son parte del conocimiento de fondo y determinan el resultado del experimento. Para Brendel, las intuiciones no gozan de ningún privilegio respecto de otras creencias relativas a la percepción, la introspección y la memoria, ni constituyen una vía privilegiada de acceso al mundo natural sino que, al igual que otras creencias son falibles.

Los escenarios imaginarios empleados en los experimentos mentales, según Brendel, apelan a intuiciones y esto es lo que constituye su carácter experimental. En síntesis, sin desafiar una de las tesis básicas de la posición argumental (la de la reconstrucción) Brendel ofrece una explicación, dentro de los límites del empirismo, del origen y del alcance del conocimiento en estos experimentos. En este marco, ofrece un criterio de confiabilidad. Las intuiciones que son la fuente de conocimiento en los experimentos mentales son falibles y pueden extraviarse si el escenario descrito en el experimento no satisface las condiciones antes mencionadas. La reconstrucción argumental permite explicitar las intuiciones que funcionan como conocimiento de fondo y evaluar si la conclusión está racionalmente justificada. Cuando las intuiciones son mal conducidas por el escenario imaginario, el resultado puede aparecer al experimentador como racionalmente justificado, cuando en realidad se basa en un razonamiento que falla en representar la relación entre variables, invoca hipótesis imposibles dados los supuestos teóricos, apela a situaciones acerca de las cuales los experimentadores no han desarrollado intuiciones o en los que los detalles desorientan al experimentador de tal manera que no puede emplear sus intuiciones confiablemente.

Algunos de los problemas que esta manera de evaluar a los experimentos mentales exitosos suscita devienen de la dificultad de evaluar la confiabilidad de las intuiciones. Dado que se supone que las intuiciones se forman en nuestro contacto con la experiencia, puede suponerse que esta relación esta mediada social y culturalmente y que como resultado de esta mediación lo que resulta conspicuo para un grupo puede no serlo para otro o puede aparecer como una

percepción sesgada. De la misma manera, la formación científica puede ser determinante del carácter prístino de algunas intuiciones, es decir, sin este tipo de formación algunos experimentadores pueden no contar en absoluto con las intuiciones relevantes.

Clatterbuck proporciona una interpretación alternativa de la tesis de la reconstrucción que incorpora de manera esencial la representación del escenario imaginario. Desde esta perspectiva, los experimentos mentales son una forma particular de inferencia inductiva que comparte algunos aspectos con las simulaciones computacionales. Las operaciones de idealización consistentes en abstraer propiedades contingentes de la descripción del escenario facilita la inferencia de conclusiones generales. El procedimiento que sigue esta forma de razonamiento puede ser esquematizado del siguiente modo:

1. El sujeto simula en su imaginación un fenómeno en un contexto *C* con el que está suficientemente familiarizado.
2. Abstrae de la descripción del escenario contingentes del contexto con el que está familiarizado.
3. Realiza la simulación de un fenómeno en un contexto idealizado con el cual no tiene experiencia de familiaridad directa.
4. Realiza una inducción que extiende las conclusiones del caso a otros contextos (Clatterbuck, 2013: 322).

Esta perspectiva comparte algunos rasgos con la posición de Brown que ubicamos dentro de las concepciones experimentales. De acuerdo con esta propuesta, la ejecución del experimento involucra una perspectiva de primera persona y la realización de una acción o la activación de un mecanismo para “ver qué sucede” (Brown, 2003: 25). Estas acciones permiten al experimentador representarse un fenómeno que exhibe de manera conspicua ciertas propiedades de clase natural. Esta clase de fenómenos *a priori* posibilita la captación intuitiva de leyes de la naturaleza. La propuesta de Clatterbuck es similar aunque no

suscribe los mismos compromisos metafísicos. Los experimentos mentales dependen de la experiencia subjetiva del experimentador, de su grado de familiaridad con el escenario, de la posibilidad de manipular los objetos de la manera descrita, y posteriormente, poder eliminar las características contingentes del caso. La inferencia que justifica racionalmente las conclusiones de los experimentos mentales es un tipo de inferencia inductiva no enumerativa. En las inferencias inductivas probabilísticas, la relación entre casos observados y no observados está dada por las características de la muestra, que habilita a proyectar una regularidad observada sobre los casos no observados. En el tipo de inferencia que Clatterbuck caracteriza como “inducciones de Dewey” la inferencia de casos observados a no observados depende de la identificación de individuos que exhiban características representativas de su clase. Experimentos como el de Galileo son instancias ejemplares de este tipo de inducción no enumerativa. La efectividad de los mismos depende de que sea posible abstraer las propiedades esenciales de los detalles contingentes de la experiencia particular. La principal dificultad reside en encontrar individuos que sean representativos de su tipo, cuyas propiedades sean proyectables a casos desconocidos.

Aunque esta perspectiva incluye de una manera esencial la representación del escenario imaginario para la obtención de nuevo conocimiento en los experimentos mentales, lo que permitiría subsanar una de las principales limitaciones de la perspectiva argumental, el tipo de inferencia invocada compromete a esta posición al menos con una concepción realista respecto de la existencia de clases naturales. El tipo de operación invocada es dificultosa en tanto sería necesario especificar los rasgos que estos casos ejemplares exhibidos en los experimentos mentales tienen y que haría posible identificarlos. Asimismo, la explicación parece no capturar todas las funciones que estos experimentos pueden desempeñar. La representación (que involucra actividades de abstracción e idealización) de un caso que exhibe propiedades típicas no es el tipo de mecanismo invocado por todos los experimentos mentales científicos exitosos.

Esta podría ser una explicación razonable del funcionamiento de los experimentos mentales constructivos pero otros casos, por ejemplo, aquellos que tienen por objetivo refutar una hipótesis, apelan a la presentación de un contraejemplo y no dependen de que representen propiedades naturales. Por otra parte, otros tipos de inferencia ampliativa podrían explicar la obtención de nuevo conocimiento en los experimentos mentales y en este sentido modelar o incorporar el mecanismo de conocimiento en cuestión a otras formas de conocimiento. El tipo de inferencia abductiva o inferencia a la mejor explicación, podría proporcionar una hipótesis explicativa de los fenómenos representados en el escenario imaginario.²³ El problema es que estas formas de inferencia fueron relegadas, al menos según las opiniones de varios teóricos relevantes, al contexto de descubrimiento, como recursos importantes para la generación de hipótesis pero carentes de valor justificatorio.

En síntesis, las perspectivas revisadas en este apartado intentan explicar el papel que la representación de un escenario imaginario tiene en los experimentos mentales asumiendo que la eficacia heurística y la justificación de los resultados no pueden ser suficientemente explicadas por la perspectiva argumental. Aunque no suscriben la tesis de acuerdo con la cual los experimentos mentales pueden ser reducidos a argumentos, sostienen que la identificación de un esquema de inferencia permite evaluar los casos y distinguir los experimentos exitosos de los espurios. Para algunos de los representantes de este punto de vista, la narrativa experimental es similar a la narrativa de un experimento real o de un experimento idealizado y la visualización o representación de un escenario desempeña un papel al facilitar la aceptación de algunas premisas necesarias para la inferencia de ciertas conclusiones. En la perspectiva “pragmática” de Goffie y Roux, los experimentos mentales introducen una proposición que genera un conflicto entre creencias. Esta es una explicación plausible de cómo la representación de una

²³ Una interpretación posible de los poderes confirmatorios de los experimentos mentales en ciencias fácticas, es atribuir a la IME el poder explicativo del mecanismo de conocimiento (véase Borge y Mettini, 2018).

situación imaginaria favorece la aceptación de una creencia que entra en conflicto con un sistema de creencias preexistente. Pero la naturaleza psicologista de esta explicación permite suponer que este efecto depende de varios factores, además de la estructura de las creencias de los experimentadores, que la explicación no alcanza a determinar. Entre estos factores podrían señalarse: la variabilidad en la vivacidad de las intuiciones a las que apela la representación visual, la incidencia de las hipótesis auxiliares en la reconfiguración de los sistemas de creencias, las diferencias en la formulación proposicional de la representación visual. Como tal, la perspectiva podría explicar la función heurística de los escenarios imaginarios, pero no su papel en la justificación de conclusiones.

De manera similar, la perspectiva de Brendel hace de la intuición sensible la fuente del conocimiento involucrado en los experimentos mentales. El criterio de confiabilidad de los experimentos mentales está determinado en parte por la familiaridad de los interlocutores con el escenario descrito. El principal problema para establecer usos legítimos e ilegítimos desde esta posición es determinar si esto efectivamente ocurre, lo que implicaría formular exhaustivamente los supuestos de fondo del experimento.

La posición de Clatterbuck otorga a la realización de un experimento en la imaginación un papel más importante. Este es el que permite extender las conclusiones a otros casos. No obstante, esta posición se compromete con una forma de realismo respecto de las clases naturales. Este compromiso puede considerarse como un elemento injustificado y, aunque no fuera así, la explicación no permite dar sentido a los casos que dependen de la presentación de un contraejemplo.

3.5 Razonamiento contrafáctico

Las perspectivas inferencialistas consideran que las operaciones de representación cumplen un papel central en la justificación de las conclusiones obtenidas a partir de la ejecución de un experimento mental. Estas posiciones desafían la tesis de la eliminación y sostienen que la actividad imaginativa desempeña un papel en la derivación de conclusiones. Otros autores han argumentado contra las tesis ontológicas de Norton (identidad y dispensabilidad) y han sostenido que los experimentos mentales son una subclase de experimentos asociados a argumentos. Desde este punto de vista, la ejecución de un experimento mental tiene lugar en la imaginación de los experimentadores y motiva la construcción de un argumento que apoya o refuta una hipótesis teórica. La reconstrucción del argumento en cuestión implica, para este punto de vista, la introducción de operadores modales. Esta regimentación permite dar sentido al tipo de supuesto contrafáctico que da lugar a la reconfiguración de un esquema de creencias. Asimismo, estas regimentaciones dan sentido a varios casos canónicos de experimentos mentales en física que emplean premisas que son explícitamente falsas porque involucran idealizaciones (abstracciones, supresión de variables) o contrafácticos (negación de enunciados legales). Se revisan a continuación estas propuestas de regimentación para evaluar su utilidad para la elaboración de criterios de confiabilidad para los experimentos mentales.

Sorensen se compromete con una perspectiva machiana respecto del origen del conocimiento proporcionado por los experimentos mentales, según la cual los estos son descendientes de los experimentos ordinarios que apelan a la intuición sensible (Sorensen, 1992: 64). Su funcionamiento se explica atendiendo a dos factores. Por una parte, la representación de un escenario imaginario genera creencias de manera similar a como los experimentos ordinarios lo hacen. Al igual que los experimentos ejecutados, los experimentos mentales emplean un diseño experimental y están asociados con argumentos que proporcionan una

interpretación del experimento en un contexto teórico. Por otra parte, los experimentos mentales desempeñan un papel fundamental en la evaluación de teorías. Esta tesis abreva en la concepción khuniana de acuerdo con la cual, estos experimentos exhiben anomalías dentro de una matriz teórica. Para Sorensen, la de Kuhn es una buena aproximación al funcionamiento de los experimentos mentales. No obstante, considera necesario matizar la idea de que hacen posible derivar contradicciones. Lo que los experimentos mentales son capaces de mostrar es que existe algún tipo de inconsistencia local que es preciso resolver abandonando alguna de las premisas a partir de las cuales se genera. El tipo de inconsistencia que son capaces de exhibir puede capturarse en términos lógicos por la estructura de una paradoja (Sorensen, 1992: 123), por lo que su funcionamiento puede comprenderse adecuadamente si se los considera *refutadores aléticos*. Como tales, su objetivo es derivar una inconsistencia de un principio teórico que Sorensen denomina “enunciado fuente”. Esta clase de supuestos se caracteriza por implicar algún tipo de consecuencia modal.²⁴ En el caso de las ciencias naturales, los enunciados fuente de los experimentos mentales son enunciados legales o nomológicos. En general, lo que el argumento muestra es que en una situación contrafáctica, una consecuencia modal del enunciado fuente produce un resultado extraño que es descartado como imposible. Existen dos clases de *refutadores aléticos*. Un *refutador de necesidad* es un supuesto diseñado para refutar una afirmación mostrando que algo, descartado como imposible por esa afirmación, es realmente posible después de todo (Sorensen, 1992: 135). Este razonamiento, asociado con un escenario contrafáctico, se formula del siguiente modo:

1. S
2. $S \supset \square I$

²⁴ Sorensen emplea “modalidad” en un sentido amplio. En el caso de los experimentos mentales en física, la modalidad es física, en el caso de los experimentos mentales en ética la modalidad es moral, en el caso de los experimentos mentales en teoría del conocimiento la modalidad es epistémica, etc.

3. $(I \ \& \ C) \Box \rightarrow W$
4. $\neg \Diamond W$
5. $\Diamond C \bullet$

En este esquema, (1) representa al enunciado fuente S , (2) al “extractor de necesidad”, es decir, a la premisa en la cual se deriva del enunciado fuente, una consecuencia necesaria I ; (3) indica que la conjunción entre esa consecuencia necesaria y cierto escenario contrafáctico tiene la consecuencia extraña W ; (4) descarta W como imposible y (5) establece que el escenario contrafáctico C es posible. Los experimentos mentales exponen la existencia de esta forma particular de inconsistencia introduciendo el escenario C . La manera más simple de evitar la paradoja consiste en eliminar el enunciado fuente, en otras palabras, la ejecución del experimento motiva el abandono del enunciado fuente. Los ejemplos que Sorensen reconstruye con este esquema corresponden a experimentos mentales en filosofía. El caso que mejor se acomoda a su interpretación son los contraejemplos de Gettier como refutadores de la definición de conocimiento como creencia verdadera justificada (Sorensen, 1992: 137). El experimento de Galileo podría reconstruirse, al menos en su parte negativa (Brown, 2011: 46), como destinado a mostrar la contradicción a la que conduce la noción aristotélica de velocidad natural:

- (1) La velocidad natural es una función del peso.
- (2) Si la velocidad natural es una función del peso, entonces, en un mismo medio material dos cuerpos de distinto peso en caída libre, *ceteris paribus*, caen con velocidades proporcionales a su peso.
- (3) En un mismo medio dos cuerpos de distinto peso en caída libre, *ceteris paribus*, caen con velocidades proporcionales a su peso y ambos cuerpos pueden ser unidos para formar un compuesto, entonces, necesariamente el cuerpo más pesado acelerará la velocidad de la caída del más liviano y el cuerpo más liviano retardará la velocidad de caída del más pesado.

- (4) Es imposible que el compuesto caiga simultáneamente más rápida y más lentamente que el cuerpo más pesado.
- (5) Es posible arrojar dos cuerpos de distinto peso unidos.

De acuerdo con esta reconstrucción, el argumento motiva el abandono de (1) debido a que la interpretación del experimento compromete a los experimentadores con la aserción de un conjunto de premisas contradictorias. Pero este análisis no explica cómo el experimento apoya la interpretación galileana del fenómeno.

Los que Sorensen denomina *refutadores de posibilidad* tienen un objetivo cognitivo diferente del de los *refutadores de necesidad*. En este caso, se deriva una consecuencia posible del enunciado fuente (2) y se afirma que es posible la conjunción de la consecuencia modal y el contenido del experimento (5).

- 1. S
- 2. $S \supset \Diamond I$
- 3. $(I \ \& \ C) \Box \rightarrow W$
- 4. $\neg \Diamond W$
- 5. $\Diamond I \supset \Diamond(I \ \& \ C) \bullet$

El propósito de esta clase de razonamientos es mostrar que el enunciado fuente tiene implicaciones modales espurias. En el caso de este esquema, como en el primero, la manera de resolver la inconsistencia consiste en rechazar el enunciado fuente. Sorensen no ofrece una reconstrucción de casos de experimentos mentales en ciencia. Sin embargo, comenta algunos como ejemplos que podrían reformularse a partir de este esquema. El experimento del gato de Schrödinger funciona, de acuerdo con Sorensen (1992: 156), “amplificando” las consecuencias absurdas del enunciado fuente, en este caso del principio de superposición de estados. Otro caso que emplea la misma estrategia es el

experimento de Lucrecio para rechazar la tesis de la finitud del universo, que podría reconstruirse siguiendo el esquema siguiente:

1. El universo es finito.
2. Si el universo es finito, entonces, es posible que un hombre viaje hasta el límite del universo.
3. Si un hombre arrojara una lanza hacia el límite del universo, entonces, necesariamente esta rebotaría o se desvanecería.
4. No es posible que una lanza arrojada hacia el límite del universo rebote y se desvanezca.
5. Es posible viajar al límite del universo y arrojar una lanza.

Aunque la propuesta de Sorensen da sentido a la intuición de acuerdo con la cual los experimentos mentales poseen un componente experimental o, al menos, explica la relación entre los supuestos contrafácticos y el argumento que acompaña el experimento, la reconstrucción, a partir de los esquemas propuestos, de los casos de experimentos mentales en ciencia resulta problemática. En ambos modelos, el segundo enunciado deriva del primero una consecuencia (metafísicamente) necesaria. Ya que este enunciado no puede ser falso, la principal crítica a esta regimentación (Häggqvist, 1996: 97) señala que el enunciado identificado como (2) en los esquemas, o bien es tautológico y en tal caso no cumple ningún papel en la reconstrucción, o bien es un enunciado lógicamente más fuerte que (1) y, por consiguiente, afirma algo que no está contenido en el primer enunciado y es el objetivo de la crítica que se realiza a partir del experimento (lo cual haría a (1) innecesario en la formulación). Por otra parte, los experimentos mentales que intentan dar apoyo a una hipótesis o que desempeñan funciones exploratorias (en casos en los cuales no existe una teorización previa) no pueden acomodarse al modelo propuesto de acuerdo con el cual la principal función de los experimentos mentales es refutar teorías. Entre estos casos, que Brown denomina experimentos mentales constructivos, podría mencionarse el experimento del balde giratorio de Newton, en el que el punto de

partida no es un supuesto teórico sino un fenómeno hipotético para el cual se conjetura una explicación teórica.

Häggqvist (1996), por su parte, argumenta que los experimentos mentales son casos hipotéticos que se proponen cumplir funciones similares a las de los experimentos ordinarios. Sostiene que la perspectiva argumental ignora una distinción fundamental entre entidades lingüísticas, como enunciados e inferencias, y entidades no lingüísticas, como eventos y procesos. A diferencia de los argumentos, los experimentos mentales no son compuestos de entidades con valor de verdad y no puede afirmarse significativamente que sean válidos o inválidos. Son más bien *eventos o procesos* que motivan enunciados. Los eventos físicos que tienen lugar en el laboratorio causan que los observadores profieran ciertos enunciados observacionales. Los eventos psicológicos que tienen lugar dentro de la mente del experimentador que realiza un experimento mental, motivan la aceptación de ciertos enunciados (Häggqvist, 1996: 86). Estos enunciados son subsecuentemente empleados como premisas de los argumentos cuyo objetivo es evaluar teorías e hipótesis.

En tanto procedimientos destinados a generar la aceptación o el rechazo de ciertos enunciados o proveer fundamentos para la creencia en algún principio, los experimentos mentales adoptan algunas formas canónicas. La forma típica de los argumentos asociados con esta clase de experimentos es la de argumentos refutatorios, donde el experimento plantea una situación contrafáctica en la que una tesis bajo escrutinio resultaría falsa (Häggqvist, 2009: 97). Existen tres modelos que pueden adoptar los argumentos asociados con los experimentos mentales. Estos esquemas de reconstrucción representan las diferentes maneras en las cuales es posible resolver la inconsistencia entre un condicional anidado que establece que, a partir de cierto principio teórico y ciertas condiciones iniciales, debería seguirse determinada consecuencia, y la representación de un escenario posible en el que la consecuencia no se da. Estos esquemas explican por qué la representación del mismo escenario (es decir, la realización del mismo

experimento mental) da lugar a diferentes argumentos. Asimismo proporcionan un criterio para identificar casos espurios.²⁵

El primer modelo para la reconstrucción del argumento ligado a un experimento mental establece que la presentación de un escenario imaginario posible dentro de un determinado marco teórico implica una contradicción que sugiere el rechazo de la teoría en cuestión. El objetivo del argumento es evaluar una teoría T , en un escenario posible C . En este contexto, una determinada consecuencia W debería darse necesariamente de acuerdo con la teoría T . No obstante, de las condiciones contrafácticas C se sigue necesariamente que la consecuencia W no se da y, por lo tanto que T es falsa. La formalización de este tipo de argumentos provista por Häggqvist es la siguiente:

1. $\diamond C$
2. $T \supset (C \square \rightarrow W)$
3. $C \square \rightarrow \neg W$
4. $\neg T \bullet$

Para Häggqvist, los modelos de reconstrucción propuestos exhiben lo que denomina la *estructura dialéctica* de los experimentos mentales, en la que un escenario posible es presentado a los experimentadores, la tesis que es objetivo del experimento es puesta en conexión con cierta consecuencia y de esto se sigue que en el escenario descrito tal consecuencia no se obtiene. Este tipo de modelo pone énfasis en las instancias que son contraejemplos de las tesis bajo consideración y supone que el caso que presenta el experimento genera la creencia en la plausibilidad del escenario contrafáctico que motiva la aceptación de la premisa 3 y de la conclusión dada la asunción de 2. Esta situación es similar a la de los experimentos ordinarios en los que la creencia en una teoría dada

²⁵ A diferencia de la propuesta de Sorensen, en la propuesta de Häggqvist los experimentos mentales funcionan como refutadores y las variantes de este esquema de inferencia corresponden a las maneras de resolver la contradicción.

implica ciertos condicionales (por ejemplo, que si se producen las condiciones iniciales C , ocurrirá el evento E) y en este sentido es guía la elección del experimento en cuestión (Häggqvist, 2009: 65).

El primer tipo de argumento presenta una inconsistencia en el conjunto $\{T, \diamond C, T \rightarrow (C \Box \rightarrow \neg W), C \rightarrow \Box \neg W\}$. Una manera de resolver esta inconsistencia es rechazando T . Los otros esquemas de argumento lidian con esta inconsistencia de formas diferentes, fundamentalmente defienden el principio teórico que el experimento ataca. El segundo modelo rechaza la idea de que en el escenario contrafáctico dado por el experimento mental se dé la consecuencia prevista por la hipótesis.

1. T
2. $\diamond C$
3. $T \supset (C \Box \rightarrow W)$
4. $\diamond C \ \& \ C \Box \rightarrow W \bullet$

La forma de resolver la inconsistencia consiste en este caso en “morder la bala” y negar que en el escenario (C) se daría la consecuencia W , aduciendo, por ejemplo, que la sensación de que la consecuencia es implausible en el escenario dado se debe a la extrañeza del escenario mismo.

Otro esquema de argumento consiste en sostener que la tesis en cuestión no está comprometida con la consecuencia W . En este caso, se afirma que la consecuencia en cuestión no es relevante como instancia refutatoria de la hipótesis. El esquema de esta forma de argumento es:

1. T
2. $\diamond C$
3. $C \Box \rightarrow \neg W$
4. $\neg(T \supset (C \Box \rightarrow W)) \bullet$

Un último esquema señala que el escenario en cuestión es imposible. Este es el caso típico de quienes mantienen una posición escéptica respecto de la ejecución de experimentos mentales en el contexto de justificación de hipótesis:

1. T
2. $T \supset (C \square \rightarrow W)$
3. $C \square \rightarrow \neg W$
4. $\neg \diamond C$

El criterio para identificar casos de experimentos que fallan depende, fundamentalmente, del grado en que las premisas están justificadas. Los experimentos mentales fallidos pueden ser, por ejemplo, del primer tipo de reconstrucción en el que no todas las premisas están justificadas y recíprocamente los experimentos mentales que adoptan los modelos representados por el segundo, tercer y cuarto esquema pueden resultar como casos especiales de esta condición (en tanto atacan la justificación de alguna de las premisas del primer esquema).

Hägqvist señala que, aunque el mecanismo que los experimentos mentales emplean para evaluar principios teóricos es similar al que emplean los experimentos ordinarios, los primeros son estrategias menos confiables. Una de sus principales debilidades se vincula con el origen de las premisas en cuestión. Las premisas especifican algún tipo de creencia. Tales creencias pueden proceder (ser inferidas) de principios teóricos, o guardar alguna relación con la percepción (pueden proceder de la memoria, ser heredadas evolutivamente de nuestros ancestros, provenir del sentido común, etc.). El origen de estas creencias, en principio, es menos confiable que en el caso de los experimentos ordinarios, donde las creencias se forman en base a la experiencia (al menos las premisas vinculadas con las descripción de las condiciones iniciales y la predicción de la creencia en cuestión). Asimismo, en los experimentos mentales, como en los reales, la fuerza persuasiva del caso depende de la capacidad de los

experimentadores de “evaluar por sí mismos” la situación en cuestión. La “obviedad” de las conclusiones es en los experimentos mentales más polémica que en el caso de los experimentos reales. Por esta misma razón, las condiciones iniciales que indican que en el escenario imaginario la conclusión implicada por la teoría no se sigue, es más vulnerable que la descripción de las condiciones iniciales de un experimento real.

Otro aspecto que socava la confiabilidad de los experimentos mentales en lo que concierne a sus funciones confirmatorias se relaciona con el vínculo entre el experimento y la teoría general o hipótesis que es objeto de examen. En los experimentos ordinarios, la relación entre la teoría, las condiciones iniciales y las consecuencias observacionales esperables es, según Häggqvist, similar a la de una inferencia deductiva apoyada por hipótesis auxiliares. En el caso de los experimentos mentales, la relación entre el contrafáctico y la teoría no es tan clara. Una interpretación posible de este vínculo es suponer que el condicional contrafáctico cae dentro del alcance modal de la teoría. La principal dificultad en este punto consiste en establecer la relevancia del escenario que introduce el condicional contrafáctico en cuestión para contrastar la teoría bajo examen.

Los modelos de Häggqvist son utilizados para reconstruir, casi exclusivamente, casos de experimentos mentales en filosofía. En estos casos, el mecanismo de suposición empleado se ajusta a los esquemas descritos porque la actividad argumental en esta disciplina es dialéctica, es decir, en general se intenta socavar la plausibilidad de algún principio para establecer una explicación alternativa. Los casos en ciencia no son tan dóciles a esta regimentación. Häggqvist (1996: 103-104) presenta como ejemplos del primer modelo el experimento de los cerebros en cubetas de Putnam (1982) y el experimento de la Tierra gemela (Putnam, 1975). En ambos casos el experimento motiva la negación del enunciado legal (o definicional) que constituye el antecedente del condicional expresado en (2). El único experimento mental científico que reconstruye siguiendo este esquema es el experimento de la caja de fotones de Einstein.

Según Haggqvist, el experimento tal como fue formulado por Einstein motiva el abandono del principio de indeterminación. La respuesta de Bohr puede ser representada por el último esquema, de acuerdo con el cual el escenario en cuestión es imposible, ya que rechaza la posibilidad postulada por el argumento de Einstein (que pueda medirse con precisión arbitraria la posición y el momento de un fotón).

Si bien la perspectiva de Häggqvist se apoya en las similitudes entre los argumentos empleados para interpretar experimentos mentales y experimentos reales, la formación de creencias y la formulación de hipótesis parecen ser diferentes en uno y otro caso. La formulación del argumento que acompaña a un experimento mental parece depender fuertemente de las consecuencias metafísicas y epistémicas que pueden seguirse de un principio teórico. En este sentido, el mecanismo es similar a la indagación especulativa de esta clase de principios. Por otra parte, las razones que Häggqvist proporciona para sostener el argumento de la diferencia ontológica entre enunciados y procesos no resultan convincentes. Simplemente, afirma que se trata de tipos diferentes de entidades sin llegar a mostrar en qué reside la diferencia específica entre la descripción de un procedimiento que se ejecuta en el laboratorio y un procedimiento puramente hipotético. Por último, su perspectiva tampoco parece hacer justicia a los casos en los cuales el escenario contrafáctico no entra en conflicto con un supuesto teórico, sino que tiene el objetivo de ilustrar un supuesto teórico.

En síntesis, las posiciones de Sorensen y Häggqvist defienden el carácter experimental de los experimentos mentales y emplean operadores modales para proporcionar una interpretación de los argumentos asociados con estos dispositivos. Estas perspectivas proporcionan un modelo plausible para reconstruir experimentos mentales en filosofía, en los que usualmente los supuestos contrafácticos se emplean para “presentar un caso” contra un principio teórico y a favor de otra interpretación. No obstante, esta posición no captura adecuadamente la estructura de todos los experimentos mentales científicos. No

todos los casos en ciencias fácticas tienen a la estructura de una paradoja. Quizás este es la principal razón por la cual resulta tan difícil reconstruir experimentos mentales científicos siguiendo estos esquemas de razonamiento contrafáctico. Por otra parte, la relación entre las premisas del argumento y la realización del experimento imaginario se propone como una explicación causal: la representación de un escenario motiva ciertas creencias o el asentimiento a ciertos enunciados. Los criterios para evaluar estos enunciados, sin embargo, son difíciles de formular. En el caso de los experimentos ordinarios, existe una serie de criterios epistémicos para calibrar el dispositivo experimental y para garantizar la objetividad de las observaciones. En estas perspectivas se arguye un origen empírico del conocimiento en cuestión, pero dado que en los experimentos mentales los escenarios en ocasiones involucran situaciones ideales o incluso físicamente imposibles, es difícil aplicar criterios análogos para determinar el valor de verdad de los enunciados que motivan. Asimismo, la relación entre el diseño experimental y la teoría en cuestión es, como el propio Häggqvist asume, es más discutible que en los experimentos ordinarios.

3.6 Conclusiones

A pesar de sus ventajas explicativas, la perspectiva argumental tiene importantes limitaciones. En primer lugar, no proporciona una explicación satisfactoria de los rasgos que distinguen a los experimentos mentales de otras formas de razonamiento hipotético. Fundamentalmente, las posiciones que insisten en la dispensabilidad del escenario experimental y sostienen que la ejecución de un experimento mental coincide con la de un argumento, no están respaldadas por la intuición general de acuerdo con la cual la realización de un experimento mental consiste en seguir un procedimiento para arribar a un resultado. Por otra parte, en muchos casos, la identificación de las premisas relevantes para la inferencia es dificultosa e involucra la clarificación de un

conjunto de supuestos de fondo, que no son necesarios en principio para derivar la conclusión en cuestión cuando el experimento preserva su narrativa experimental. La perspectiva argumental subestima la eficacia heurística de los experimentos mentales. Dado que no considera que los rasgos experimentales sean un elemento relevante de los poderes epistémicos de estos experimentos, no proporciona una explicación de porqué los resultados de un argumento son obtenidos de manera más fácil y rápida por un experimento mental.

Este punto de vista tampoco proporciona una interpretación convincente de la significación evidencial de los experimentos mentales. Algunos casos históricamente relevantes, como el de Galileo o el experimento de la caja de fotones de Einstein y la reinterpretación del mismo por parte de Bohr, han dado apoyo a principios teóricos de un modo más cercano a la evidencia empírica que al argumento teórico. Por otra parte, aunque el mecanismo en virtud del cual los experimentos mentales proporcionan nuevo conocimiento no es completamente análogo al de los experimentos ordinarios, la posibilidad de ser replicados bajo principios teóricos diferentes a aquellos en el contexto de los cuales fueron inicialmente formulados y la confirmación de los resultados de experimentos mentales realizados como experimentos reales, son características que los asemejan a los experimentos ordinarios. Asimismo, la posibilidad de separar la descripción del escenario, las condiciones iniciales, las operaciones y el escenario en cuestión de su interpretación teórica son razones para considerar que poseen un aspecto experimental que es epistémicamente importante o que al menos no es despreciable.

Las dificultades que enfrentan las perspectivas argumental e inferencial para dar cuenta de la manera en que la información empírica es reformulada a través de los argumentos procede, respectivamente, del reduccionismo y de la explicación insuficiente del papel de los mecanismos de visualización en estas posiciones. En los esquemas de reformulación no hay explicaciones del papel que

las abstracciones e idealizaciones desempeñan en la obtención de las conclusiones.

Para Norton, los detalles de la situación son irrelevantes para la conclusión. Pero en el experimento mental de Galileo, la abstracción del rozamiento y la falta de peso de la cuerda que une los dos cuerpos podrían considerarse detalles como el color o el tipo de objeto en cuestión. No obstante, estas características son relevantes para la obtención de una conclusión general. El supuesto de que solo los argumentos pueden proporcionar un esquema confiable para explicar el nuevo conocimiento es indispensable para estas perspectivas. Apelar a otros mecanismos de formación de creencias puede dar lugar a una explicación más completa de cómo la representación de un escenario imaginario influye en la obtención de ciertas conclusiones. El tipo de explicaciones en las cuales el escenario descrito desempeña un papel relevante en la formación de creencias apela a un modo de explicación psicologista que no logra dar cuenta del papel de la representación en la justificación de las conclusiones del argumento. No obstante el trabajo de reconstrucción que realizan las perspectivas argumentales e inferenciales devela una serie de supuestos relevantes para examinar la función que los experimentos mentales pueden desempeñar en la elección de teorías. La identificación de premisas relevantes, así como de un esquema de inferencia, permite dar sentido a los casos como el de la caja de fotones donde hay dos interpretaciones del funcionamiento del mismo experimento.

SEGUNDA PARTE

Los fundamentos del conocimiento en los
experimentos mentales

CAPÍTULO 4

El origen del conocimiento en los experimentos mentales

4.1 Introducción

Los intentos de proporcionar una explicación acerca del funcionamiento de los experimentos mentales asumen como principal desafío dar respuesta al interrogante que fuera formulado por Kuhn ([1962] 2004: 232): “¿Cómo es (...) que, confiando exclusivamente en datos familiares, se puede llegar con un experimento imaginario a una nueva comprensión de la naturaleza?”. El examen de casos tomados de la historia de la ciencia muestra que, al menos en algunas ocasiones, los experimentos mentales son capaces de proporcionar conocimiento nuevo sobre el mundo natural. El análisis filosófico de esta clase de experimentos ha intentado determinar la procedencia del nuevo conocimiento (explicando asimismo en qué sentido tal conocimiento es nuevo) y el mecanismo en virtud del cual es obtenido. Las posiciones existentes pueden ser reunidas en tres grupos, considerando el tipo de compromisos epistémicos y metafísicos que asumen para proporcionar tal explicación. Las agrupaciones coinciden con perspectivas filosóficas tradicionales sobre el fundamento del conocimiento.

1. Empiristas
2. Aprioristas
 - a) Racionalistas
 - b) Trascendentalistas

3. Constructivistas²⁶

Las posiciones en el primer grupo sostienen que el único fundamento incontrovertible de conocimiento del mundo natural es la experiencia sensible y que, por lo tanto, la única explicación plausible de cómo obtenemos nuevo conocimiento a partir de la ejecución de un experimento mental es que éste es el producto de la reorganización de creencias empíricas. Desde esta perspectiva, un experimento mental proporciona conocimiento nuevo porque a partir de su ejecución la información empírica previamente acumulada es reorganizada de manera que pueden extraerse de ellas conclusiones nuevas (en un sentido epistémico, no necesariamente lógico).

Las segundas, que agrupamos bajo el nombre de *aprioristas* pueden subdividirse en dos variantes de acuerdo con algunas diferencias en los compromisos epistémicos y metafísicos que asumen. La línea *racionalista* representada por Brown (2011) se compromete con una forma de racionalismo en relación con el mecanismo de adquisición de conocimiento y con un realismo de corte platónico en lo que respecta al objeto de conocimiento. Propone que los experimentos mentales permiten a los experimentadores trascender el conocimiento empírico y captar intuitivamente entidades abstractas. En el caso de los experimentos mentales en ciencias fácticas, estas entidades son las leyes de la naturaleza. En algunos casos es posible captar directamente las leyes de la naturaleza de la misma manera que es posible captar verdades matemáticas. Para este punto de vista, algunos experimentos mentales proporcionan nuevo conocimiento en un sentido radical, en tanto que no puede derivarse de otros conocimientos presentes en el conjunto de creencias previas del experimentador.

²⁶ Grundman (2018: 342) ofrece una clasificación similar en la que denomina empiristas a las primeras y racionalistas a las segundas, entre las que distingue perspectivas trascendentalistas (Buzzoni, 2013), conceptualistas (Ludwig, 2007) y platonistas (Brown, 2011).

La segunda variante dentro de los enfoques *aprioristas*, que denominaremos *trascendentalistas*, se compromete con la noción de conocimiento *a priori* funcional, pero no con una concepción platonista de las leyes de la naturaleza. Según esta perspectiva, la capacidad de plantear preguntas a la naturaleza constituye una condición de posibilidad de la experimentación. La capacidad de producir hipótesis que anticipan el comportamiento de los fenómenos atestigua la existencia de una forma de conocimiento *a priori* funcional, simultáneamente producto de la facultad especulativa y dependiente del conocimiento empírico adquirido (no particularmente de la familiaridad del experimentador con el mundo sensible, sino de su conocimiento de hechos científicos bien establecidos y leyes aceptadas como verdaderas). Debido a que esta facultad es falible, las proposiciones que se obtienen como resultado de la ejecución de un experimento mental deben ser comparadas con los resultados de un experimento real análogo. En síntesis, para este punto de vista los experimentos mentales proporcionan nuevo conocimiento empleando la capacidad de suponer o conjeturar (que es una facultad *a priori* funcional) porque anticipan el comportamiento de la naturaleza sin introducir nuevo conocimiento empírico.

Los enfoques que denominamos *constructivistas* sostienen que los experimentos mentales pueden revelar los límites de nuestro conocimiento del mundo. Para ellos, la novedad del conocimiento que esta clase de experimentos puede proporcionar se debe a la corrección o ampliación del esquema conceptual del sujeto cognoscente. De acuerdo con esta perspectiva, los experimentadores aprenden acerca de su sistema conceptual y del mundo simultáneamente. La efectividad de algunos experimentos mentales para ampliar el conocimiento se basa en la mutua dependencia entre conceptos y expectativas empíricas. Los experimentos mentales exitosos pueden proporcionar evidencia a favor o en contra de un principio teórico porque en ellos tiene lugar la presentación de un fenómeno que ilustra un principio general (y en este sentido encaja en el esquema

conceptual del experimentador) o bien constituye un contraejemplo (porque no puede ser explicado en el marco conceptual del experimentador).

La cuestión de determinar el origen y el mecanismo de adquisición de nuevo conocimiento en los experimentos mentales constituye uno de los nodos de su análisis filosófico. Al igual que la disputa tradicional entre empiristas y racionalistas en teoría del conocimiento, se trata de una cuestión probablemente irresoluble. Aunque no es posible determinar en principio qué constituiría una respuesta totalmente satisfactoria a este problema, es posible adelantar criterios convenientes para evaluar las propuestas existentes. Asumimos que una interpretación que satisface tales criterios es la que presenta una ventaja explicativa por sobre las demás. Una explicación satisfactoria acerca del fundamento del conocimiento en los experimentos mentales debería, en primer lugar, esclarecer el enigma planteado por Kuhn, es decir, proporcionar una interpretación explícita que muestre cómo una actividad especulativa permite obtener nueva información sobre el mundo físico. En este sentido, es conveniente adoptar un *criterio de claridad* para la valoración de los enfoques, que consiste en preferir las interpretaciones que eviten la ambigüedad en las caracterizaciones e identifiquen de manera perspicua el mecanismo de producción de nuevo conocimiento en cuestión. En segundo lugar, dado que enmarcamos nuestra investigación en la epistemología de la experimentación contemporánea, es deseable que las perspectivas en cuestión eviten asumir compromisos metafísicos innecesarios o superfluos para dar una explicación del fundamento del conocimiento de los experimentos mentales. Por lo que aplicaremos un *criterio de parsimonia* al examen de las interpretaciones existentes. En tercer lugar, suponiendo que los experimentos mentales son un tipo de exploración del mundo que comparte rasgos con otras estrategias de investigación científica, la explicación sobre sus poderes epistémicos debería poder incorporar el mecanismo de generación de nuevo conocimiento como una especie de

mecanismos más generales de producción o ampliación de conocimiento.²⁷ Por lo que adoptaremos un *criterio de incorporación* de la explicación del funcionamiento de los experimentos mentales a otros mecanismos de adquisición, ampliación y corrección de conocimiento.

Es el objetivo de este capítulo abordar de manera general el problema del fundamento del conocimiento en los experimentos mentales evaluando las ventajas y desventajas de asumir una epistemología que pueda enmarcarse en alguna de las líneas esbozadas. Con este propósito se presentan las principales tesis y argumentos de cada una de las posiciones y se evalúan parcialmente de acuerdo a los criterios sugeridos. Finalmente, las perspectivas se valoran comparativamente y se propone la adopción de una de ellas en vistas a sus ventajas explicativas.

4.2 Perspectivas empiristas

Norton (1991, 1996, 2004 y 2013) ha ofrecido una serie de argumentos para sostener que el nuevo conocimiento sobre el mundo natural que los experimentos mentales pueden proporcionar se basa en información que ya se encuentra disponible en el sistema de creencias del experimentador. De acuerdo con esta perspectiva, los experimentos mentales permiten, mediante alguna clase de inferencia, obtener nuevo conocimiento a partir de la reorganización del conocimiento empírico almacenado. Desde este punto de vista, la apariencia

²⁷ En la defensa de una posición inferencialista sobre el tipo de conocimiento involucrado en los experimentos mentales, Clatterbuck (2012) establece tres criterios de evaluación que pueden ser puestos en relación con los aquí descritos. Ellos son: modelado, parsimonia y unificación. El primero establece una demanda que desde nuestro punto de vista es difícil de evaluar a menos que se emplee comparativamente, a saber: que la perspectiva modele el proceso que *de hecho* los experimentadores usan para obtener sus conclusiones (Clatterbuck, 2012: 323). El segundo es un principio común en la epistemología empirista, la simplicidad de la explicación y modestia de los compromisos metafísicos y el tercero es esencialmente similar al que caracterizamos como la incorporación del mecanismo a instancias más generales de adquisición de conocimiento, aunque a diferencia de la propuesta de Clatterbuck, las restringimos a instancias de adquisición de conocimiento científico.

experimental es un disfraz y puede prescindirse de ella ya que el mecanismo de generación de conocimiento subyacente a los experimentos mentales es inferencial. Prueba de ello es que los experimentos mentales científicos pueden ser reconstruidos como argumentos en los que aparecen premisas que hacen referencia al conocimiento empírico (informes observacionales o principios de sentido común formados en con base en la experiencia) y premisas que enuncian principios físicos (leyes o hipótesis en general). A partir de este tipo de premisas se realizan inferencias cuyas conclusiones permiten ampliar o corregir nuestro conocimiento del mundo.

Aunque para Norton el mecanismo de adquisición de conocimiento nuevo se identifica con las inferencias lógicas, y los criterios de confiabilidad de los experimentos mentales pueden ser derivados de los criterios de validez de las inferencias, es la autoridad de la experiencia la que posibilita la obtención de consecuencias aplicables al mundo. En otros términos, basándonos en nuestra familiaridad con el mundo proporcionamos soluciones a las paradojas o los enigmas presentados por los experimentos mentales. Así, por ejemplo, en el caso del experimento mental de Lucrecio podemos derivar una solución a la aparente paradoja acerca de la infinitud del espacio recurriendo a nuestra experiencia del mundo. El experimento mental plantea el problema de decidir si el espacio es finito o infinito imaginando una situación en la cual nos encontraríamos en el límite espacial del universo. Lucrecio, ([1975]) intenta mostrar que el espacio es infinito y para ello propone imaginarnos en el límite del universo intentando arrojar una lanza hacia dicho límite. Luego propone el siguiente razonamiento: si la lanza atravesara el límite, entonces el límite no es tal, si la lanza rebotara hacia nosotros entonces habría algo más allá de este límite.

Para Norton, nuestra experiencia previa nos permite derivar la solución correcta al enigma:

[...] el espacio de nuestra experiencia nunca se agota; nunca hemos visto un límite en el espacio más allá del cual no pudiéramos pasar, a menos que ya hubiera algo más allá del límite para obstruirnos. (Norton 2004: 50)

La propuesta de Norton presenta una serie de virtudes. En primer lugar, explica el alcance epistémico de los experimentos mentales apelando a formas de conocimiento ordinario: el conocimiento se obtiene a partir de la percepción y es “almacenado” en la memoria. En segundo lugar, proporciona un criterio de confiabilidad para evaluar experimentos mentales. Y en tercer lugar, ofrece un modelo de reconstrucción de los experimentos sobre la base de esquemas inductivos y deductivos de razonamiento, que procede a partir de la explicitación de todas las premisas, incluidas aquellas que permanecen implícitas en la forma experimental del argumento. En principio, esta podría considerarse la explicación más plausible del funcionamiento de los experimentos mentales, fundamentalmente porque parece satisfacer los criterios de evaluación propuestos: es una explicación clara, integra el mecanismo de adquisición de conocimiento a otros y es parsimoniosa en tanto no se compromete con entidades ni capacidades cognitivas extravagantes. No obstante, si se la examina detenidamente, pueden señalarse algunos puntos oscuros y limitaciones.

Ante todo, Norton no ofrece ningún argumento explícito para justificar la adopción del empirismo como la perspectiva que proporciona la mejor explicación disponible del origen del conocimiento en los experimentos mentales. Se limita a sugerir que, dado que el empirismo es la epistemología dominante en la filosofía de la ciencia actual, una exégesis de los experimentos mentales que se acomode a esta epistemología debería aceptarse por *default* (Norton, 1996: 333). También señala que, dado que los experimentos mentales científicos pueden reconstruirse como argumentos inductivos o deductivos basados en información empírica, sería una extraordinaria coincidencia que otro tipo de conocimiento, uno exótico y misterioso, se ocultara bajo la clara forma de un argumento. Aunque estos puntos pueden considerarse como motivaciones para aceptar la posición de Norton, no constituyen razones concluyentes para adoptar el empirismo en la explicación del

funcionamiento de los experimentos mentales. La explicación de Norton acerca de cómo el conocimiento empírico acumulado opera en la formación de nuevo conocimiento es muy escueta. Si bien postula un mecanismo de generación de nuevo conocimiento, no aclara cómo la información sobre el mundo empírico acumulada, que estaba de alguna manera inactiva en el sistema de creencias del experimentador, desempeña un papel en la formación de nuevas creencias explícitas sobre el mundo natural resultado del experimento mental.

Algunos filósofos han señalado que la perspectiva de Norton falla en identificar el elemento distintivo de estas prácticas. Clatterbuck, por ejemplo, señala que aunque la interpretación de Norton es clara, solo ofrece un argumento negativo para sostener que los experimentos mentales son argumentos (Clatterbuck 2013: 317). Esta observación permite ampliar el argumento sobre las limitaciones de la explicación de Norton. La interpretación argumental no implica la adopción de la epistemología empirista, aunque es compatible con ella. El razonamiento que lleva a Norton a sostener la interpretación de los experimentos mentales científicos como argumentos puede ser reconstruido del siguiente modo. En primer lugar, Norton supone que en los experimentos reales el nuevo conocimiento es fruto de la intervención en la naturaleza, de la producción y observación de nuevos fenómenos y que a partir de la interpretación de los resultados experimentales es posible extraer consecuencias acerca de las teorías o hipótesis existentes. En segundo lugar, asume que dado que los experimentos mentales no introducen nuevos datos producto de la intervención y manipulación de fenómenos, no son experimentos. Norton señala explícitamente que la tesis de que los experimentos mentales científicos son argumentos depende del supuesto de acuerdo con el cual: “el pensamiento no puede conjurar nuevo conocimiento (empírico)” (Norton. 2004: 50). En tercer lugar, supone que como no son experimentos, los experimentos mentales no pueden ser otra cosa que argumentos. Finalmente, infiere que el modelo para estudiar la contribución epistémica de los experimentos mentales es la reconstrucción argumental. Lo que

puede seguirse de este análisis es que las razones de Norton para asimilar los experimentos mentales a los argumentos se derivan de una analogía negativa con los experimentos reales. Pero nada se dice en esta interpretación sobre el fundamento empírico del conocimiento en este proceso.

Una razón adicional para dudar de la interpretación de Norton es que no especifica la manera en que el conocimiento empírico que se encuentra disponible en el sistema de creencias del experimentador es recuperado en la realización de una inferencia. En otras palabras, no queda claro en su explicación cómo en la ejecución de un experimento mental, un proceso que se caracteriza por su aparente instantaneidad, los experimentadores logran tener acceso al conocimiento relevante para la realización de una inferencia. Norton supone que la ejecución de un experimento mental, en cuanto actividad psicológica, coincide con la ejecución de una inferencia (Norton, 2004: 51). Dado que, en muchos casos, la información empírica en la que se funda la verdad de algunas de las premisas, está implícita, Norton debería aceptar que en tales casos las inferencias se realizan también de manera implícita y que las premisas y el esquema de razonamiento solo se hacen explícitos cuando se reconstruye la narrativa del experimento y sus premisas formulan adecuadamente. Aquí cabe preguntarse cómo una premisa implícita conduce de manera espontánea a una conclusión nueva y cómo la información empírica de repente constituye información relevante para arribar a una conclusión que antes de la realización del experimento mental era inalcanzable.

Por otra parte, es importante señalar que para Norton la capacidad de visualizar los escenarios descritos en la narrativa del experimento mental no resulta crucial para alcanzar las conclusiones que se puedan extraer de ellos. Desde nuestro punto de vista, este elemento, que podría resultar explicativo de cómo se recupera la información empírica necesaria para la justificación de las conclusiones, queda relegado a un papel puramente heurístico. Según Norton, dado que es un hecho que podemos imaginar de manera espuria, el tipo de

representación mental asociada a los experimentos mentales no agrega ningún elemento a sus poderes epistémicos. En general, la visualización, la representación mental y otras habilidades imaginativas vinculadas a la ejecución de un experimento de este tipo se consideran irrelevantes desde el punto de vista epistémico, por lo que en su explicación no recurre a ellas para exponer cómo tienen lugar los procesos de recuperación de información y reorganización de la misma en relación con principios teóricos. Para Norton, se trata más bien de un elemento accesorio; prueba de ello es que muchos de los detalles de la visualización como el color de los objetos, pueden ser reemplazados por otros sin alterar el resultado de la inferencia. En consecuencia, podría afirmarse que, desde la perspectiva de Norton, la información empírica necesaria para la obtención de nuevo conocimiento debe encontrarse disponible o accesible de alguna manera en el sistema de creencias del experimentador; de otro modo, no podría realizarse una inferencia. Si la realización de un experimento mental coincide con una inferencia, la información empírica debería estar simultáneamente disponible y proposicionalmente formulada y a la vez implícita y, de cierto modo, inaccesible al experimentador.

Si bien estas observaciones no muestran que la explicación de Norton sea inconsistente o errónea, evidencian la fragilidad de algunos aspectos de su argumento. Aunque la posición empirista se presenta como una explicación parsimoniosa y clara del proceso de adquisición de nuevo conocimiento en los experimentos mentales, no parece hacer justicia a ciertos rasgos de esta clase de experimentos, como la instantaneidad con la que se obtienen las conclusiones o el carácter imaginativo de su narrativa. Asimismo, la adopción del empirismo por razones de economía no libra a esta perspectiva de la opacidad en la explicación del origen del conocimiento en cuestión.

Otra perspectiva que da cuenta del origen y el mecanismo de adquisición del nuevo conocimiento en los experimentos mentales, que el propio Norton menciona como antecedente de la suya, es la de Mach (1897). Si bien los

supuestos que configuran la posición de Mach son sustancialmente diferentes de los de Norton, ambas posiciones comparten un núcleo empirista.

Mach fue el primero en elaborar una caracterización explícita del funcionamiento de los experimentos mentales en ciencia. La misma depende de dos grandes supuestos, por un lado la doctrina sensacionalista de la percepción y por otro la premisa de acuerdo con la cual experimentos mentales y reales son dos instancias del método de la variación concomitante. Esta posición sostiene que una masa de conocimientos empíricos se encuentra disponible, aunque de manera no analizada e inarticulada, en el sistema de creencias de los sujetos. Mach sostiene que este conocimiento procede tanto de la experiencia pasada de los individuos como de la experiencia heredada de sus ancestros y que ambas fuentes constituyen una forma de conocimiento instintivo. El nuevo conocimiento en los experimentos mentales se apoya en este conocimiento instintivo y tiene un fundamento *a posteriori* cuando se funda en las experiencias pasadas que se forma el experimentador a partir del contacto con el mundo, y un fundamento *a priori* cuando se funda en el conocimiento que configura al sujeto cognoscente como parte de la especie.

Según Mach, la imaginación desempeña también una función importante en la ampliación del conocimiento. Esta facultad hace posible la exploración de diferentes escenarios en tanto permite combinar libremente diferentes elementos. Al confrontar estas representaciones con el conocimiento empírico acumulado, el experimentador evalúa a algunas como empíricamente imposibles. Así, por ejemplo, es posible imaginar una serie de planos inclinados totalmente lisos en los que se desplaza sin fricción una cadena, pero el conocimiento empírico de los experimentadores opera negativamente sobre esta representación, mostrando que no hay experiencia previa de máquinas de movimiento perpetuo. Para Mach, la imaginación en los experimentos mentales científicos se encuentra constreñida por el conocimiento empírico disponible. Es la autoridad del conocimiento

empírico la que garantiza el nuevo conocimiento que puede obtenerse como resultado de los experimentos mentales.

Con todo, la explicación del funcionamiento de los experimentos mentales propuesta por Mach es oscura en varios aspectos. En primer lugar, no resulta clara la relación entre el origen y la justificación del conocimiento que se deriva de la ejecución de un experimento mental. Aunque se admitiera que en el sistema de creencias de los sujetos hay algunas de carácter empírico cuyo origen es innato, y por lo tanto *a priori*, el conocimiento que debería obtenerse como resultado de la ejecución (e interpretación) de un experimento mental es *sintético a posteriori*. Por esa razón, el carácter heredado o innato de las creencias que sirven de base al nuevo conocimiento no lo justifica por su carácter instintivo o innato sino por su carácter empírico.

En segundo lugar, la relación entre el conocimiento empírico acumulado y la modificación de las creencias como resultado de la ejecución del experimento mental también es confusa. Aunque la experiencia acumulada parece ejercer una fuerza compulsiva en los experimentos mentales, excluyendo algunos escenarios como empíricamente imposibles, la realización de experimentos reales análogos a los mentales se hace necesaria en algunos casos en los que los resultados son inciertos (Mach, [1897] 1973: 455). Desde el punto de vista empirista, la experiencia (y en el caso de Mach, la sensación) siempre tienen la última palabra. Esta peculiaridad socava la relación entre el origen del conocimiento y su adecuación empírica, una conexión que de mantenerse sería una gran ventaja explicativa de la versión del empirismo que Mach defiende. Por otra parte, este supuesto no es útil para explicar el funcionamiento de los experimentos mentales que operan con cierto tipo de supuestos contrafácticos. El conocimiento empírico acumulado parece no desempeñar un papel en el examen de casos en los que se pide al experimentador extraer consecuencias de la extrapolación de principios teóricos que rigen el comportamiento del mundo subatómico a fenómenos macroscópicos. No parece posible, en principio, que la experiencia entendida

como derivada del contacto con el mundo cotidiano, actúe como limitante de los escenarios posibles.

En tercer lugar, el carácter implícito del conocimiento en cuestión también es un elemento que oscurece la explicación de Mach. El hecho de que no seamos conscientes del conocimiento empírico o de las creencias acerca del comportamiento del mundo natural que intervienen en los experimentos mentales, hace difícil distinguir los casos en los que el conocimiento empírico implícito le proporciona contenido empírico a las proposiciones que formulamos a partir de los experimentos mentales de los casos en los que no lo hace.

Por otra parte, aunque la perspectiva de Mach se enmarca dentro de la epistemología empirista, y como tal podría presumirse que proporciona una explicación parsimoniosa del funcionamiento de los experimentos mentales, asume compromisos epistémicos innecesarios. El evolucionismo que Mach adopta es superfluo en cuanto no cumple un papel en la explicación del funcionamiento de los experimentos mentales. Como señalamos, las creencias derivadas de la ejecución de un experimento mental están justificadas solo si son consistentes con el conocimiento empírico acumulado. Es decir, suponer que algunas de esas creencias sean heredadas filogenéticamente no parece aclarar cómo obtenemos conocimiento de los experimentos mentales.

Otra cuestión problemática en la explicación de Mach es que adopta un punto de vista subjetivista. Para este filósofo, la ejecución de un experimento mental es un acto introspectivo, en el que el sujeto debe representarse a sí mismo como realizando las operaciones descritas en la narrativa. La ausencia de referencia a criterios internos a partir de los cuales determinar si la ejecución se realizó "correctamente" o no, hace difícil diferenciar el tipo de actividad vinculada con la ejecución de un experimento mental de tipo de actividades imaginativas como la fantasía, el sueño, la narración de ficción.

Una posición intermedia entre la perspectiva machiana y la interpretación argumental de Norton es la de Sorensen (1992). Esta propuesta explicativa se acerca a la posición de Mach en cuanto asume un punto de partida evolucionista y naturalista, pero se diferencia de ésta en que rechaza el sensacionalismo en favor de una forma de confiabilismo respecto de la justificación del conocimiento. Por otra parte, se parece a la perspectiva de Norton en cuanto sugiere emplear esquemas de reconstrucción argumental como criterio de fiabilidad para los experimentos mentales.

Para Sorensen, los experimentos mentales evolucionaron a partir de los ordinarios por un proceso de atenuación, entendido como un proceso de abstracción. Esto supone que un experimento puede ser definido en general como un procedimiento para establecer relaciones entre variables. Los experimentos mentales son, entonces, procedimientos dependientes de la capacidad de manipular variables en la imaginación y, en este sentido, son casos límite de los experimentos reales. La habilidad de realizar experimentos mentales es una habilidad altamente teórica adquirida como producto de un proceso de evolución. De acuerdo con este paralelismo, los experimentos reales plantean y responden preguntas sobre relaciones específicas entre variables independientes y dependientes a través de un proceso de manipulación. Los experimentos mentales, por su parte, buscan lograr lo mismo por la mera reflexión sobre un diseño experimental.

De acuerdo con Sorensen, el nuevo conocimiento que los experimentos mentales pueden proporcionar se debe a la reestructuración o reorganización de las creencias del experimentador. El procedimiento por el cual se genera este nuevo conocimiento puede ser reconstruido en un argumento. Esto no implica que los experimentos mentales sean en sí mismos argumentos. Según este punto de vista, el mejor modelo disponible para explicar el funcionamiento de los experimentos mentales es el *modelo de limpieza (cleasing model)* que muestra como

[...] los experimentos mentales mejoran nuestro (sistema de) pensamiento al exponer sus irracionalidades, lo que nos conduce finalmente a una mayor racionalidad. (Sorensen, 1992: 110)

Desde esta perspectiva, los experimentos mentales muestran que una hipótesis teórica tiene ciertas consecuencias modales que son falsas en algún mundo posible. Es decir, el experimento mental procede proporcionando un escenario hipotético en el cual un determinado principio teórico es falso. Este procedimiento puede reconstruirse como un argumento modal del que se deriva una contradicción. A diferencia de los mecanismos psicológicos que dan lugar al nuevo conocimiento, las inferencias son menos oscuras y permiten esclarecer tanto los principios teóricos como los enunciados con contenido empírico involucrados en el experimento. No obstante, para Sorensen, el mecanismo de justificación de creencias de los experimentos mentales no es argumental, sino experimental.

La posición de Sorensen es difícil de evaluar de manera global ya que incorpora muchos elementos cuya plausibilidad debe valorarse por separado. Asume una epistemología empirista y evolucionista acerca del fundamento del conocimiento; interpreta los experimentos mentales como una clase de experimentos; ofrece una regimentación modal para la reconstrucción argumental; y, finalmente, supone que es posible proporcionar una teoría unificada del funcionamiento de los experimentos mentales en ciencias y filosofía. Es necesario evaluar, además, si todas estas tesis son consistentes entre sí.

Ante todo, podría decirse que la de Sorensen es una explicación parsimoniosa del funcionamiento de los experimentos mentales ya que, aunque no sostiene que los experimentos mentales funcionen simplemente reorganizando la información empírica disponible, tampoco mantiene que habiliten alguna facultad que trascienda el conocimiento empírico. Funcionan sobre la base de la habilidad evolutivamente adquirida de especular sobre el comportamiento de los fenómenos y de derivar a partir de este ejercicio

consecuencias para principios teóricos que se consideran suficientemente justificados.

Una de las mayores dificultades de esta perspectiva reside en la oscuridad con la que se explica el proceso de ampliación de conocimiento que tiene lugar en los experimentos mentales. Por un lado, Sorensen asume que se trata de un proceso lo suficientemente semejante a la experimentación como para ser considerado de la misma clase. Se trata de un proceso que incluiría la representación de una situación o un diseño experimental y la manipulación mental de variables. La garantía de que este ejercicio es legítimo o confiable se encuentra en su carácter evolutivo: los sujetos adquieren gradualmente la capacidad de omitir (la descripción de) ciertos aspectos superfluos sobre la naturaleza del mundo, de idealizar o de no realizar ciertos ensayos en la medida en la que acumulan experiencia. Por otro lado, este proceso conduce a la adecuación empírica de las creencias y a un aumento de la racionalidad (consistencia en el sistema de creencias) vinculando las afirmaciones teóricas pertinentes con los resultados experimentales. Este segundo elemento parece no depender del primero y dado que puede ser reconstruido como una inferencia, no hay buenas razones para suponer que depende de las mismas habilidades. Una interpretación alternativa podría consistir en dividir el proceso de adquisición en dos etapas o fases, una de carácter imaginativo, que consiste en la representación de un escenario posible o de un aparato experimental, y la otra de carácter inferencial, que consiste en derivar de este ejercicio consecuencias para una teoría. Esta modificación tendría el beneficio de integrar el mecanismo de adquisición de conocimiento a otras formas comúnmente admitidas de ampliación de conocimiento.

Un rasgo común de las perspectivas empiristas sostiene que la experiencia del mundo configura en los experimentadores una suerte de *intuición empírica*. Según este punto de vista, compartido con algunas diferencias por Brendel (2005) y Clatterbuck (2013), los experimentos mentales representan situaciones con las

que los experimentadores están familiarizados. Esto garantiza que las conclusiones alcanzadas en la ejecución de un experimento mental sean aplicables al mundo físico.

Para Brendel, las intuiciones son actitudes mentales proposicionales que van acompañadas de un fuerte sentido de certeza. En los experimentos mentales, son parte del conocimiento de fondo y determinan inconscientemente nuestra interpretación. Algunas intuiciones son relativamente estables y compartidas, lo que se debe al hecho de que pertenecemos a la misma especie biológica y a comunidades culturales y científicas con algún conocimiento común (Brendel, 2005: 96).

Para Clatterbuck, el conocimiento empírico acumulado por los experimentadores garantiza el funcionamiento correcto de los experimentos mentales, pero la formación de creencias a los que estos dan lugar depende de un tipo particular de inferencia inductiva. El conocimiento empírico acumulado garantiza la extensión a los casos futuros a partir de una componente objetiva y otra subjetiva. En cuanto al elemento subjetivo, el nuevo conocimiento depende de la experiencia previa de los sujetos, por ejemplo, para realizar afirmaciones sobre el comportamiento de un objeto dentro de un marco inercial es necesario que el sujeto haya experimentado esa situación. El experimento le permite al experimentador formarse explícitamente conocimiento sobre su experiencia previa invitándolo a simular y prestar atención a los detalles relevantes de esa experiencia. El éxito de estos primeros pasos del experimento mental dependerá crucialmente de si el experimentador tiene la experiencia previa relevante y de si puede manipular mentalmente sus detalles para llegar a un caso más puro e idealizado. En los casos en que el individuo no tiene conocimiento experiencial previo adecuado, el experimento no tendrá éxito. El componente objetivo que garantiza la confiabilidad de los experimentos mentales está dado por el vínculo entre los casos observados y los no observados. Este vínculo depende de que el experimentador identifique las propiedades relevantes en los casos de los que

tiene experiencia de manera que sean *proyectables* a los casos de los que no tiene experiencia. Esta proyección, sin duda, se funda en una inferencia de carácter inductivo y, como tal, resulta siempre falible. El tipo de conocimiento nuevo que los experimentos mentales pueden proporcionar consiste en ubicar los fenómenos desconocidos en la clase de fenómenos conocidos (Clatterbuck, 2013: 324).

En general, esta perspectiva puede ser criticada en dos aspectos. Uno se relaciona con la subjetividad y la relatividad de la intuición empírica, que descansa principalmente en la experiencia individual que sirve de base a su formación. Si bien este elemento teórico podría explicar la instantaneidad con la cual nos representamos un escenario descrito por un experimento mental y derivamos algunas conclusiones sobre el comportamiento de los fenómenos, no resulta evidente que las intuiciones así entendidas operen como trasfondo en casos de experimentos mentales que no se refieren a la experiencia cotidiana (como, por ejemplo, aquellos que tratan acerca del comportamiento cuántico de las entidades microscópicas). La pertenencia a una determinada comunidad epistémica también determina el papel pueden desempeñar las intuiciones en la ejecución de un experimento mental. La intuición del físico experimental difiere de la del lego, configurada por el sentido común. Por otra parte, identificar los principios teóricos relevantes parece depender más bien de la posesión del conocimiento teórico en cuestión que del hecho de haber tenido las experiencias previas relevantes para el caso. El segundo aspecto que puede criticarse en esta perspectiva se refiere a la posibilidad de admitir la importancia del conocimiento de fondo sin comprometerse con la existencia de la intuición empírica (de un tipo especial, diferente de la percepción ordinaria pero derivada de ella). La noción de intuición misma suscita un conjunto de problemas filosóficos que parece innecesario incorporar a la discusión sobre la epistemología de los experimentos mentales.

4.3.1 Perspectivas Aprioristas I: Racionalismo

Ørsted ([1811] 1998), Koyré (1973) y Brown (2001 y 2011) comparten, con algunas diferencias, un punto de vista racionalista acerca del mecanismo de adquisición de nuevo conocimiento en los experimentos mentales. En general, estos autores sostienen que estos experimentos son capaces de proporcionar conocimiento del mundo natural, prescindiendo de datos empíricos nuevos y coinciden en concebir a las matemáticas como paradigma de esta forma de conocimiento. Aunque estas posiciones tienen un trasfondo común, presentan importantes diferencias. El primer punto de divergencia se encuentra en lo que en cada caso se asocia con la noción de conocimiento *a priori*. Aunque ninguno de los autores ofrece una definición explícita, la caracterización del conocimiento *a priori* como aquel conocimiento cuya justificación es independiente de la experiencia podría aplicarse a las tres variantes dentro la posición racionalista. Lo que diferencia a las posiciones es la interpretación del mecanismo a través del cual se arriba a esta forma de conocimiento. Para el caso de Ørsted y Koyré los experimentos mentales son procedimientos en los que la imaginación cumple un papel importante en la abstracción o supresión de alguna propiedad de los objetos o procesos y la razón cumple un rol relevante en la derivación de consecuencias a partir del examen de una situación idealizada. Para Ørsted (1811: 295) en particular, aunque se trata de una actividad especulativa, en los experimentos mentales la mente no es completamente contemplativa sino que es activa al derivar deductivamente, de una situación hipotética, consecuencias que son aplicables a los fenómenos del mundo real. En el caso de Koyré, el uso de los experimentos mentales en Galileo es representativo del pensamiento moderno, caracterizado por la matematización del espacio y la precisión de los resultados. Desde su punto de vista, en sus experimentos mentales Galileo encarna en objetos imaginarios las exigencias de la teoría. De este modo, logra “concretar” la teoría y comprender lo real sensible como una desviación del modelo puro que nos ofrece:

[...] es imposible producir una superficie plana que sea verdaderamente plana, no hay ni puede haber cuerpos perfectamente rígidos o cuerpos perfectamente elásticos. Entre el dato empírico y el objeto teórico siempre queda una distancia imposible de salvar. La imaginación suprime esa separación. No se preocupa por las limitaciones que nos impone lo real. Realiza lo ideal e incluso lo imposible. Opera con objetos teóricamente perfectos. (Koyré, [1973] 1991: 207)

Brown, en cambio, identifica el mecanismo de adquisición de conocimiento en los experimentos mentales con una forma de intuición racional, a la que concibe como una forma de conocimiento no inferencial. Como se intentará mostrar más adelante, la principal dificultad que esta posición enfrenta es proporcionar una explicación consistente de cómo una práctica análoga a la experimentación real habilita la contemplación de entidades platónicas.

Otra diferencia importante entre los puntos de vista reunidos que agrupamos bajo la concepción trascendentalista, se encuentra en la noción de justificación que aparece en los tratamientos de Koyré y Brown. Los experimentos mentales de Galileo pueden, según la interpretación de Koyré, confirmar teorías encarnando en objetos ideales, lo que es imposible en los casos concretos. Hacen posible de este modo, “comprender lo real sensible como una desviación que el modelo puro nos ofrece” (Koyré, [1973] 1991: 250). Aunque Koyré no elabora la noción de justificación que subyace a estos casos, una manera de caracterizar la relación evidencial es considerar a la ejecución de un experimento mental como un ejercicio especulativo que proporciona un caso paradigmático de una ley general. Para Brown los experimentos mentales en física incluyen un elemento emparentado con la experiencia. Por lo general, en ellos el experimentador debe “manipular” un escenario imaginario y “visualizar” el resultado (Brown, 2011: 17-18). A diferencia de los experimentos ordinarios que proporcionan datos a partir de los cuales los científicos abstraen las características de los fenómenos, los experimentos mentales son capaces de presentar fenómenos puros, por lo que poseen poderes epistémicos extraordinarios: pueden justificar directamente principios teóricos.

Entre las perspectivas racionalistas, la de Brown es la que merece mayor atención ya que se desarrolla en el marco de un tratamiento sistemático de los experimentos mentales y es la única en la que se elaboran argumentos para sostener una versión del racionalismo como forma de conocimiento típica de algunos experimentos mentales. Para defender su posición, Brown apela a una analogía entre los métodos empleados en matemática y los métodos empleados en la ciencia física; defiende una concepción realista de las leyes y elabora una noción de fenómeno *a priori*. Brown (2001 y 2011) sostiene una versión radical del apriorismo, de acuerdo con la cual algunos experimentos mentales en física, los experimentos mentales platónicos, comparten rasgos esenciales con las estrategias metodológicas utilizadas en matemáticas. El empleo de diagramas y visualizaciones es una metodología viable para apoyar hipótesis (a la par de las demostraciones matemáticas tradicionales) debido a que, para Brown, estas formas de representación habilitan una vía no inferencial de conocimiento de verdades, proporcionando evidencia para los teoremas. El argumento que Brown emplea para justificar la hipótesis de que los experimentos mentales son capaces de probar teorías, supone que lo que es aplicable a las estrategias metodológicas en matemáticas también es aplicable a las ciencias fácticas. Llamaremos a esta tesis el *argumento de la transitividad de la evidencia*. Desde este punto de vista, la evidencia es transitiva, lo que implica que si un diagrama es evidencia de un teorema (o al menos, evidencia de la existencia de una prueba formal) un experimento mental puede ser evidencia para un principio teórico:

Por lo tanto, si la 'prueba breve' es evidencia de la existencia de una prueba formal y la derivación formal es evidencia para el teorema, entonces la 'prueba breve' es evidencia para el teorema. De manera similar, el experimento mental de Galileo es evidencia para la afirmación de que todos los cuerpos caen a la misma velocidad. (Brown, 2011: 110)

Como las visualizaciones en matemática abrevian la prueba de un teorema, los experimentos mentales en física funcionan proporcionando imágenes (fenómenos *a priori*) que facilitan la intelección de leyes naturales. Una parte

importante de la argumentación de Brown a favor de una interpretación apriorista de los experimentos mentales se apoya en esta supuesta semejanza entre los experimentos mentales y los diagramas y visualizaciones en matemática.

Un examen del funcionamiento de las visualizaciones en matemática excede los límites de este trabajo, no obstante, podrían señalarse algunas limitaciones explicativas de esta analogía entre experimentos mentales y visualizaciones en el contexto de una prueba matemática. Por una parte, como se desarrolla en los próximos párrafos la noción de conocimiento *a priori* empleada por Brown involucra un procedimiento análogo al de la percepción sensible. La experiencia visual que proporciona evidencia a favor de un principio o ley es conocimiento *a posteriori*, lo que resulta problemático para defender la tesis de acuerdo con la cual el conocimiento involucrado constituye una forma de conocimiento *a priori*. Por otra parte, la perspectiva argumental ha argüido que la posibilidad de reformular los experimentos mentales platónicos como argumentos proporciona una explicación que integra el mecanismo empleado por estos experimentos a otras formas de adquisición de conocimiento. A esta objeción podría agregarse que la posibilidad de formular proposicionalmente las premisas y las conclusiones y de identificar el mecanismo de inferencia, relegaría a una función puramente heurística el papel de las visualizaciones. Dada la oscuridad del mecanismo invocado por Brown y la perspicuidad de las inferencias lógicas, parece más razonable aceptar una explicación inferencialista del mecanismo de formación de creencias. En este caso, al argumento de la transitividad de Brown podría objetarse que la visualización facilita la inferencia pero no justifica las conclusiones. Es lo que ocurre, por ejemplo, en el caso de la demostración de teoremas de la geometría de Euclides apelando a figuras y diagramas: la prueba de un teorema depende de su deducción a partir de los postulados y axiomas, no de la visualización de las figuras.

Brown sostiene que la intuición de leyes es una explicación aceptable de cómo los experimentos mentales platónicos proporcionan conocimiento nuevo

del mundo natural sin intervenir activamente en él, es decir, sin introducir datos empíricos nuevos. Desde este punto de vista, en casos excepcionales, como el del experimento de los cuerpos en caída libre de Galileo, obtenemos conocimiento a través de un mecanismo no inferencial diferente de la experiencia, en virtud del cual accedemos directamente a leyes de la naturaleza.

La transición de la teoría aristotélica a la teoría galileana no es un caso en el que se realiza un simple ajuste en la vieja teoría. Quizás la transición era la más simple, pero no había ninguna razón para hacerla. (Brown, 2011: 100)

Es debido a que la intuición racional proporciona un acceso directo a entidades abstractas que los experimentos mentales *platónicos* como el de Galileo justifican hipótesis o teorías. Pese a la ausencia de nuevos datos observacionales, estos posibilitan la evaluación de teorías y aumentan el grado de creencia racional en una hipótesis postulada. En palabras de Brown, que los experimentos mentales proporcionan conocimiento *a priori* significa que algunas leyes de la naturaleza pueden ser vistas “de la misma manera que algunos objetos matemáticos puede ser vistos” (Brown, 2011: 98). A su vez, la existencia independiente de leyes de la naturaleza, le da a los experimentadores “algo que ver” (Brown, 2011: 98).

En su interpretación de los experimentos mentales platónicos Brown introduce una versión epistemológica del platonismo nomológico. En esta perspectiva, la posibilidad del conocimiento *a priori* está relacionada con una concepción realista de las leyes del siguiente modo: dado que supone que las leyes son relaciones entre universales que existen de manera independiente del mundo físico, se asume que es posible conocerlas intuitivamente de la misma manera en que se captan los axiomas matemáticos. Según esta posición, las leyes no supervienen a los hechos, sino que consisten en relaciones entre propiedades. Su perspectiva se funda en la concepción de las leyes desarrollada por Dretske (1977), Tooley (1977) y Armstrong (1978 y 1983) (en adelante DTA). Las regularidades expresadas por una ley se sostienen en virtud de una relación

irreductible entre universales, denominada relación de *necesitación*. Las relaciones de *necesitación* son universales de segundo orden que solo están contingentemente instanciadas, por lo que su fuerza modal está restringida a un subconjunto de mundos posibles. Esto implica que la relación de *necesitación* entre universales puede implicar regularidades empíricas pero no está implicada por ellas. De modo que, es verdadero que $N(F, G) \rightarrow (\forall x)(Fx \rightarrow Gx)$, pero no es verdadera la conversa $(\forall x)(Fx \rightarrow Gx) \rightarrow N(F, G)$, donde F y G son universales y N es la relación de *necesitación* que los liga. En las versiones de Dretske y Armstrong, las leyes no existen en los mundos posibles en los cuales los universales en cuestión no están instanciados, pero Brown adopta una versión platonista de esta tesis, también suscrita por Tooley (1977: 671), de acuerdo con la cual las leyes existen independientemente de que sean o no instanciadas. A esta perspectiva Brown le agrega un componente epistémico que no forma parte de la propuesta general de la DTA. Según Brown, aunque usualmente accedemos a las leyes por medio del conocimiento de sus instancias, en ciertos casos especiales podemos evitar esta mediación y acceder directamente a los universales relacionados.

La perspectiva de Armstrong, Dretske y Tooley sobre las leyes de la naturaleza es pura metafísica. Proporciona una explicación de la naturaleza de las leyes y explica las regularidades que se dan. No obstante, la manera en que aprendemos acerca de las leyes es similar a la del empirista acérrimo: por las instancias; vemos cuervos, nunca "lo cuervo". Yo quiero agregar un aspecto epistemológico a esta explicación metafísica de las leyes. Es una pena tener este aparato metafísico disponible y no ponerlo a trabajar. La afirmación que deseo defender es esta: en circunstancias muy especiales podemos ver las leyes de la naturaleza: no las regularidades sino los patrones abstractos mismos. Esto nos da conocimiento *a priori* del mundo físico. ¿En qué circunstancias especiales? En los experimentos mentales; aquí es cuando podemos ver en el reino de lo abstracto. (Brown, 1994: 98)

Brown justifica la posibilidad de ganar conocimiento nuevo de manera independiente de la experiencia sensible a partir de la analogía entre la captación intuitiva y la percepción visual. Aunque se ha argüido que, a diferencia de la percepción sensible, la intuición racional es una forma de conocimiento

completamente misterioso, Brown señala que la percepción de objetos físicos es un proceso solo parcialmente explicado y que la misma falencia podría admitirse en una explicación del conocimiento de objetos abstractos (Brown, 2011: 108). Llamaremos a este argumento: *argumento de la parcialidad de la explicación de la percepción visual*. Según su perspectiva, entendemos cómo los fotones emitidos por un objeto interactúan con los conos y bastones que se encuentran en la retina y cómo una señal se transmite a través del nervio óptico hasta la corteza visual, pero no entendemos cómo este proceso causa creencias acerca de los objetos físicos. Por lo tanto, la misma carencia en la explicación es admisible en la postulación de conocimiento intuitivo de verdades abstractas.

Dado que la noción de conocimiento *a priori* de entidades abstractas es sumamente problemática desde el punto de vista de la teoría causal del conocimiento, ampliamente aceptada en la epistemología contemporánea, Brown señala las limitaciones explicativas que, desde su punto de vista tiene, esta perspectiva. Llamaremos a este: *argumento de la deficiencia explicativa de la teoría causal del conocimiento*. Sostiene que la teoría causal resulta problemática para dar cuenta de casos que son generalmente admitidos como conocimiento, tales como las generalizaciones y los casos cadenas de interacción causal con objetos muy remotos. Por lo que es plausible admitir una forma de interacción con objetos abstractos que operan como causantes de ciertas creencias. El paradigma de esta forma de conocimiento es el matemático. La intuición es la mejor explicación disponible de nuestro conocimiento de verdades matemáticas, por lo que, de acuerdo con Brown, debemos seguir a Gödel en este punto y comprometernos con la existencia de entidades abstractas como causa de nuestras creencias acerca del mundo natural:

[...] a pesar de su distancia con la experiencia sensible, tenemos también algo como una percepción de los objetos de la teoría de conjuntos, como se ve a partir del hecho de que los axiomas nos compelen a creer su verdad. No veo ninguna razón por la que deberíamos tener menos confianza en este tipo de percepción,

es decir, en la intuición matemática, que en la percepción sensorial. (Gödel, 1947: 268)

Aquí Brown agrega un elemento a su concepción apriorista y afirma que, a pesar de que la captación de entidades abstractas es *a priori*, es posible tener creencias falsas acerca de entidades abstractas aunque estas puedan ser captadas directamente. Para ilustrar esta afirmación toma un ejemplo de la teoría de conjuntos. Del axioma de comprensión de Cantor, de acuerdo con el cual para cualquier condición hay un conjunto de objetos que la satisface, es posible derivar contradicciones si se intenta construir conjuntos bajo determinadas condiciones, como mostró Russell. La paradoja de Russell condujo, según Brown a la revisión del axioma y del concepto de conjunto y a un consecuente cambio conceptual. De este ejemplo se sigue que aunque la captación *a priori* de objetos matemáticos sea falible, este aspecto no afecta al platonismo. El ojo de la mente, según Brown, es falible, como la percepción sensible es falible, por lo tanto, apriorismo y falibilismo son compatibles. (Brown, 2008: 24). En el contexto de las ciencias naturales, la captación de leyes, entendidas como universales de segundo nivel, también está sujeta a errores. Por lo que en definitiva la perspectiva de Brown se compromete con una facultad de conocimiento *a priori* capaz de captar leyes de la naturaleza que es al mismo tiempo falible.

A pesar de los intentos de Brown por defender su perspectiva de las posibles objeciones, existen dificultades importantes en su argumentación. Uno de los puntos en el que convergen las críticas es en que hay diferencias entre la captación de entidades abstractas y la percepción sensible (véase, por ejemplo Norton, 1996: 359). La principal diferencia entre ambas formas de conocimiento reside en que la intuición *a priori* carece de criterios de corrección, a diferencia de la percepción sensible que está gobernada por una variedad de regularidades que hacen posible ejercer un control sobre su confiabilidad. Estas regularidades permiten explicar, por ejemplo, las ilusiones ópticas. La intuición de entidades abstractas carece de estos criterios y es, por tanto, ininteligible (Häggqvist, 1996:

63-64). Por consiguiente, la postulación de este mecanismo de conocimiento no permite explicar el conocimiento mundo físico. El mismo Brown admite que esta forma de intuición es falible, pero no formula criterios para detectar ilusiones o para corregir errores, de modo que no es posible determinar cuándo la captación de entidades abstractas es exitosa y cuando no lo es.

Otra dificultad importante en la argumentación de Brown reside en la naturaleza del vínculo entre lo que denomina “fenómenos *a priori*” y las leyes de la naturaleza. En general, la caracterización de esta relación es, cuando menos, oscura. Brown parece identificar la teoría del fenómeno de Bogen y Woodward (1988) con un constructivismo según el cual los fenómenos son producto de un trabajo de abstracción de las regularidades a partir de la multiplicidad que se presenta en la percepción. Asume que “los científicos construyen fenómenos a partir de una gran masa de datos singularizando lo que toman por clases naturales genuinas” (Brown, 1994: 125). Esta interpretación es compatible con la postulación del conocimiento *a priori* de las leyes entendidas como universales de segundo orden, pero no queda claro de qué manera, según Brown, los fenómenos que representan clases naturales están conectados con las leyes. Una interpretación posible de este vínculo consiste en considerar que lo que se intuye en los experimentos mentales platónicos son fenómenos que ejemplifican clases naturales o exhiben instancias “ideales” de las leyes. Esto implicaría que el conocimiento *a priori* de las leyes consiste en realizar una especie de inferencia inductiva desde el fenómeno concebido *a priori* a la ley. Esta es una interpretación plausible de cómo obtenemos conocimiento en el marco conceptual presentado, pero el propio Brown no expone este vínculo de manera suficientemente inteligible. Por otra parte, parece contradecir la naturaleza *a priori* (no inferencial) del conocimiento que los experimentos mentales platónicos pueden proporcionar de acuerdo con Brown.

Un problema adicional en la interpretación de la noción de fenómeno *a priori* realizada por Brown es que extrae una conclusión que no se sigue del marco

teórico referido, esto es, de la teoría del fenómeno de Bogen y Woodward. Según esta perspectiva, los datos son registros públicos producidos por mediciones y experimentos, que sirven como evidencia de la existencia de un fenómeno o de su posesión de ciertas características. Los datos guardan una relación causal con los fenómenos pero reflejan, además, aspectos idiosincráticos de las operaciones de medición y experimentación de las que son el resultado. Estos aspectos se vinculan, entre otras cosas, con las características de la situación experimental, la función que se espera que los datos cumplan respecto de la teoría, su grado de accesibilidad, los criterios de confiabilidad, la facilidad con la que pueden ser identificados, clasificados, medidos, analizados, o reproducidos. En el proceso de interpretación de los datos que sirven de evidencia del fenómeno está involucrado el uso de argumentos, técnicas de análisis y patrones de razonamiento contrafáctico, independientes del conjunto de supuestos teóricos (Bogen y Woodward, 1988: 306). Desde esta perspectiva, no se postulan procesos de conocimiento de los fenómenos que prescindan de los datos. Por lo tanto, si bien sería posible derivar, de la diferencia entre datos y fenómenos, la conclusión de que estos últimos pueden existir con independencia de nuestro conocimiento acerca de los primeros, esto no implica, desde el marco teórico establecido, que podamos conocerlos prescindiendo de los datos que son evidencia de los fenómenos.

Una última dificultad en la argumentación de Brown es la falta de especificación de la relación entre las operaciones de captación racional y visualización que forman parte de la explicación del mecanismo de producción de nuevo conocimiento en los experimentos mentales. Al respecto, no precisa si la captación de leyes y la visualización de fenómenos se realizan en una misma operación o si se trata de dos operaciones diferentes. Esta explicación tiene una importancia fundamental dado que Brown asume que conocimiento *a priori* significa captación intuitiva de entidades que existen de manera independiente (Brown, 2011: 108). En su argumentación expone por un lado una alteridad entre

datos y fenómenos y por otro una similitud entre percibir datos de la experiencia y visualizar un fenómeno. Afirma que los datos de la observación no cumplen funciones relevantes en la justificación de hipótesis científicas y sostiene que esta función corresponde a los fenómenos, entendidos como modos estables en los que se articula la naturaleza. Por otro lado, asume que en los experimentos mentales visualizamos fenómenos. Aunque Brown no aclara en qué consiste la “visualización de los fenómenos”, podría interpretarse que esta operación es semejante a la percepción de una disposición espacio temporal de objetos. Si lo que ocurre en los experimentos mentales platónicos es una captación *a priori* de propiedades de clase natural, entonces, el aspecto “visualizable” de los fenómenos no es relevante, dentro del marco conceptual establecido por el propio Brown, para afirmar que es posible el conocimiento *a priori* de fenómenos. Para sostener esta idea basta con postular que es posible captar intuitivamente regularidades naturales.

En síntesis, la explicación del mecanismo de conocimiento *a priori* postulado por Brown resulta deficiente por varias razones. En cuanto a la analogía invocada, su empleo no resulta esclarecedor ya que se funda en un argumento por *default*: la misma oscuridad con que se explica la percepción visual debe ser admitida en la explicación del conocimiento *a priori*. Asimismo, la falta de criterios de corrección para detectar errores en la percepción de entidades platónicas no contribuye a aclarar el mecanismo de adquisición de nuevo conocimiento que se postula y es un elemento más para socavar la analogía propuesta. Por otra parte, las características del conocimiento *a priori* difícilmente puedan ser atribuidas al mecanismo de conocimiento empleado en los experimentos mentales platónicos: lo que se visualiza en estos casos es un fenómeno que ejemplifica una ley, por lo que el conocimiento de las leyes relevantes puede ser interpretado como el producto de una inferencia. La postulación de fenómenos *a priori* como intermediarios en la captación de leyes naturales, lejos de favorecer la aceptación de la tesis del conocimiento *a priori*, es una razón adicional para dudar de su

plausibilidad. Por último, la actividad de “visualización” de fenómenos *a priori* entendida como una actividad análoga a la percepción de una disposición espacio temporal de objetos, no parece cumplir ninguna función relevante en la intuición de leyes.

Otras razones para considerar que la perspectiva de Brown presenta dificultades importantes son su compromiso con una metafísica inflacionaria y la falta de integración con otras formas de adquisición de conocimiento en la práctica científica. Brown sostiene que los experimentos mentales habilitan una vía de conocimiento extraordinario, trascendente de nuestra relación con el mundo empírico pero capaz de justificar hipótesis que se refieren a éste. Esta facultad, además, requiere como objeto de conocimiento la existencia de leyes entendidas como universales de segundo orden, entidades platónicas subsistentes a los fenómenos que las instancian. Su intento de asimilar esta vía de acceso al conocimiento de verdades matemáticas es, cuando menos, dudoso. Es importante señalar, además, que este mecanismo de generación de nuevo conocimiento del mundo físico es excepcional y solo tiene lugar en los experimentos mentales platónicos, poco frecuentes en la práctica científica.²⁸ En todos los demás casos, el mecanismo de adquisición de nuevo conocimiento es asimilable a procesos más usuales que dependen de conocimiento sobre el mundo empírico adquirido por medios ordinarios (la percepción sensible y en general el contacto con el mundo) y que no requieren de la concurrencias de facultades que trasciendan el conocimiento empírico. Brown afirma que incluso en los casos en los que los experimentos mentales admitan una reconstrucción argumental, no es posible reducirlos a inferencias (Brown 2004a: 35). Dado que los experimentos mentales apelan a representaciones (imaginar cierta disposición espacio temporal de objetos) nos llevan “desde una percepción a una proposición”. No obstante, estas afirmaciones no proporcionan ningún argumento

²⁸ Brown (2004b y 2011) solo presenta un número muy reducido de ejemplos de experimentos mentales platónicos: el experimento de los cuerpos en caída libre de Galileo, el experimento de la *vis viva* de Leibniz y el experimento EPR.

para sostener que el recurso a la visualización es indispensable para la obtención de conclusiones en los experimentos mentales y que no se trata de un elemento puramente heurístico.

4.3.2 Perspectivas aprioristas II: Trascendentalismo

Marco Buzzoni (2010, 2012, 2013, 2017 y 2018) ha defendido una interpretación de acuerdo con la cual el conocimiento en los experimentos mentales en ciencias fácticas depende de una forma de conocimiento *a priori* funcional. De acuerdo con su enfoque, el tipo de conocimiento que los experimentos mentales pueden proporcionar es sintético *a priori* y tiene su fundamento en la facultad de la razón (Buzzoni, 2012: 95). La interpretación funcional del *a priori* da lugar a una lectura trascendental-operacional de los experimentos mentales. Estos experimentos son condiciones de posibilidad de los reales ya que, si la mente no pudiera razonar con independencia de la experiencia, seríamos incapaces de formular hipótesis y de diseñar experimentos para contrastarlas (Buzzoni, 2010: 5). Aunque la distinción entre experimentos reales y mentales no puede suprimirse, ya que consiste básicamente en la distinción entre el dominio reflexivo de la mente y la realidad, el carácter funcional del conocimiento *a priori* implica que a un nivel metodológico ambos tipos de experimentos funcionan del mismo modo. Esto es así porque un experimento mental anticipa una situación experimental hipotética. En base al conocimiento previo sabemos que ciertas intervenciones en el dispositivo experimental afectarán el valor de determinadas variables. En este sentido, el conocimiento previo, que procede de las intervenciones operacionales sobre la realidad, hace superflua la ejecución del experimento. Dado que los experimentos mentales carecen de contenido empírico, lo adquieren por referencia explícita o implícita a la experiencia.

En un sentido similar al que intuiciones y conceptos se complementan en la perspectiva kantiana, Buzzoni sostiene que existe una mutua dependencia entre los experimentos mentales y reales. Los primeros son condición de posibilidad de los segundos y los segundos proporcionan las bases de confirmación para los principios que sirven de fundamento a los primeros. En este sentido, Buzzoni reinterpreta el *dictum* kantiano “los pensamientos sin contenido son vacíos; las intuiciones sin conceptos son ciegos” (Kant 1781, A51/B75) para aplicarlo al conocimiento experimental: “experimentos mentales (empíricos) sin experimentos reales son vacíos, experimentos reales sin experimentos mentales son (ciegos)” (Buzzoni, 2012: 100).

De acuerdo con este punto de vista, los experimentos científicos son preguntas planteadas a la naturaleza. En ellos el experimentador no desempeña el papel de observador pasivo, sino el de juez que compele a la naturaleza a responder las preguntas que ha formulado. En virtud de una forma de conocimiento especulativa, el experimentador es capaz de anticipar hipotéticamente el veredicto de la experiencia.

La perspectiva de Buzzoni representa un intento audaz por responder al interrogante kuhiano acerca de cómo es posible adquirir especulativamente nuevo conocimiento sobre el mundo físico. Acierta en recuperar un aspecto esencial de los experimentos mentales: funcionan razonando sobre casos particulares y, por esa razón, dependen de elementos concretos que son, al menos en principio, reproducibles en situaciones espaciotemporales (Buzzoni, 2012: 93). En este sentido, su vínculo con los experimentos ordinarios es claro y su diferencia con éstos también lo es: los experimentos mentales consisten en anticipar una situación posible que en ocasiones incluye la descripción de un dispositivo experimental más o menos complejo y carecen de la intervención en la naturaleza, que proporciona una “respuesta” concreta a la “pregunta” planteada por el experimento. De acuerdo con Buzzoni, si entendemos la noción de *a priori* como condición de posibilidad de la experiencia, algunos experimentos mentales

permiten anticipar la respuesta correcta, pero el veredicto de la experiencia será siempre complementario como criterio de confiabilidad para la exploración hipotética que posibilitan los experimentos mentales. En esto reside su complementariedad.

Buzzoni (2010 y 2012) señala la ausencia de interpretaciones kantianas acerca del funcionamiento de los experimentos mentales y se propone suplir esta vacancia. Sugiere que Ørsted (1777-1851) fue representante de esta perspectiva, la que presenta como superadora de la disputa entre empirismo y racionalismo en el análisis epistémico de los experimentos mentales. Asimismo, enfatiza las diferencias entre su posición y la que denominamos constructivista (representada inicialmente por Kuhn). Es importante señalar que la perspectiva de Buzzoni se presenta como una aplicación de la epistemología kantiana al examen del funcionamiento de los experimentos mentales científicos, pero que no se ocupa de los experimentos mentales filosóficos formulados o criticados por Kant (véase Virvidakis, 2012: 127-142).

No obstante su originalidad, la perspectiva no parece arrojar luz sobre el origen y el mecanismo de obtención de nuevo conocimiento. Desde este punto de vista los experimentos mentales son estrategias metodológicas que aplican el método de la variación concomitante para averiguar como un aparato experimental varía en respuesta a ciertas intervenciones específicas y que dependen de ciertas hipótesis subyacentes (Buzzoni, 2009: 5). Si bien esta caracterización de los experimentos mentales que podría considerarse aceptable: no resulta explicativo que la habilidad de anticipar el comportamiento de la naturaleza dependa de estructuras trascendentales que configuran toda la experiencia posible. Es decir, la misma interpretación sobre cómo ganamos conocimiento en los experimentos mentales podría sostenerse bajo un supuesto empirista, esto es, podría explicarse nuestra capacidad de anticiparnos a la experiencia apelando a la experiencia pasada o acumulada sobre el comportamiento de los sistemas físicos o sobre las relaciones entre variables

particulares. La apelación a la estrategia trascendental no parece esclarecer el mecanismo de obtención de nuevo conocimiento, al menos en los casos de experimentos mentales realizables.

4.3 Perspectivas constructivistas

Las perspectivas constructivistas sugieren que los experimentos mentales pueden contribuir a una nueva comprensión del mundo natural porque son capaces de operar reformas en el aparato conceptual de los científicos. Esta posición, sustanciada inicialmente por Kuhn (1964), fue posteriormente recuperada y ampliada por Gendler (1998, 2000, 2002 y 2007). Su tesis principal es que los experimentos mentales eficaces son capaces de revelar lagunas en el sistema conceptual de los experimentadores, forzándolos a reconocer contradicciones en un enfoque teórico o mostrando su incapacidad para dar sentido a un fenómeno que debería ser explicado por la teoría. Denominamos a esta posición *constructivismo* para referirnos a la observación de Kuhn, según la cual en los experimentos mentales los científicos aprenden simultáneamente algo nuevo acerca del mundo y acerca de su sistema conceptual (Kuhn, [1964] 1977: 38). Dado que esta clase de experimentos no introduce nuevos datos, podría suponerse que su función se restringe a la detección de inconsistencias. No obstante, dada la dependencia que la perspectiva kuhniana introduce entre el contenido empírico de las teorías y el paradigma respecto del cual los datos resultan relevantes, el efecto de un experimento mental que exhibe algún tipo de inconsistencia en una teoría o la inaplicabilidad de un concepto es, a la vez, un nuevo conocimiento sobre el sistema conceptual y sobre el mundo. Este efecto es comparable al que producen las observaciones y los experimentos reales cuando revelan la existencia de fenómenos hasta entonces desconocidos. Cuando la

naturaleza no se ajusta a las expectativas (predicciones) teóricas, esto puede sugerir revisiones en el sistema teórico.

Kuhn ([1964] 1977) interpreta a partir de consideraciones históricas, el papel que los experimentos mentales han desempeñado respecto de teorías físicas y determina sobre la base de este análisis ciertas condiciones mínimas de confiabilidad para estos experimentos. Con este propósito, revisa un experimento mental de Galileo, presentado con la finalidad de derivar contradicciones en la noción aristotélica de movimiento. Kuhn revisa el experimento mental que Galileo propone en la Primera Jornada de los *Diálogos sobre los dos máximos sistemas del mundo* (Galileo [1632] 1967: 23). En el diálogo Salviati les pide a sus interlocutores que imaginen dos planos *CB* vertical y *CA* inclinado, levantados hasta la misma distancia vertical sobre un plano horizontal *AB*. A lo largo de estos hay que imaginar dos cuerpos que ruedan sin fricción desde un punto de partida común *C*. Cuando los cuerpos que se deslizan llegan a los puntos *A* y *B* éstos habrían adquirido el mismo ímpetu o velocidad, esto es, la velocidad necesaria para devolverlos al punto de partida. Aquí se pregunta: ¿cuál de los cuerpos se mueve más rápido? De acuerdo con Kuhn, ([1964] 1977: 31), empleando el concepto de velocidad aristotélico nos vemos forzados a admitir que el movimiento a lo largo del plano perpendicular es, al mismo tiempo, más rápido, igual y más lento que el movimiento a lo largo del plano inclinado. Según esta interpretación, Galileo quiere que sus lectores se percaten de que la velocidad no debiera atribuirse al conjunto del movimiento sino más bien a sus partes. La discusión es una propedéutica para la discusión del movimiento uniforme y el acelerado. De acuerdo con Aristóteles, el movimiento es cualquier tipo de cambio. A partir de esta noción más básica, define la noción de “más veloz” que es objeto de la crítica galileana y que Kuhn recupera como caso de análisis. Para Aristóteles, “la más rápida de dos cosas atraviesa una mayor magnitud en un tiempo igual, una magnitud igual en menos tiempo y una menor magnitud en menos tiempo” (*Física 2*: 232a28-31). Sintéticamente, lo que el experimento muestra es que de la

noción Aristotélica de velocidad depende de un criterio borroso para identificar el movimiento más rápido cuando se intenta determinar comparativamente el movimiento de móviles en planos de diferente longitud, diferente inclinación y dividiendo el movimiento en distintos segmentos. La interpretación galileana sugiere que es necesario distinguir entre velocidad instantánea y velocidad promedio:

Más rápido significa algo distinto cuando se le aplica, por un lado, a la comparación de la rapidez instantánea del movimiento en instantes determinados y, por el otro, a la comparación de los tiempos necesarios para que se complete el total de los movimientos especificados. (Kuhn, [1964] 1977: 35)

Sobre la base de este examen, Kuhn sostiene que de la ejecución de este experimento se aprende algo acerca de los conceptos y también algo acerca del mundo. En este caso, al aprender algo acerca del concepto de velocidad, los lectores de Galileo aprenden también algo acerca de cómo se mueven los cuerpos. Esto ocurre porque el experimento exhibe un tipo de inconsistencia local en el concepto aristotélico de velocidad: los defectos de la noción no estriban en su inconsistencia lógica sino en que dicha noción no encaja en la totalidad de la estructura fina del mundo al cual se pretendía aplicar. Kuhn concluye que, desde el punto de vista histórico, experimentos como el de Galileo son herramientas analíticas esenciales que se emplean durante los procesos de crisis paradigmática y contribuyen a promover la reforma conceptual básica. El resultado de esta clase de experimentos puede ser el mismo que el de la revolución científica: hacer posible que el científico emplee como parte integral de su conocimiento una serie de premisas que no le eran accesibles antes de la ejecución del experimento mental. La condición de adecuación que Kuhn deriva de este examen es la siguiente: los experimentos mentales eficaces presentan una paradoja (entendiendo por ello una discordancia entre el aparato conceptual vigente y la situación imaginada) porque permiten a los experimentadores emplear los conceptos ordinarios de la misma manera en que los emplea habitualmente y deducir de la ejecución del experimento (entendiendo por ello la

lectura de la narrativa y la consideración de las diferentes situaciones y preguntas planteadas por ésta) un conflicto entre el experimento mental y la naturaleza.

Gendler (1998, 2000, 2002, 2004 y 2013) aporta otros elementos a la explicación kuhniana del origen y el mecanismo de obtención de nuevo conocimiento en los experimentos mentales. De acuerdo con esta perspectiva, los experimentos mentales pueden operar una reconfiguración de los compromisos conceptuales de los científicos a partir de la presentación de un fenómeno de una manera novedosa. En el caso del experimento de Galileo de los cuerpos en caída libre:

Lo que hace el Galileo es proporcionar al aristotélico un espacio conceptual para una nueva noción del tipo de cosa que la velocidad natural podría ser: una constante independientemente comprobable en lugar de una función de algo más primitivo (es decir, más que una función del peso). De esta manera, al permitir que el aristotélico dé sentido a un concepto previamente incomprensible, el experimento mental lo ha llevado a una creencia nueva. (Gendler, 1998: 412)

De acuerdo con Gendler, los escenarios imaginarios introducidos por los experimentos mentales proporcionan contextos dentro de los cuales se puede dar sentido a distinciones conceptuales previamente incomprensibles. Esto sucede cuando, como en el caso de Galileo, dos características que se unen constantemente en nuestras representaciones (en este caso, las variables de velocidad y peso) se separan imaginativamente en el escenario. La narrativa del experimento hace que el experimentador se concentre en los detalles y distinga aplicaciones adecuadas e inadecuadas de los conceptos en cuestión. La justificación de las creencias formadas a partir de la ejecución del experimento mental depende de la participación constructiva del experimentador en el escenario imaginario. En este sentido, la visualización del aparato experimental y la modificación de variables en la imaginación desempeñan un papel central en la formación de nuevas creencias. En el caso en cuestión, la realización de una serie de acciones (reconstruir el experimento en la imaginación, imaginar los cuerpos separados en caída libre y el compuesto en caída libre) nos da noticia del tipo de

creencias involucradas y de inconsistencias entre ellas (en este caso, que no podemos imaginar discontinuidades radicales en la caída de los cuerpos).

Para Gendler, la contemplación de un escenario imaginario puede dar lugar a nuevo conocimiento acerca de las características contingentes del mundo natural. El *puzzle* acerca de cómo estos experimentos, que no introducen datos nuevos y que, por lo tanto, no se apoyan en una facultad perceptual, pueden conducir a nuevos conocimientos sobre características contingentes de la realidad, se resuelve mostrando que son capaces de reorganizar el espacio conceptual. En este sentido, ponen a disposición de los experimentadores, creencias que, aunque configuradas empíricamente, no se encontraban directamente disponibles. El experimento mental de Galileo, muestra a los aristotélicos que es posible concebir fenómenos (el comportamiento de un compuesto de dos objetos de distinto peso en caída libre) a los cuales el concepto de velocidad natural aristotélica no permite dar sentido.

La explicación constructivista enfrenta algunos problemas serios de inteligibilidad. El primero y más importante se vincula con la posibilidad de que los experimentos mentales proporcionen contraejemplos a las hipótesis científicas. En pocas palabras, el problema es el de determinar bajo qué condiciones estamos, desde la perspectiva constructivista, en posesión de un enunciado falso si toda experiencia es dependiente de un marco conceptual. En general el constructivismo supone que los datos de la experiencia que se consideran relevantes para una hipótesis están determinados por el marco paradigmático. Los rasgos perceptivos que un paradigma ilumina hasta el punto de mostrar sus regularidades, cambian cuando cambian los compromisos del científico con el paradigma. Por lo tanto, no pueden, desde este punto de vista encontrarse datos brutos de la experiencia, sino “más bien algo recogido con dificultad por el científico hasta que la experiencia inmediata esté determinada por el paradigma” (Kuhn, [1962] (2004: 227). Bajo este supuesto, resulta difícil comprender cómo los

experimentos mentales pueden defraudar las expectativas empíricas a partir de la representación de fenómenos.

Respecto de la noción de velocidad aristotélica, el propio Kuhn señala:

Nada podría salir mal en su ciencia o lógica debido a su aplicación del concepto. En cambio, dada nuestra experiencia más amplia y nuestro aparato conceptual más rico, probablemente diríamos que, consciente o inconscientemente, había incorporado en su concepto de velocidad su expectativa de que solo movimientos uniformes ocurrirían en su mundo. Es decir, llegaríamos a la conclusión de que su concepto funcionó en parte como una ley de la naturaleza, una ley que era regularmente satisfecha en su mundo, pero eso solo sería ocasionalmente satisfecho en el nuestro. (Kuhn, [1964] 1977: 44)

Lo que afirma el pasaje citado es que, dado el papel legislativo de los conceptos, no era posible que los aristotélicos (o Aristóteles mismo) vieran una inadecuación en la aplicación de su concepto de velocidad porque vivían “en un mundo en el que todos los movimientos eran uniformes o cuasi uniformes” (Kuhn, [1964] 1977: 46). Dadas estas consideraciones, sumadas al criterio de acuerdo con el cual en un experimento mental confiable los experimentadores deben poder usar los conceptos en el mismo sentido en que los venían empleando, es difícil ver cómo los experimentos mentales pueden modificar los compromisos conceptuales de los científicos. La hipótesis de los constructivistas es que esta clase de experimentos pone a disposición de los experimentadores información empírica no articulada por el paradigma (Kuhn [1964] 1977: 52). Los experimentos reales proporcionan información sobre el comportamiento de la naturaleza que puede defraudar las expectativas empíricas configuradas por el concepto porque aportan información nueva e inesperada. De esa manera, pueden mostrar la resistencia de la naturaleza para adecuarse a la teoría. Pero, dado que los experimentos mentales no proporcionan datos nuevos, no proceden mostrando la inadecuación entre la experiencia y la teoría. Que puedan hacerlo o no, depende de que los compromisos de los científicos se vean afectados por otros factores que motivan el análisis crítico de las creencias. Esto implica que la

reconceptualización de la experiencia es más bien un producto de la transformación de las condiciones de trabajo bajo un paradigma (por ejemplo, la falta de confianza en el paradigma). En síntesis, en el marco constructivista la experiencia no es independiente del marco conceptual, solo es significativa en el contexto de un paradigma que la articula, por lo que la experiencia (proporcionada por la experimentación ordinaria o recuperada en el experimento mental) no puede constituir evidencia en contra de ninguna hipótesis si el marco conceptual no permite identificarla como relevante.

Existe una dificultad adicional en el alcance de la interpretación constructivista. No todos los experimentos mentales implican una reforma del sistema conceptual del científico. Algunos simplemente ilustran hipótesis o incluso se proponen proporcionar evidencia a favor de una teoría. No podría apelarse en estos casos a la idea de que esta clase de experimentos produce una reorganización conceptual en los términos propuestos, esto es, sosteniendo que el caso presentado por el experimento mental proporciona un contraejemplo a la teoría de fondo. Es decir, el mecanismo no parece integrar la forma en que el conocimiento se amplía en otros tipos de experimentos mentales ni en otras formas ordinarias de conocimiento.

Por último, la perspectiva constructivista parece implicar que los experimentadores poseen un sistema de creencias jerarquizadas y articuladas y que, al mismo tiempo, poseen algunas creencias que no se encuentran disponibles. Estas últimas son activadas por la presentación del escenario descrito en la narrativa de los experimentos mentales. Es difícil ver cómo el ejercicio imaginativo pone a disposición el conocimiento que no se encuentra disponible y con ello manifiesta una contradicción entre el conjunto inicial de creencias y aquellas que permanecieron, hasta ese momento, inaccesibles para el experimentador.

4.4 La razón y la experiencia como fuentes de conocimiento

La discusión acerca del fundamento del conocimiento del mundo natural excede ampliamente la investigación de los poderes epistémicos de los experimentos mentales. No obstante, es una cuestión transversal al problema acerca de su funcionamiento. En principio, una explicación completa de cómo es posible que esta clase de experimentos proporcione conocimiento nuevo descansa sobre una serie de compromisos metafísicos y epistémicos que es preciso justificar. Dentro de las perspectivas existentes, las que suscriben una interpretación *apriorista* han aportado argumentos para justificar la asunción de estos compromisos. Las perspectivas que abogan por el empirismo, considerado como la epistemología aceptada por *default* en la filosofía de la ciencia contemporánea, en general no expusieron argumentos a favor de esta postura.

Aunque existen numerosas objeciones al apriorismo, y en particular al platonismo con el que la posición de Brown se asocia, y aunque la adopción de una perspectiva empirista ha permanecido fuera de la discusión (se ha problematizado la interpretación argumental de los experimentos mentales pero no su fundamento empirista), esta última no está libre de problemas. Aun suponiendo que el empirismo es, *prima facie*, la mejor alternativa para explicar el origen de nuestras creencias acerca del mundo físico, es una característica de los experimentos mentales no presentar datos nuevos. Por lo que si se asume, como lo hace Norton, que el mecanismo de generación de nuevas creencias es inferencial, entonces, no es inconsistente rechazar en empirismo de fondo y sostener que los experimentos mentales son argumentos disfrazados.

Por otra parte, el hecho de que los experimentos mentales puedan ser reconstruidos como argumentos, o que las mismas conclusiones puedan ser alcanzadas a través de un argumento, no es una razón suficiente para creer que el mecanismo de generación de nuevas creencias en los experimentos mentales es puramente argumentativo. La presunción de Norton es que, dado que la

reconstrucción argumental es posible, los experimentos mentales son argumentos encubiertos; sin embargo, el mero hecho de que podamos alcanzar los mismos resultados por otros medios no nos revela la naturaleza esencial de la operación. Al respecto, Brown ha ofrecido el siguiente argumento:

Por ejemplo, la mayoría de nosotros diría que hacemos juicios sobre el tamaño relativo de otras personas en función de nuestra percepción de sus características geométricas. Supongamos que Norton afirma que esto no es así y que lo que realmente hacemos es contar moléculas. Luego, muestra que cada vez que juzgamos que *A* es más grande que *B*, lo hacemos a partir de una reconstrucción (laboriosa): *A* tiene más moléculas que *B*. Incluso una tasa de éxito del 100% en la reconstrucción de juicios de tamaño no refuta la afirmación del sentido común de acuerdo a la cual juzgamos el tamaño mediante la geometría visual, no la aritmética. (Brown, 2002: 1134)

En síntesis, la perspectiva de Norton no proporciona argumentos positivos para adoptar el empirismo como epistemología de base y tampoco ofrece razones para aceptar la tesis de la reconstrucción (de acuerdo con la cual los experimentos mentales pueden ser reconstruidos como argumentos y por ello puede considerarse que son argumentos encubiertos). Aun suponiendo que no hay ninguna forma de conocimiento exótico subyacente en los experimentos mentales, es necesario proporcionar alguna razón para aceptar la tesis empirista como explicativa del enigma del nuevo conocimiento en esta clase de experimentos. Asimismo, es preciso mostrar que, como sostiene Norton, la perspectiva argumental es el único modelo posible para explicar el mecanismo de generación de creencias en los experimentos mentales, es decir, mostrar que al prescindir de la intervención en el mundo natural la única manera de obtener nuevo conocimiento sobre este es razonando argumentativamente.

Las posiciones aprioristas, aunque desarrollan argumentos para justificar la aceptación de los compromisos epistémicos y metafísicos asociados, no logran modelar, mediante el paradigma del conocimiento matemático, el tipo de conocimiento del mundo natural que tiene lugar en los experimentos mentales. En particular, Brown no explica suficientemente en qué consiste el conocimiento a

priori del mundo natural ni determina claramente la relación que supone entre la visualización de fenómenos *a priori* y la captación intuitiva de leyes de la naturaleza. Se limita a sostener que existe una analogía entre el conocimiento de verdades matemáticas y el conocimiento intuitivo de leyes naturales. Esta estrategia, lejos de aclarar en qué consiste este tipo de conocimiento, admite que la percepción de entidades platónicas es opaca. Si esta analogía se sostiene, los criterios para resolver los desacuerdos perceptivos acerca de los fenómenos físicos también deberían ser aplicables a la resolución de desacuerdos entre intuiciones intelectuales discordantes, o, en todo caso, la teoría de Brown debería especificar criterios análogos de corrección para la intuición intelectual. Si un axioma de una teoría matemática resulta evidentemente verdadero para un sujeto y evidentemente falso, o al menos muy dudoso, para otro, no se advierte de qué manera podría superarse el desacuerdo apelando únicamente a la intuición. El conflicto entre intuiciones constituye, así, una poderosa razón para dudar de la confiabilidad de la intuición intelectual. En general, la falta de inteligibilidad de la captación de entidades platónicas se ha considerado una razón suficiente para descartarla como vía de conocimiento.

Las perspectivas aprioristas se presentan como las rivales más serias del empirismo. Norton (2004: 56) ha sostenido que esto es así dado que el apriorismo, en particular el que defiende Brown, es la única posición irreductible a la interpretación argumental porque es la única alternativa que postula que esta clase de prácticas posee, en ocasiones, poderes epistémicos extraordinarios.

Una manera alternativa de explicar cómo obtenemos conocimiento nuevo acerca del mundo en los experimentos mentales, sin postular capacidades epistémicas extraordinarias ni entidades extravagantes como objeto de conocimiento, es apelando a la estructura conceptual de nuestras creencias. Tal postura sostiene que, en algunos casos, esta clase de experimentos muestra a partir de la presentación de un caso, que el empleo de conceptos es inadecuado o que un concepto no se subsume bajo otro (desafiando un supuesto teórico). La

principal dificultad de esta perspectiva reside en esclarecer de qué manera la inspección de una situación imaginaria puede presentar un caso que desafíe las expectativas empíricas, si estas son configuradas por nuestro sistema conceptual. Desde este punto de vista, no es posible argumentar que estos experimentos sean capaces de proporcionar apoyo evidencial a hipótesis científicas, dado que el mecanismo invocado, en caso de poder justificarse, solo explica el funcionamiento de los experimentos mentales con funciones refutatorias.

En síntesis, todas las perspectivas disponibles fallan individualmente en satisfacer alguno de los criterios generales de evaluación y, comparativamente, no presentan ventajas explicativas sustantivas respecto de sus rivales. La perspectiva empirista, en principio el mejor candidato para los propósitos de determinar los poderes justificatorios de los experimentos mentales, resulta insatisfactoria en cualquiera de sus versiones : en la versión de Norton es demasiado parsimoniosa (no aporta razones suficientes para adoptar el empirismo ni el experimentalismo); en la versión de Mach se funda en un subjetivismo que hace difícil integrar la explicación a mecanismos generales de razonamiento científico; en la versión de Sorensen es una explicación oscura en cuanto a que el mecanismo psicológico de generación de creencias se supone en cierto sentido inefable y solo expresable en la reconstrucción lógica; en la versión de Clatterbuck y Brendel depende de la noción de intuición empírica, por lo que se compromete con una noción filosóficamente oscura. El apriorismo, el candidato menos deseable dentro del marco filosófico adoptado en este trabajo, presenta serias dificultades para la aceptación, la mayoría de ellas dependientes de la asunción de compromisos metafísicos superfluos y de la oscuridad de las explicaciones proporcionadas. Por último, las perspectivas constructivistas no parecen útiles a los propósitos de esta investigación, fundamentalmente porque no permiten integrar el mecanismo de obtención de nuevo conocimiento a otras prácticas de razonamiento científico.

4.6 Conclusiones

La evaluación de las explicaciones disponibles muestra que en general resultan insatisfactorias por la falta de claridad acerca de los vínculos que se establecen entre el fundamento del conocimiento invocado y el mecanismo a partir del cual se obtienen nuevas creencias luego de la ejecución de un experimento mental. En los casos en los que se adoptan compromisos muy modestos, como sucede en las posiciones empiristas, la apelación a mecanismos de inferencia lógica no parece corresponderse con el procedimiento que de hecho se lleva a cabo en los experimentos mentales para obtener conocimiento nuevo sobre el mundo natural. Las perspectivas aprioristas que adoptan compromisos epistémicos y metafísicos extravagantes, no pueden explicar con claridad el funcionamiento de las capacidades invocadas para proporcionar conocimiento sobre el mundo empírico. La perspectiva constructivista, por su parte, no consigue explicar la aplicabilidad de las conclusiones a casos no contemplados por la teoría ya que sujeta la significatividad de la experiencia a los esquemas conceptuales. Como conclusión de este examen, se esboza a continuación un enfoque que se propone superar las limitaciones de las propuestas explicativas existentes. Esta perspectiva postula que el empirismo es la posición más plausible para explicar el origen del conocimiento sobre el mundo natural, explicita el mecanismo de obtención de nuevo conocimiento apelando a la imaginación y muestra cómo esta facultad se relaciona con otros procesos cognitivos y específicamente, con el razonamiento científico.

Debido a que los usos de los experimentos mentales atestiguan que sus resultados proporcionan algún tipo de conocimiento del mundo natural, es razonable suponer que dicho conocimiento tiene un fundamento en la experiencia y que es en virtud de este fundamento que pueden proporcionar evidencia para contrastar hipótesis o teorías físicas. Si los experimentos mentales nos permiten extraer conclusiones acerca del comportamiento de los fenómenos físicos,

entonces, dichas conclusiones son proposiciones de carácter sintético y el acceso a ellas, *a posteriori*. Si suponemos que la ejecución de (algunos) experimentos mentales amplía el conocimiento sobre la naturaleza, podemos conjeturar que dicho conocimiento, dado que no se presentan datos nuevos, es el producto de un proceso de ampliación o corrección del conocimiento previamente disponible. Para explicar de qué manera dicha ampliación es posible, es razonable suponer que este proceso tiene lugar con la concurrencia de otras facultades, además de la percepción que permite acceder al conocimiento del mundo sensible. La imaginación es una facultad que permite explicar el funcionamiento de las inferencias ampliativas y que tiene un papel importante en el razonamiento científico. Aunque no es posible definirla de manera unívoca, podría ser caracterizada brevemente como la capacidad de formular o suponer posibilidades (Williamson, 2016: 5). En su aplicación al método científico, es posible entender que desempeña un papel en los casos en los que el método de ensayo y error no puede llevarse a cabo. Este ejercicio no es completamente libre, se encuentra limitado y orientado por sus objetivos y por el conjunto de supuestos teóricos y empíricos que condicionan su empleo. Si bien puede resultar problemático suponer que es posible aprender algo relevante a partir de un ejercicio exclusivamente imaginativo, es plausible argumentar que este proceso se da y que pueden establecerse condiciones de adecuación para detectar casos espurios. Dadas ciertas condiciones iniciales empíricas y teóricamente orientadas, el desarrollo posible del escenario imaginado está condicionado por estas de modo que sería posible distinguir cuándo las consecuencias que se derivan de esta situación inicial son ilegítimas. En este sentido, aunque se trata de una facultad falible, tanto como la percepción y la memoria, el procedimiento imaginativo (bajo las condiciones previamente estipuladas) constituye un método confiable para formar creencias verdaderas acerca de lo que sucedería en circunstancias hipotéticas (véase Williamson, 2016: 8). En el caso de los experimentos mentales lo que compele a la aceptación de ciertas conclusiones y al rechazo de otras es que la decisión final acerca de 'qué sucedería' no depende de que el

experimentador imagine libre y voluntariamente, sino de lo que, dadas las condiciones antecedentes y el marco teórico asumido sucedería en tales y cuales escenarios. De manera similar a como podemos ganar conocimiento acerca de las relaciones espaciales sin intervenir directamente sobre los objetos, la imaginación permite obtener conocimiento sobre los fenómenos físicos, en condiciones especiales, cuando está orientada por supuestos teóricos bien definidos y suficientemente informada por la experiencia previa.

CAPÍTULO 5

Experimentos mentales, representación y modelos científicos

5.1 Introducción

Las perspectivas que interpretan a los experimentos mentales como condiciones de posibilidad de los experimentos ordinarios (Sorensen, 1992) o como límites ideales de los experimentos reales (Laymon, 1991) son algunas de las líneas que permiten conectar el análisis de estos experimentos con la epistemología de los modelos. Desde estos puntos de vista, los experimentos mentales describen situaciones idealizadas que proporcionan información acerca de cómo las leyes científicas gobiernan el mundo físico. De manera similar, los modelos científicos integran fenómenos a contextos teóricos centrándose en algunos aspectos de los sistemas que representan. Humphreys (1993: 220) por ejemplo, interpreta a los experimentos mentales como exploraciones y refinamientos de los modelos teóricos.

En la ciencia moderna la construcción de modelos es fuente de conocimiento del mundo físico. Los modelos científicos son de gran importancia especialmente para comprender sistemas que resultan inaccesibles a la experimentación real como en el caso de los objetos de estudio de la astrofísica o la cosmología. La relevancia de los modelos y las simulaciones computacionales en los procesos de experimentación ha llevado a cuestionar el rol de la materialidad en el estatus epistémico de los resultados experimentales (Morrison, 2009). La ausencia de intervención en el mundo material es un rasgo que las simulaciones y modelos computacionales comparten con los

experimentos mentales, por lo que explorar las estrategias empleadas para proveer una justificación de las creencias en los resultados de las simulaciones puede resultar de utilidad para proporcionar argumentos análogos para explicar cómo los experimentos mentales son capaces de justificar sus conclusiones.

La noción de modelo científico ha sido ampliamente tematizada en la literatura filosófica de los últimos años (véase por ejemplo Weisberg, 2013, Morrison, 2015, Gelfert, 2016 y Downes, 2021). Entre otras cuestiones, se ha discutido la naturaleza de los modelos, las funciones que desempeñan y el estatus del conocimiento que pueden proporcionar. Aunque no existe un consenso generalizado sobre la extensión del concepto de modelo científico, se han realizado distinciones importantes entre tipos de modelos y marcos teóricos para el abordaje de la cuestión. Interesa para los propósitos de esta investigación, retomar la cuestión acerca de cómo los modelos vinculan principios teóricos con un dominio de fenómenos. La elucidación de este vínculo estuvo ligada, en la discusión filosófica reciente, al desarrollo de una noción inteligible de representación científica. En términos generales, se asume que los modelos científicos representan algún aspecto del mundo o conjunto de fenómenos, usualmente denominado “objetivos” (*targets*) de la representación, a través de ciertos objetos abstractos o concretos identificados como la “fuente” (*source*) de la representación. La noción de representación científica es objeto de una amplia discusión y existen pocos acuerdos entre los filósofos de la ciencia acerca de qué constituye una representación adecuada de un fenómeno (véase Frigg & Nguyen, 2020). En este examen nos apoyaremos en una noción de representación científica centrada en el uso de modelos en la práctica científica. Esta perspectiva permite dar sentido al hecho de que los experimentos mentales sean capaces de proporcionar conocimiento nuevo del mundo físico sin incorporar datos nuevos. Retomaremos la noción inferencial de representación, de acuerdo con la cual los modelos científicos posibilitan una forma de razonamiento sustitutivo (*surrogate reasoning*). Desde este enfoque, los experimentos mentales pueden ser

interpretados como una clase de modelos que posibilita la obtención de nuevo conocimiento del mundo físico por medio de la realización de inferencias sobre el objetivo de la representación a través de la fuente (el modelo).

Algo que experimentos mentales y modelos tienen en común es que proporcionan acceso a determinados fenómenos que en muchos casos son de una complejidad tal que no puede ser descrita o controlada sin operaciones previas de simplificación y abstracción. Las abstracciones, simplificaciones e idealizaciones se utilizan tanto en las descripciones de experimentos reales, como en la construcción de modelos computacionales y en los experimentos mentales. La eficacia heurística de los estos últimos se explica, desde algunas perspectivas, por el empleo de estas estrategias. A la atención a ciertos elementos relevantes del sistema representado por la narrativa del experimento y la abstracción de los detalles que no son sustanciales para la representación del fenómeno en cuestión, se asocia la celeridad con la que los experimentadores derivan conclusiones en los experimentos mentales. Asimismo, las distorsiones introducidas en las situaciones representadas por los experimentos mentales facilitan el tratamiento de los problemas complejos.

En síntesis, la construcción de modelos de los fenómenos que permiten aplicar principios teóricos a sistemas concretos del mundo real es una actividad que ha sido diferenciada de la teorización y de la experimentación. La exégesis de los poderes epistémicos de los experimentos mentales aparece íntimamente vinculada a sus propiedades de representación y modelización. Interpretar a los experimentos mentales como una especie de modelos permitiría explicar sus rasgos más importantes, sustrayéndolos de la analogía con experimentación real y con las inferencias lógicas a partir de las cuales han sido explicados.

Se argumenta en este capítulo que los experimentos mentales dependen de un tipo de actividad de representación que es común a la actividad de modelización y que permite vincular el conocimiento teórico con el conocimiento

empírico disponible. Se muestra, además, que los experimentos mentales pueden concebirse como una clase de modelos científicos. Finalmente, se arguye que la comparación entre experimentos mentales, modelos mentales y simulaciones computacionales proporciona una clave para explicar un aspecto no tematizado explícitamente de su funcionamiento: el carácter dinámico del modelo y su papel en la comprensión de la evolución temporal de un sistema físico.

5.2 Sobre la noción de modelo científico

El propósito de esta sección es presentar la discusión filosófica acerca de los modelos científicos y mostrar de qué manera se relaciona con el problema epistemológico de los experimentos mentales.

Si bien los puntos de coincidencia entre los filósofos abocados al estudio de los modelos científicos son pocos, una de las características que tuvo en los últimos años el examen de esta noción es que se centró en la práctica científica, puntualmente en la construcción y aplicación de modelos. El estudio de casos ha funcionado como motivación del análisis filosófico y como elemento de evaluación de las conceptualizaciones de los modelos, como lo ejemplifica el análisis que realiza Bailer-Jones (2003: 59) del modelo de la estructura estelar con el objetivo de examinar la capacidad representativa de los modelos.

Por otra parte, algunos filósofos que adoptaron esta perspectiva de análisis les han atribuido a los modelos un papel mediador entre la teoría y los fenómenos. Particularmente los trabajos de Mary Morgan y Margaret Morrison (1999) y Nancy Cartwright (1999) defienden esta interpretación. Desde este punto de vista, los modelos son entidades híbridas en cuya construcción intervienen diversos elementos: hipótesis pertenecientes a diferentes teorías, datos empíricos y estrategias de modelización. Este enfoque les atribuye a los modelos poderes epistémicos que exceden el papel heurístico al que tradicionalmente habían sido

relegados. Los modelos como mediadores son capaces de representar aspectos del mundo físico, ejemplificando los conceptos abstractos que aparecen en las teorías. Son al mismo tiempo herramientas imprescindibles de la práctica científica porque posibilitan una amplificación de los recursos disponibles para obtener conocimiento del mundo natural. Algunas clases de modelos pueden funcionar como instrumentos de medición, aumentando nuestra capacidad de observación de los fenómenos. En algunos casos, los modelos pueden incluso facilitar la intervención sobre fenómenos en principio inaccesibles (Morgan y Morrison, 1999: 19).

Los experimentos mentales han sido asociados a modelos fenoménicos idealizados (Laymon, 1991) y desde algunas perspectivas sus funciones se han vinculado a la articulación conceptual y a la ilustración de hipótesis teóricas.²⁹ Recientemente, se ha argumentado que los experimentos mentales pueden ser considerados como precursores de las simulaciones computacionales o de los experimentos virtuales (véanse Chandrasekaran y Neressian, 2012 y Lenhard, 2018).

La noción general de modelo científico que comparten las teorizaciones antes mencionadas es fructífera para elucidar las características de los experimentos mentales. Los experimentos mentales pueden ser interpretados como una clase de modelos científicos sobre la base de tres aspectos fundamentales que comparten con estos. En primer lugar, hacen a los sistemas físicos accesibles porque pueden *suplantarlos*. De este modo, posibilitan una forma de razonamientos *sustitutivo* que permite a los experimentadores derivar conclusiones acerca del mundo natural a partir de la exploración del comportamiento del modelo. En segundo lugar, *hacen a problemas complejos tratables* a través del empleo de estrategias de idealización que permiten la simplificación por la concentración en algunos aspectos, la abstracción y la

²⁹ Como sostienen por ejemplo Häggqvist (1996) y Sorensen (1992).

supresión de información no relevante. Por último, cumplen un importante papel en la *aplicación de teorías*, como mediadores entre los fenómenos y los principios teóricos.

El primero de los aspectos mencionados conduce directamente a uno de los problemas centrales de la epistemología de los modelos científicos. Si bien no existe un consenso sobre la definición de modelo científico, se suelen caracterizar de manera muy general como entidades que *representan de manera idealizada* un fenómeno o dominio de fenómenos. Que los modelos puedan suplantar un sistema físico depende en principio de su capacidad de representar (adecuadamente) algunos aspectos de éste. Por consiguiente, es necesario proporcionar alguna noción de representación que permita explicar en qué sentido los experimentos mentales representan fenómenos, de manera análoga a como los modelos científicos lo hacen.

Un primer elemento para abogar a favor de la semejanza entre experimentos mentales y modelos consiste en sostener que tanto los experimentos mentales como los modelos científicos representan porque permiten derivar conclusiones acerca del comportamiento del mundo natural. Algunas versiones de esta interpretación no se comprometen con la idea de que los modelos son reducibles a los enunciados empleados para expresar las proposiciones acerca de los fenómenos implicadas por el modelo. Esta perspectiva resulta de utilidad para defender una posición no reduccionista de los experimentos mentales. Ya se ha mostrado que la perspectiva que reduce sus poderes epistémicos a los de las inferencias lógicas y la verdad de los enunciados que expresan proposiciones sobre el mundo natural tiene limitaciones para dar cuenta de varios aspectos del funcionamiento de esta clase de experimentos. La celeridad con que se arriba a ciertas conclusiones parece no coincidir con la realización de procesos inferenciales. Asimismo, algunos experimentos mentales han cumplido un papel en el apoyo a principios teóricos que no parece corresponderse con el de los argumentos teóricos.

La perspectiva defendida por Bailer-Jones (2003) combina una concepción realista de la noción de representación en los modelos, es decir, defiende la idea de que los modelos son acerca del mundo, y una perspectiva no reduccionista de su funcionamiento, es decir, sostiene que los modelos no son reducibles a las proposiciones por ellos implicadas. Según su punto de vista, “un modelo científico es una descripción interpretativa de un fenómeno (proceso u objeto) que facilita el acceso intelectual o perceptual a ese fenómeno” (Bailer-Jones, 2003: 61). El tipo de descripción que los modelos son capaces de proporcionar excede la mera descripción fenomenológica, explora un trasfondo teórico relevante y emplea una gama de herramientas representacionales externas como ecuaciones, parcelas de datos empíricos, esquemas, objetos y enunciados del lenguaje ordinario, entre otros recursos. Los modelos usan estas herramientas e implican proposiciones.³⁰ Por lo que la cuestión acerca de cómo los modelos representan no puede centrarse solo en la explicación de esta relación en términos de proposiciones.

La noción de representación que para Bailer-Jones permite dar una interpretación de los modelos concierne a la práctica empírica y no requiere (sustancialmente) del concepto formal de modelo. En esta clave, el vínculo entre fenómenos y modelos se concibe de una manera particular. Por una parte, dado que los modelos no son formulados para describir todos los aspectos del fenómeno al cual son aplicados, no es posible determinar de antemano cuántas proposiciones (verdaderas o falsas) son implicadas por el modelo. Por otra parte, dado que los modelos involucran idealizaciones y abstracciones, un cierto número de proposiciones falsas pueden seguirse del modelo. Cuántas proposiciones falsas se deben admitir y cuáles proposiciones son verdaderas respecto del fenómeno original son cuestiones que deben ser decididas por razones pragmáticas. Asimismo, la función de los modelos y la selección de los aspectos del fenómeno modelados por éstos son cuestiones que conciernen a la práctica

³⁰ “Implicar” para Bailer-Jones (2003: 61) no tiene el sentido de implicación lógica sino el sentido de acuerdo con el cual los modelos contienen proposiciones o pueden ser transcritos en proposiciones que expresan su mensaje.

científica; puntualmente, competen a los usuarios del modelo. Por otra parte, los modelos son evaluados en relación con los datos para los cuales proporcionan una explicación y pueden ser actualizados a partir de una revisión crítica basada en nueva información relevante. En síntesis, desde esta perspectiva, el análisis epistemológico de los modelos científicos no puede reducirse a la interpretación de los mismos como proposiciones, incluye aspectos pragmáticos y otorga un papel importante a los agentes o usuarios de los modelos. Además, se sirve de una noción minimal de representación, a saber: el modelo representa cuando implica proposiciones que son verdaderas acerca del fenómeno.

La noción de modelo científico adoptada por Bailer-Jones permite dar sentido a varios rasgos de los experimentos mentales. En primer lugar, los experimentos mentales emplean recursos para representar fenómenos que no se limitan a la capacidad expresiva de las proposiciones lingüísticas. Muchos experimentos mentales, como por ejemplo, el experimento del Plano inclinado de Stevin, incluyen diagramas que facilitan la visualización del sistema descrito en la narrativa. En segundo lugar, como ha mostrado la perspectiva argumental defendida principalmente por Norton, muchos experimentos mentales pueden ser reconstruidos como argumentos a partir de la explicitación (o formulación de las premisas) y de la identificación de las inferencias involucradas. No obstante, este proceso no es unívoco y en algunos casos la reconstrucción de los argumentos varía, fundamentalmente respecto de la formulación de las premisas involucradas. El experimento de los cuerpos en caída libre de Galileo es el que ha dado lugar a un mayor número de reconstrucciones. Véanse, por ejemplo, las diferencias entre las reconstrucciones que realizan Norton (1996: 340-345), Brendel (2005: 93-94), Picha (2011: 164), Gendler (2012: 403-406). O las diferentes interpretaciones de las premisas en el experimento de Stevin por parte de Mach ([1893]1960: 24-32), Norton (1996: 349-351) y Brown (2011: 3-5). Este hecho podría ser tomado como una razón para no reducir los experimentos mentales a las proposiciones en las que pueden ser traducidos o a las proposiciones en las

que pueden expresarse sus resultados o conclusiones. Por otro lado, todos los experimentos mentales se centran en algunos aspectos del fenómeno a representar. En las narrativas de esta clase de experimentos solo se describen ciertos aspectos del fenómeno, en el mencionado caso de Galileo, por ejemplo, se describen la forma y composición de los cuerpos, pero no su color o su sabor.

Por su parte, Mauricio Suarez (2004) ha defendido una concepción deflacionaria de la representación que permite articular la interpretación de los experimentos mentales como modelos científicos.³¹ De acuerdo con Suarez, los modelos no representan a sus objetivos en virtud de su semejanza. Su perspectiva puede ser sintetizada del siguiente modo: un modelo (*A*) representa un sistema (*B*) si (y solo si) la fuerza representacional de *A* señala a *B* y *A* permite a agentes competentes e informados derivar inferencias acerca de *B*. Esta noción de representación involucra dos condiciones. La primera estipula que el modelo es usado por alguien (un usuario) para representar un sistema. Suarez denomina a esta condición “fuerza representativa”. La segunda condición establece que el modelo permite al usuario realizar inferencias del modelo al sistema. El autor llama a esta condición “capacidad inferencial”. Para Suarez, no hay aspectos en la noción de representación que permitan proporcionar una definición en términos de condiciones necesarias y suficientes. Argumenta que tal definición no puede proporcionarse en términos de semejanza e isomorfismo, como tradicionalmente intentó elucidarse la relación lógica entre la fuente y el objetivo de la representación. Por lo que, la mejor respuesta que es posible proporcionar a la pregunta: “¿en virtud de qué los modelos representan sistemas?” es que los modelos representan porque los agentes son capaces de extraer conclusiones

³¹ De acuerdo con Suarez (2014: 38), el enfoque deflacionario se caracteriza por no proveer una definición sustancial de representación, esto es, en términos de condiciones necesarias y suficientes. El enfoque deflacionario sostiene que las funciones de la representación no pueden ser definidas *a priori*, independientemente de la práctica representacional sino que solo pueden ser completamente articulados contra el trasfondo de esta práctica. Por lo que su perspectiva solo se propone proveer algunas condiciones necesarias para la representación.

sobre los sistemas a partir de una forma de razonamiento sustitutivo, esto es, derivando conclusiones acerca del comportamiento de los fenómenos a partir del modelo. Dichas inferencias pueden ser deductivas o inductivas y no necesariamente son enunciados verdaderos acerca del target.³² Esta noción de representación logra establecer la direccionalidad de esta relación pero solo proporciona un criterio para distinguir representaciones epistémicas de representaciones por mera estipulación o denotación arbitraria. La cuestión que queda por decidir es la de cuándo las inferencias sobre los objetivos son verdaderas. La noción deflacionaria de representación no permite establecer en qué grado el modelo es una representación adecuada o inadecuada de su objetivo. Para Suarez, la representación es una actividad colectiva y son los agentes y usuarios de los modelos los que a partir de sus intereses y propósitos y de un trasfondo de conocimiento establecen los criterios de corrección para los modelos.

La perspectiva deflacionaria de Suarez permite explicar el tipo de razonamiento sustitutivo que los experimentos mentales hacen posible. En principio, el escenario descrito en el experimento mental tiene por objetivo representar un sistema físico (o un dominio de fenómenos) y permitir a los experimentadores realizar inferencias sobre ese sistema físico o dominio de fenómenos. Así, por ejemplo, Galileo en el *Diálogo sobre los dos máximos sistemas del mundo* ([1632] 1967: 186-188), propone una situación para observar el comportamiento de los objetos en un marco de referencia inercial. El experimento describe a dos observadores situados dentro de la cabina de un barco liberando animales voladores o arrojándose mutuamente una pelota. El objetivo del experimento es mostrar que no es posible para un observador dentro de un marco de referencia en movimiento uniforme diferenciar entre éste y uno en reposo. A partir del comportamiento del modelo (en este caso de los objetos en el escenario

³² Para un examen más detallado de la noción de representación deflacionaria de Suárez y su comparación con otras perspectivas de la representación ver Cassini y Redmond (2021: 28-32).

descrito) es posible realizar inferencias válidas sobre el comportamiento del sistema objetivo. La conclusión se aplica a sistemas físicos reales, por ejemplo, a la Tierra en rotación, considerada como un sistema inercial. En principio, podría decirse que satisface las dos condiciones establecidas por Suarez para que un modelo represente: el experimento posee fuerza representativa y capacidad inferencial.

Gabriele Contessa (2007), por su parte, revisa la noción de razonamiento sustitutivo y vuelve sobre la propuesta inicial de Chris Swoyer (1991) recuperada por Suarez (2004). El examen de la propuesta de Contessa permite atender a algunos puntos problemáticos en la perspectiva deflacionaria de la representación que empleamos como marco para interpretar a los experimentos mentales como modelos. Especialmente, permite elaborar algunos criterios de adecuación para la representación. Contessa señala que Suarez no proporciona elementos para esclarecer la naturaleza del vínculo de representación. Su propuesta es que la habilidad de los usuarios de realizar inferencias desde el modelo al objetivo puede ser explicada por el hecho de que los usuarios interpretan el modelo en términos del sistema objetivo (pero no viceversa). Para argumentar a favor de esta perspectiva, Contessa distingue la noción de representación epistémica de otras dos que son cercanas: la noción de denotación y a la de representación epistémica fiel (*faithful*). En relación al primer concepto, señala que algo puede denotar otra cosa sin que exista entre estos objetos una relación de representación. Así, por ejemplo, el logo del Metro de Londres denota al sistema de subterráneos de Londres pero no lo representa. Una representación epistémica, por su parte, permite realizar inferencias correctas (*sound*) aunque no necesariamente verdaderas para un sistema a partir de un modelo. Por ejemplo, un mapa antiguo del subterráneo de Londres es una representación epistémica de éste aunque algunas de las inferencias que se pueden realizar a partir de este mapa no son verdaderas respecto del sistema de metro actual.

Dadas estas distinciones, Contessa entiende que el razonamiento sustitutivo es una condición suficiente para la representación epistémica pero no para una representación epistémica fiel. Para que un modelo represente fielmente un objetivo, las inferencias que es posible realizar desde el vehículo al objetivo deben ser verdaderas además de ser correctas. Aquí Contessa (2006: 54) diferencia grados en la fidelidad de la representación: una representación es completamente fiel si todas las inferencias (válidas) del modelo al sistema son verdaderas, es completamente infiel si ninguna de las inferencias es verdadera, y representa inadecuadamente (algunos aspectos del sistema objetivo) si algunas inferencias son verdaderas y otras falsas. Los usuarios de un modelo adoptan una representación interpretativa de este si toman al modelo como un todo, es decir si consideran a los componentes del modelo y a algunas de sus propiedades y relaciones como estando en lugar de los componentes, propiedades y relaciones del sistema en cuestión. Esta noción de representación interpretativa permite, según Contessa, explicar en virtud de qué algunas inferencias son legítimas y otras no. La legitimidad está dada por las relaciones de denotación que se establecen entre los elementos del modelo y los del sistema representado, es decir, están determinadas por el vínculo que se establece entre los elementos, propiedades y relaciones del modelo que están en lugar de los elementos, propiedades y relaciones del sistema.

La noción de representación interpretativa es, desde este punto de vista, más fundamental que la de razonamiento subrogativo. Es en virtud de que los usuarios interpretan el modelo en términos del objetivo que pueden realizar inferencias del modelo al objetivo. Si la interpretación no fuera una condición suficiente para la representación, habría casos en los cuales los usuarios adoptarían una interpretación en términos del sistema y sin embargo el modelo fallara en representar el sistema, algo que Contessa considera que no es posible que ocurra. Puede haber representaciones inadecuadas que, no obstante, son representaciones dada la condición de representación epistémica.

Contessa (2007) agrega algunos elementos importantes para explicar cómo los experimentos mentales son capaces de representar fenómenos físicos. Dada la dificultad de la cuestión, no sería posible examinar suficientemente el problema de si es posible proporcionar una noción sustantiva de representación. No obstante, puede afirmarse que la noción de interpretación propuesta por Contessa resulta esclarecedora de la relación de representación para el caso de los experimentos mentales. Una de las dificultades más importantes de las concepciones representacionistas de los modelos es la de proporcionar criterios para identificar representaciones incorrectas o inadecuadas. La noción de interpretación analítica propuesta por Contessa proporciona un criterio para identificar inferencias válidas desde el modelo al fenómeno. Presentada de manera simplificada, sostiene que este vínculo depende de que los usuarios tomen hechos acerca del vehículo como estando en lugar de hechos acerca del *target*. Si bien por sí sola esta condición es incapaz de proveer un criterio para distinguir las conclusiones verdaderas acerca de fenómeno de las conclusiones falsas, es un avance sobre este problema en tanto nos permite identificar cuándo un modelo es una representación epistémica de un fenómeno.

Los escenarios descritos en los experimentos mentales habilitan ciertas inferencias y bloquean otras en virtud de que los usuarios los emplean interpretando ciertos objetos propiedades y relaciones del modelo como estando en lugar de ciertos objetos, propiedades y relaciones que posee un determinado fenómeno. Así, por ejemplo, Newton ([1728] 1934: 512-513) argumenta que la fuerza gravitacional es la misma que rige el comportamiento de los objetos en las cercanías de la Tierra y de la Luna, comparando el movimiento de un proyectil con el movimiento de este satélite. El experimento mental describe una bala de cañón arrojada desde lo alto de una montaña con una velocidad inicial que se incrementa paulatinamente hasta que es lanzada con una velocidad inicial suficiente para que represe al punto de inicio. No todos los elementos del sistema objetivo puedan encontrarse en el modelo. El experimento mental supone por ejemplo, que la Tierra

no rota y que no hay fricción entre el proyectil y el aire. La conclusión de acuerdo con la cual cuando la velocidad inicial sobrepasa un valor crítico el proyectil volverá al punto inicial desde la cual se disparó, representa (y en este caso explica) el comportamiento del fenómeno. El usuario del modelo asume que el comportamiento de la bala de cañón puede representar el comportamiento de la Luna, y que la gravedad ejerce una fuerza atractiva sobre ambos objetos.

El segundo elemento para argumentar que los experimentos mentales son un tipo de modelos, es que permiten que ciertos sistemas físicos muy remotos o muy pequeños sean accesibles y que algunos problemas muy complejos sean tratables. Esta es una función comúnmente aceptada de los modelos. Los experimentos mentales emplean estrategias de idealización, fundamentalmente la abstracción y la supresión de ciertos rasgos de los fenómenos, para simplificar o hacer posible la representación.

La cuestión acerca de qué clase de estrategias de modelización utilizan los experimentos mentales se tratará en el siguiente apartado. Una interpretación posible acerca de cómo los experimentos mentales representan es suponer que el tipo de idealizaciones utilizadas en estos contextos posibilitan procesos de desidealización que pueden volver a los modelos empleados en los experimentos mentales crecientemente verdaderos. McMullin (1985) sostiene que este tipo de estrategias de idealización es comúnmente empleada en los modelos científicos. Desde esta perspectiva, es posible acercarse de forma cada vez más precisa al comportamiento del modelo al comportamiento del sistema específico representado por éste.

Por otro lado, una característica de los experimentos mentales es representar selectivamente ciertos aspectos de los fenómenos. Si bien hacen posible inferir conclusiones acerca de fenómenos, es un hecho que no todos los elementos presentes en los sistemas reales están presentes en la descripción del escenario imaginario y que ciertos elementos descritos en el experimento mental

no representan nada. Una interpretación posible de cómo los experimentos mentales representan ciertos fenómenos es que lo hacen porque son modelos que tienen un dominio no vacío de aplicación. Weisberg (2013: 90), por ejemplo, afirma que el objetivo de los modelos no son los fenómenos en sí mismos sino abstracciones de los fenómenos. Como consecuencia, el mismo fenómeno puede originar muchos “objetivos” diferentes dependiendo del subconjunto de propiedades seleccionadas. La cuestión abierta en este punto de vista es en base a qué criterios los científicos realizan esta selección.

Cassini (2018) propone que la construcción de un modelo comienza con un problema, prosigue con la determinación del propósito del modelo y la especificación de qué tipo de soluciones se espera de su empleo y en qué grado el modelo se aproxima a los fenómenos reales. Una vez que el problema es definido es posible identificar el objetivo del modelo destinado a resolver el problema específico (Cassini, 2018: 193). Esta perspectiva, además de iluminar de manera general cómo los experimentos mentales son capaces propiciar inferencias con conclusiones verdaderas sobre los fenómenos, permite explicar los casos que emplean contrafácticos. En muchos experimentos mentales no hay fenómenos reales que puedan ser considerados como objetivos de la representación. No obstante, algunos experimentos mentales de este tipo, como es el caso del experimento del Demonio de Maxwell, han permitido explorar el alcance y los límites de ciertas leyes y teorías acerca de la naturaleza.

Un tercer elemento para sustanciar la hipótesis de que los experimentos mentales son un tipo de modelos científicos, es que desempeñan funciones análogas a las de los modelos en la práctica científica. Muchos experimentos mentales desempeñaron un papel importante en la articulación de principios teóricos. La concepción de los modelos como mediadores, desarrollada paralelamente por Cartwright (1999) y Morgan y Morrison (1999) proporciona un marco teórico para la interpretación de los experimentos mentales como mediadores.

Cartwright (1999) ha sostenido que una de las principales funciones de los modelos es ajustar las descripciones del dominio empírico con las representaciones matemáticas que proceden de la teoría. Para Cartwright (1983), las leyes fundamentales son de naturaleza abstracta, por lo que no pueden describir lo que ocurre en el mundo natural excepto que sean interpretadas como leyes *ceteris paribus*. Los principios fundamentales, particularmente en las teorías físicas, proporcionan relaciones puramente abstractas entre conceptos. Por esta razón, los científicos construyen modelos que pueden ser verdaderos de los objetos en el mundo, una capacidad que las leyes generales no tienen. Lo hacen a través de las leyes fenomenológicas que son descriptivas (a diferencia de las generales que son explicativas) y permiten vincular las apariencias con el mundo real. En trabajos más recientes, Cartwright (1999) propone que el tipo de orden que las leyes establecen, no se encuentra en el mundo empírico. El mundo es "veteado" (*dappled*), en él las relaciones causales no son inmediatas, el tipo de arreglo denominado "máquina nomológica" es "una disposición fija de componentes o factores, con capacidades estables que, en el tipo correcto de entorno con una operación repetida, dará lugar al tipo de comportamiento regular que representamos en nuestras leyes científicas" (Cartwright, 1999: 253). A este tipo de herramientas pertenecen los modelos.

Morgan y Morrison (1999) también sostienen que los modelos funcionan como mediadores y establecen una multiplicidad de maneras en las que pueden desempeñar esta tarea: como dispositivos para la aplicación de teorías, como instrumentos de medición, como herramientas para corregir teorías o para examinar sus alcances, como instrumentos para intervenir en el mundo y como objetos de investigación para explorar procesos para los cuales no tenemos una buena explicación. Varias de estas funciones pueden ser puestas en paralelo con las que algunos experimentos mentales desempeñaron en la historia de la ciencia. Morgan y Morrison afirman que los modelos poseen una autonomía relativa respecto de las teorías. Haciéndose eco de la aseveración de Hacking (1983), de

acuerdo con la cual los experimentos tienen vida propia en la práctica científica, argumentan que la multiplicidad de funciones que los modelos desempeñan es evidencia de su independencia respecto de las teorías. Asimismo, sostienen que la construcción de modelos procede de manera independiente respecto de los datos provenientes del mundo empírico. Así, por ejemplo el modelo del péndulo puede ser comprendido como el modelo de un tipo de movimiento descrito por la mecánica newtoniana aplicado a un dominio de objetos físicos:

Queremos usar el péndulo para medir la fuerza gravitacional y, en ese sentido, el proceso no comienza con una teoría, sino con un péndulo real. Pero también necesitamos una explicación teórica altamente detallada de cómo funciona en todos sus aspectos físicos. La mecánica newtoniana proporciona todas las piezas necesarias para describir el movimiento del péndulo, pero las leyes de la teoría no se pueden aplicar directamente al objeto. Las leyes describen varios tipos de movimiento en circunstancias idealizadas, pero aún necesitamos algo separado que nos permita aplicar estas leyes a objetos concretos. El modelo del péndulo desempeña este papel; proporciona un contexto más o menos idealizado donde se aplica la teoría. A partir de un modelo inicialmente idealizado, podemos construir las correcciones apropiadas para que el modelo se convierta en una representación cada vez más realista del péndulo real. (Morgan y Morrison, 1999: 15)

El punto de vista de Cartwright y Morgan y Morrison permite, a la vez, explicar la naturaleza *sui generis* de los modelos e iluminar la comprensión del funcionamiento de los experimentos mentales. Desde esta perspectiva, los modelos no son completamente independientes de toda teoría, aunque tampoco son constitutivos de una teoría determinada. Asimismo, su función en la práctica científica, lejos de ser despreciable, es central: permiten tanto representar como intervenir sobre los fenómenos. Si bien estas autoras no proporcionan una noción exhaustiva de representación, señalan que los modelos representan no porque imitan sino porque traducen a otras formas la naturaleza de un sistema y una teoría y lo hacen encarnando parcialmente al sistema en cuestión (Morgan y Morrison, 1999: 27).

Muchos ejemplos de experimentos mentales en la historia de la ciencia han desempeñado funciones mediadoras. Los que Brown (2011: 33) ha denominado *experimentos mentales constructivos* cumplen funciones que encajan con la conceptualización de los modelos como mediadores. Esta clase de experimentos puede ilustrar las consecuencias de una teoría que de otro modo resultarían altamente contraintuitivas, conjeturar fenómenos hipotéticos para los cuales no existe una teoría desarrollada, proponer una explicación igualmente hipotética de un fenómeno o postular fenómenos cuya existencia no resulta problemática y proporcionar una interpretación teórica para el caso (en ocasiones también presentar una nueva interpretación teórica que da sentido al caso).

En la concepción de Norton esta clase de experimentos se clasifican como del Tipo 2. En ellos se presenta un caso en cuya descripción se incluye una serie de detalles particulares acerca de un sistema físico imaginado y se presenta una hipótesis que explica el caso y que se aplicaría a muchos otros sistemas similares al presentado en el experimento. Esta clase de experimentos opera presentando un ejemplo que se considera como caso típico y eliminando, en la conclusión del experimento, todos los detalles particulares inesenciales descritos en la presentación del caso (Norton 1991: 131). Aunque la posición de Norton es deflacionaria respecto de los poderes epistémicos de los experimentos mentales, sostiene que los casos que se subsumen bajo esta categoría desempeñan un papel explicativo importante en el razonamiento científico. Norton y Brown coinciden en que un caso que se destaca por cumplir las funciones atribuidas a los experimentos mentales, del Tipo 2 y constructivos en sus respectivas clasificaciones, es el experimento del ascensor propuesto por Einstein ([1930] 1942: 199-242) para ilustrar el principio de equivalencia entre la masa inercial y la masa gravitatoria de un cuerpo.

5.3 Los modelos mentales

Las perspectivas que interpretan a los experimentos mentales como modelos mentales coinciden en que estos funcionan generando una representación mental de las relaciones causales, temporales y espaciales entre eventos y entidades que tienen lugar en los fenómenos. Estos modelos emplean como fundamento del nuevo conocimiento el conocimiento empírico acumulado de manera previa por los experimentadores y permiten sistematizarlo e integrarlo a un aparato teórico. Para generar representaciones confiables de los fenómenos, los modelos mentales apelan a habilidades de razonamiento comunes a otros procesos cognitivos como la anticipación, la visualización, la imaginación y la predicción. La selección de los elementos a representar es una estrategia típica en la construcción de esta clase de modelos. Aunque no todos los análisis coinciden en atribuir a las estrategias de idealización la misma importancia en la construcción de un modelo mental, concuerdan en que simplificar las situaciones representadas y aislar algunos elementos relevantes son estrategias típicamente empleadas en el modelado mental, ya que el recurso a la abstracción o simplificación permite presentar las situaciones imaginarias como paradigmáticas de la aplicación de un principio teórico.

El propósito de esta sección es evaluar estas interpretaciones de los experimentos mentales como una clase particular de modelos. Se muestra que, aunque la noción de modelo mental es de utilidad para explicar la producción de nuevo conocimiento sobre el mundo natural en ausencia de datos nuevos, el carácter psicologista de algunos de estas explicaciones obstaculiza la posibilidad de elaborar criterios objetivos de adecuación. Desde este punto de vista, la confiabilidad de un modelado mental depende de la combinación adecuada del conocimiento teórico previo, de intuiciones empíricas adecuadas y de la aplicación de estrategias de representación que en cada caso varían de acuerdo con la formación y conocimiento de fondo del experimentador.

La interpretación de los experimentos mentales como modelos mentales fue introducida de manera independiente por Nancy Nersessian (1991) y Nenad Mišćević (1992). Ambas perspectivas se apoyan en las investigaciones en psicología cognitiva desarrolladas por Phillip Johnson-Laird (1980 y 1982) e intentan sustentar una propuesta de base empirista que se propone ampliar la explicación proporcionada por Mach (1897) sobre el papel del conocimiento empírico acumulado en la generación de nuevo conocimiento en los experimentos mentales. Para Mišćević (1992) los experimentos mentales funcionan reorganizando información empírica. El tipo de representación que tiene lugar en estos experimentos está mediado por la imaginación y constreñido por la experiencia:

El experimentador mental puede usar la deducción, el razonamiento inductivo y las matemáticas (generalmente, aritmética y geometría), para imaginar libremente cosas y luego reorganizar estos elementos imaginados. La materia con la que trabaja el sujeto es experiencial, y sus conclusiones serán, en última instancia, derivadas de la experiencia. Su física será la habitual, *a posteriori*. (Mišćević, 1992: 215)

El conocimiento cotidiano al que apelan los experimentos mentales es producto de un largo proceso que involucra la jerarquización de un conjunto de creencias. Estas creencias son obtenidas a partir de diferentes mecanismos: perceptivos, inferenciales, de prueba y error, de evaluación de alternativas, etc.; se ordenan a partir de reglas de inferencia condicional y son evaluadas mediante contrastación empírica. Este conocimiento acumulado se activa en la construcción de un modelo mental y permite anticipar el resultado más plausible de una situación imaginaria planteada por la narrativa del experimento mental.

Desde la perspectiva de Mišćević, el principal desafío que una explicación filosófica del funcionamiento de los experimentos mentales debe afrontar es el de dar cuenta de cómo una descripción narrativa de una situación hipotética permite pasar de manera casi instantánea a conclusiones acerca del comportamiento de los fenómenos naturales. Una explicación plausible de cómo esto ocurre es que,

aunque los experimentos mentales carecen del elemento de intervención en el mundo natural, la posibilidad de establecer de manera clara las condiciones de contorno de las situaciones representadas permite que los experimentadores atiendan a ciertos aspectos de un fenómeno, aislando en la imaginación los factores perturbadores. Realizar un experimento mental consiste, entonces, en una forma particular de manipulación de esta clase de representaciones.

Por otra parte, los modelos mentales involucrados en la realización de un experimento mental son, por lo general, de carácter dinámico. Esta característica también es explicativa, dentro del enfoque de Mišćević, de la instantaneidad con la que se pueden derivar conclusiones sobre los fenómenos naturales a partir de la ejecución de un experimento mental. Mediante la construcción de un modelo mental es posible resolver problemas acerca de relaciones espaciales de una manera mucho más directa que mediante la formulación de un argumento. La ventaja de esta clase de modelos es que proporcionan una visión de la situación muy cercana a la experiencia cotidiana. Los ejemplos típicos de modelos mentales son los que están involucrados en la comprensión de historias y la planificación ordinaria de actividades. Cuando un lector encuentra en un relato la descripción de una situación, construye un modelo, es decir, una imagen cuasi-espacial de ella (Mišćević, 1992: 221). Lo mismo ocurre en los experimentos mentales. Los casos que proporcionan una representación adecuada de un problema, permiten a los sujetos representarse una situación problemática que pueden manipular a partir de la movilización de ciertos recursos cognitivos (habilidades cognitivas, conocimiento implícito, creencias perceptuales). Las representaciones cuasi-espaciales (que en ocasiones involucran de manera central representaciones visuales) son, desde este punto de vista, más fáciles de manipular que las representaciones teóricas o aritméticas.

En síntesis, los modelos mentales que se construyen en los experimentos mentales permiten solucionar problemas planteados por las descripciones de situaciones hipotéticas filtrando alternativas a través del conocimiento empírico

acumulado (Ver Mišćević, 1992: 225). La manipulación de esta clase de escenarios es *cuasi-experiencial* en el sentido de que produce creencias no inferenciales similares a las perceptivas. Estos rasgos explican la superioridad heurística de los experimentos mentales respecto de los argumentos y dan sentido a la producción de conocimiento nuevo en ausencia de nuevos datos.

Para Nersessian (1992, 1995, 1999, 2012 y 2018) los experimentos mentales funcionan también a partir de la construcción de modelos mentales. Este es el mismo tipo de actividad que tiene lugar en otros procesos de modelado que se emplean habitualmente en las ciencias físicas. La construcción de un modelo en física consiste desde este punto de vista en la representación una estructura o proceso físico para lo cual se requiere el conocimiento de los principios básicos que gobiernan ese dominio de fenómenos:

[...] un modelo mental es un análogo estructural de un mundo real o situación, evento o proceso imaginario que una persona construye para razonar dentro de la mente. Lo que significa para un modelo mental ser un análogo estructural, es que encarna una representación de las relaciones espaciales y temporales entre la estructura causal y la conexión de los eventos y las entidades representadas. (Nersessian, 1995: 209)

Desde este punto de vista, el modelado mental es una parte esencial de la práctica científica. Gracias a este ejercicio cognitivo, Newton por ejemplo, pudo ver los puntos en común entre los movimientos de los planetas y los proyectiles que le permitieron formular una representación matemática unificada del movimiento (véase Nersessian 1995: 211-212). Estas prácticas constituyen una extensión de la experimentación y se basan en la capacidad de razonar por medio de modelos analógicos. Estrictamente, la formulación de un experimento mental es un proceso que consta de dos instancias:

- 1) La construcción de un modelo dinámico en la mente del experimentador, quien imagina una secuencia de eventos y procesos y anticipa un resultado.

2) La elaboración de una narrativa para describir el escenario y comunicarlo a otros, esto es, para que otros construyan y pongan a correr la simulación correspondiente y presumiblemente obtengan el mismo resultado.

La comunicación de los experimentos mentales depende de una descripción narrativa mientras que la construcción de los modelos en la imaginación depende fundamentalmente del conocimiento empírico, que no necesariamente se encuentra disponible proposicionalmente. La función de la narrativa de un experimento mental es guiar al lector en la construcción de un análogo estructural de la situación descrita por él y en la realización de inferencias a partir de la simulación de eventos y procesos representados en él. La situación construida hereda su 'fuerza empírica'³³ del hecho de ser abstraída de las experiencias y actividades en el mundo físico y del conocimiento, conceptualizaciones, y supuestos acerca del mundo que el experimentador posea. La confiabilidad del modelo depende de que las conclusiones que se derivan de la ejecución del experimento mental tengan aplicación en los fenómenos reales (véase Nersessian 1992: 297).

Aunque las propuestas de Mišćević y Nersessian proporcionan una explicación de los procesos psicológicos vinculados a la ejecución de un experimento mental, no es posible derivar de este enfoque criterios objetivos de adecuación para los experimentos mentales. Las características que debe tener la narrativa que describe una serie de eventos o procesos a partir de los cuales se construye un modelo mental capaz de representar un fenómeno son fundamentalmente dos: evocar el conocimiento empírico y teórico disponible en el sistema de creencias del experimentador y apelar a las habilidades cognitivas de representación visual, suposición y anticipación pertinentes para configurar el modelo. No obstante, estas características no son suficientes para distinguir los

³³ Nersessian entiende por esto la cualidad de ser una representación vívida de un fenómeno.

casos en los cuales la narrativa del experimento permite construir un modelo mental apropiado del sistema objetivo. La perspectiva no proporciona indicios de cómo las habilidades de representación visual, suposición y anticipación pueden ser constreñidas para proporcionar representaciones adecuadas.

La perspectiva de Cooper (2005) se propone superar algunas dificultades presentes en los marcos interpretativos propuestos por Mišćević y Nersessian. Estas posiciones dependen de la noción de modelo mental definido como un "análogo estructural de la situación descrita" (Nersessian, 1992: 297) o como una representación cuasi-espacial de una serie de eventos o procesos (Mišćević, 1992: 220). Para Cooper, en cambio, la existencia de modelos mentales postulada sobre la base de la evidencia proporcionada por una serie de estudios en psicología cognitiva,³⁴ es dudosa y no constituye un requisito explicativo para sostener la interpretación de acuerdo con la cual los experimentos mentales funcionan mediante la construcción de esta clase de modelos. La noción de modelo mental no es, desde este punto de vista, un requisito indispensable para la elaboración de una representación con la cual pueda anticiparse el comportamiento de un sistema físico.

Para Cooper, diferentes tipos de modelos, como los modelos matemáticos o los modelos a escala, posibilitan en principio el mismo tipo de razonamientos y emplean los mismos recursos cognitivos que los experimentos mentales (imaginación, conocimiento empírico acumulado, capacidad de realizar inferencias). Cooper considera que la noción de modelo mental es demasiado estrecha para abarcar todas estas prácticas, que son esencialmente del mismo tipo que los experimentos mentales. Por esa razón, adopta una noción más amplia de modelo. Un modelo, según Cooper, consiste en una representación de un sistema o de una secuencia de eventos a partir de un conjunto de supuestos de fondo bien definidos que restringen su construcción. Los experimentos mentales

³⁴ Se refiere a las investigaciones de Johnson-Laird (1980 y 1982) en las que se apoyan los trabajos Mišćević y Nersessian.

funcionan como plantillas de mundos posibles (Cooper, 2005: 336), en los cuales la posibilidad, física o lógica, limita la plausibilidad del escenario. Estos experimentos permiten obtener conocimiento contingente sobre el mundo, dado un trasfondo de supuestos que operan limitando la consistencia interna de la representación. En palabras de Cooper:

Al contestar las preguntas del tipo "¿qué pasaría si...?", predecimos cómo se comportarían las entidades imaginarias de la misma manera que predecimos cómo se comportarán las entidades reales. A veces tendremos leyes explícitas que regirán cómo las entidades del tipo que estamos imaginando actúan en los tipos de situación que estamos imaginando. Por lo tanto, podemos predecir qué tan rápido caerían las masas imaginarias bajo la gravedad de la misma manera que podemos predecir qué tan rápido caerán las masas reales. Conectamos los valores relevantes en ecuaciones y calculamos la predicción. Si las masas son reales o imaginarias no hay diferencia. También podemos emplear la comprensión tácita de las leyes que no podríamos formular estrictamente. A veces, las respuestas a las preguntas de "¿qué pasaría si...?" son proporcionadas por leyes implícitas contenidas en las implicaciones de los conceptos que estamos usando. (Cooper, 2005: 336)

La noción de modelo de Cooper es más amplia que la de modelo mental que restringe el funcionamiento de los experimentos mentales a la construcción de modelos en la mente del experimentador. Asimismo, permite elaborar un criterio de corrección para evaluar experimentos mentales exitosos. De acuerdo con este punto de vista, un experimento mental confiable permite construir un modelo coherente de un sistema. La coherencia del modelo está determinada por la posibilidad de que los supuestos teóricos que constriñen la representación den sentido a la situación hipotética representada.

Una manera de evaluar las propuestas de Mišćević, Nersessian y Cooper es por medio del examen de un experimento mental que es explicado por las tres propuestas: el experimento de los cuerpos en caída libre de Galileo. Para Mišćević, el experimento de Galileo involucra un razonamiento acerca de la relación entre el peso de un cuerpo y su velocidad de la caída. El experimentador debe construir tres modelos empleando como premisa en todos los casos el supuesto

aristotélico de acuerdo con el cual la velocidad de la caída es proporcional al peso. Un primer modelo se emplea para representar la relación entre la velocidad de la caída y el peso del cuerpo más pesado y otro para representar este vínculo para el cuerpo más liviano. Luego el experimentador debe construir otro modelo en el que considere cómo caería un compuesto de un cuerpo más pesado y otro más liviano en relación con otros dos cuerpos no ligados que caen libremente, el más pesado y del más liviano. Esta operación exige que se integren los dos primeros modelos. Al intentar realizar esta integración, el experimentador advierte que es imposible acomodar consistentemente los presupuestos sobre el comportamiento de los modelos sobre cada cuerpo en caída libre al comportamiento de un cuerpo compuesto por ambos; de este modo, concluye que los modelos producen una contradicción con el supuesto inicial de acuerdo con la cual la velocidad es proporcional al peso (véase Miščević, 1992: 222).

Para Nersessian, la narrativa del experimento de Galileo describe la evolución de un sistema y guía a los experimentadores en la construcción de un modelo dinámico en el que se deben imaginar objetos de diferente peso cayendo de manera individual y formando un compuesto. A partir de estas representaciones, los experimentadores deben derivar conclusiones sobre el comportamiento de los sistemas reales comparando las situaciones en las que se dejan caer cuerpos de diferente peso separadamente y combinados (véase Nersessian, 2018: 371).

Para Cooper, el experimento de Galileo muestra que lo que el principio aristotélico predice, esto es, que los cuerpos caen con una velocidad proporcional a su peso, es imposible. Un conjunto de supuestos de fondo, entre ellos: que el color, la forma y la composición química no tienen efecto sobre la velocidad a la que cae una masa, operan constriñendo las opciones disponibles para explicar el fenómeno (véase Cooper, 2005: 340-341). El experimento muestra que el único modelo consistente es el que establece que la velocidad de la caída de los cuerpos no es una función del peso.

Aunque las perspectivas son similares, podemos identificar tres análisis diferentes del caso. Para Mišćević, consiste en la construcción de dos modelos y en la integración de estos en un tercero. De esta operación se deriva el rechazo de un marco teórico en la construcción de un modelo integrador de los dos primeros. Para Nerssesian, el experimento consiste en la construcción de un modelo mental a partir del cual el experimentador debe inferir la inconsistencia de un marco teórico para explicar el comportamiento de un sistema real. El experimentador puede derivar consecuencias que se aplican a los sistemas reales porque el experimento presenta un caso típico. Para Cooper, el experimento consiste en la construcción de un modelo que integre de manera coherente el comportamiento de los tres sistemas (el cuerpo pesado en caída libre, el cuerpo liviano en caída libre y el compuesto en caída libre) a partir de un conjunto de supuestos de fondo que limitan la construcción de dicho modelo.

En relación con los análisis del caso, identificamos una importante limitación en la presentación del mismo. Si bien el experimento de Galileo puede describirse a partir de una serie de actividades que es posible imaginar y que contribuyen a inferir conclusiones, esto no es todo lo que involucra la ejecución del experimento. Esto es claro si comparamos el ejemplo con otros casos en los cuales el “resultado” del experimento (o más bien las conclusiones que se siguen de este) no depende esencialmente de la representación de una situación espacio-temporal, sino de que el experimentador reconozca la relación entre las variables y los principios teóricos relevantes. En algunos casos, el tipo de razonamiento que el experimento propone depende de que el experimentador pueda apelar a conocimiento experto. Diversos experimentos en mecánica cuántica, como el caso del experimento mental de Feynman (1965) analizado en el Capítulo 2 de este trabajo, por ejemplo, no apelan directamente a razonamientos basados en el conocimiento perceptual, sino más bien a analogías con fenómenos conocidos. En el caso del Feynman, el experimento propone imaginar el tipo de patrón de distribución de los impactos que podría obtenerse en un escenario como el

descrito si se lanzaran proyectiles a través de una pared que tiene dos rendijas. Pero esta visualización desempeña un papel auxiliar en la obtención de conclusiones ya que para arribar a las conclusiones adecuadas el experimentador debe tener ciertos conocimientos previos (por ejemplo, los resultados de la experimentación previa, como el experimento de interferencia de Young), y aplicar ciertos principios relevantes que permiten articular el comportamiento del sistema en cada una de las partes del experimento (como el principio de incertidumbre).

En segundo lugar, consideramos que los análisis del caso no logran identificar el papel específico que los principios físicos, lógicos y geométricos desempeñan en el funcionamiento de los experimentos mentales. Estos funcionan como reglas para establecer la evolución de un fenómeno dadas ciertas condiciones iniciales y anticipar un resultado en un escenario imaginario. Dichos principios no son reglas psicológicas (excepto que se afirme la coincidencia de dichas reglas con las leyes lógicas u otro tipo de principios que se consideren universalmente válidos) sino reglas que se aplican explícitamente en el razonamiento científico. Las reglas inferenciales establecen qué conclusiones pueden derivarse correctamente, los principios geométricos restringen las representaciones espaciales y la construcción de diagramas y los principios teóricos determinan relaciones causales o de coexistencia entre propiedades físicas. Estos principios diferencian el tipo de representación que tiene lugar en los experimentos mentales de las representaciones a las que puede dar lugar una narrativa de ficción.

Del enfoque que interpreta a los experimentos mentales como modelos mentales rechazamos fundamentalmente su ontología. Aunque podría sostenerse que los experimentos mentales derivan de una habilidad evolutivamente desarrollada vinculada con la capacidad perceptiva e imaginativa, la construcción de modelos en los experimentos mentales no está centrada en las habilidades perceptuales sino en la aplicación de principios teóricos a fenómenos o sistemas idealizados. Desde nuestro punto de vista, el tipo de actividad que tiene lugar en

los experimentos mentales forma parte de una práctica sofisticada que depende de la selección de algunos aspectos del sistema, de la aplicación de principios teóricos y del conocimiento experto del experimentador.

No obstante, hay algunos aspectos de los enfoques de los modelos mentales que consideramos que son esclarecedoras del funcionamiento de los experimentos mentales. La interpretación de Nerssesian, de acuerdo con la cual los experimentos mentales consisten en una forma de razonamiento analógico, concuerda con el examen de los casos y las observaciones que hemos realizado hasta el momento sobre el funcionamiento de esta clase de experimentos. Asimismo, el aspecto dinámico de la representación que señala Nerssesian, es decir, la descripción de la evolución temporal de un sistema, resulta útil para explicar el elemento heurístico de los experimentos mentales.

La narración delimita cuáles son las transiciones específicas que gobiernan lo que ocurre. La construcción y realización del experimento hace uso de mecanismos de inferencia, representaciones existentes y conocimiento científico y general del mundo para hacer transformaciones realistas de un posible estado físico al siguiente. (Nerssesian, 1999: 19)

Las perspectivas sobre los modelos mentales hacen referencia, además, al hecho de que los experimentos mentales emplean una forma de razonamiento sustitutivo que consiste en la representación en la imaginación de la evolución de un fenómeno y que a partir de esta en algunos casos es posible inferir conclusiones de las cuales no se pueden obtener confirmaciones empíricas. Consideramos que este aspecto se relaciona con los principios físicos en los que se enmarca el experimento y a lo que nos permitan inferir los conceptos empleados en ellos, así como también al estado del conocimiento técnico. En el citado caso del experimento de la doble rendija, Feynman pudo anticipar resultados en este sentido, aunque solo mucho tiempo después pudo obtenerse confirmación empírica de sus conclusiones.

Por último, la sugerencia de Cooper, de acuerdo con la cual los experimentos mentales permiten construir un modelo consistente de un fenómeno, se aproxima a nuestra interpretación de los experimentos mentales. Es necesario aclarar, en vistas a proporcionar criterios de confiabilidad de los experimentos mentales, que el modelo podría considerarse consistente con el conocimiento experimental previo, los supuestos teóricos aceptados, y ciertos compromisos metafísicos.

5.4 Las simulaciones computacionales

El extendido uso de simulaciones computacionales en el contexto de justificación de teorías que tratan acerca de fenómenos sobre los cuales la experimentación real no es posible motiva su comparación con los experimentos mentales. Dentro de estos dominios, las funciones de las primeras son en apariencia similares a las de éstos últimos. Recientemente, ha habido una avalancha de nuevas aplicaciones de las teorías científicas, fundamentalmente en el campo de la física, ayudada en gran parte por la creciente capacidad de utilizar técnicas digitales para construir modelos dóciles de mayor complejidad usando los recursos teóricos disponibles. Algunas posiciones atribuyen a las simulaciones un carácter similar al de la experimentación real y consideran que su ejecución puede contribuir a la justificación de una teoría científica. Esta perspectiva se apoya en las similitudes entre las simulaciones y los experimentos y cuestiona el papel esencial de la interacción causal con el fenómeno como razón para atribuir un estatus epistemológico privilegiado de los experimentos reales (Morrison, 2009 y 2015). La tesis básica es que las simulaciones nos permiten crear el tipo de ambiente controlado en el que podemos variar las condiciones iniciales, los valores de los parámetros, etc., y en este sentido funcionan como el equipamiento de laboratorio usado para medir propiedades de los fenómenos físicos.

Aunque no existe una noción unificada de simulación computacional y de hecho varias nociones emparentadas se solapan (modelos computacionales, experimentos virtuales, experimentos simulados), las simulaciones podrían caracterizarse de manera general como instrumentos que tienen la función de representar la estructura o el comportamiento de un determinado sistema físico. En ellas, una secuencia de eventos sirve como representación de otra secuencia de eventos.

Por "simulación por computadora" (o simplemente "simulación", para abreviar) nos referimos a un método que involucra de manera crucial la ejecución de un código de computadora, que traza al menos soluciones aproximadas a ecuaciones que son partes esenciales de un modelo de la evolución dinámica de algún sistema objetivo real o imaginario en condiciones reales o contrafácticas. De esta manera, el comportamiento dinámico dentro del sistema de destino se imita utilizando una computadora digital. (Beisbart, 2019: 39)

La ejecución de un experimento mental por su parte, puede describirse como un ejercicio imaginativo que consiste en la representación de un escenario posible del que se pretenden derivar algunas consecuencias acerca del comportamiento de fenómenos físicos. Una de sus características fundamentales es que, a partir de la descripción de un conjunto de condiciones iniciales, los experimentadores deben activar un mecanismo o manipular alguna variable y "ver qué sucede". A partir de estas caracterizaciones generales, puede decirse que los experimentos mentales y las simulaciones computacionales comparten al menos dos rasgos importantes: 1) se proponen proveer información acerca del comportamiento del mundo natural en ausencia de intervención activa sobre éste; y 2) dependen de la construcción de modelos dinámicos que se proponen simular la evolución de un sistema físico.

El examen filosófico de las simulaciones computacionales se realizó fundamentalmente desde los marcos teóricos proporcionados por la filosofía de la experimentación. Desde este punto de vista, las simulaciones suelen concebirse como un tipo especial de práctica experimental. La noción general de simulación

computacional que formulamos anteriormente es una versión de la propuesta por Hartmann (1996):

[...] una simulación resulta cuando las ecuaciones de un modelo dinámico son resueltas. Este modelo es diseñado para imitar la evolución de un sistema temporal real. Para poner el mismo punto de otra manera, una simulación imita un proceso mediante otro proceso. En esta definición el término 'proceso' se refiere solamente a algún objeto o sistema cuyo estado cambia en el tiempo. Si la simulación corre en una computadora se denomina simulación computacional. (Hartmann, 1996: 88)

Esta línea interpretativa ha sido criticada por reducir la complejidad metodológica de las simulaciones computacionales al menos en dos sentidos. En primer lugar, porque supone que una simulación computacional resulta de la implementación directa en una computadora de un modelo matemático, aunque los modelos de simulación deben ser escritos en un lenguaje computacional apropiado para poder ser implementados. Así, de representar sistemas continuos se pasa a representar procesos discretos (Durán, 2018: 129). En segundo lugar, porque, por lo general, no se dispone de modelos empíricos completos para representar los sistemas que se quieren simular. Por el contrario, las simulaciones se sirven de múltiples modelos parciales que individualmente solo representan una porción del objetivo.

La noción de "plantilla computacional" propuesta por Humphreys (2004: 79) proporciona, de acuerdo con algunos críticos, un marco teórico más apropiado para comprender esta práctica científica y para explicitar la relación entre simulaciones y modelos científicos. El esquema de Humphreys incluye la especificación de la ontología de la representación (objetos y relaciones), las estrategias de corrección empleadas para aproximar el modelo a los datos empíricos, una interpretación y una representación (generalmente gráfica, en forma de función o numérica). La característica más importante de este enfoque es que pone el énfasis en los elementos particulares de las simulaciones, ausentes en otras prácticas, que, al menos en principio, los otros enfoques

existentes en la filosofía de la ciencia son incapaces de explicar. El rasgo más saliente de las simulaciones es el paso de representaciones abstractas de inferencias lógicas y matemáticas a procesos temporales de cálculo y representación (véase Durán, 2018: 137).

Aunque desde un punto de vista general existen semejanzas entre los experimentos mentales y las simulaciones computacionales, es claro que estas últimas emplean un conjunto de recursos y estrategias que no están presentes en los experimentos mentales. Una diferencia metodológica fundamental entre estas prácticas es la ausencia de algoritmos formales en los experimentos mentales. Aunque en cierto sentido puede comprenderse a esta clase de experimentos y a las simulaciones como maneras de explorar mundos hipotéticos, los experimentos mentales apelan principalmente a procesos cognitivos vinculados a la representación visual mientras que las simulaciones descansan en iteraciones automatizadas de algoritmos formales (véase Lenhard, 2018: 554). Esta diferencia impone diferentes restricciones en la capacidad de representación

Aunque puede considerarse que experimentos mentales, modelos físicos y modelos computacionales conforman un espectro de prácticas de razonamiento sustitutivo, los experimentos mentales emplean vehículos de representación diferentes de los utilizados en los modelos físicos y las simulaciones computacionales. Los experimentos mentales se construyen en base a algunos elementos (que podrían considerarse) concretos: intuiciones empíricas, representaciones visuales, diagramas y otros rudimentos de naturaleza abstracta: leyes científicas, ecuaciones matemáticas. Los vehículos de representación empleados en los experimentos mentales implican limitaciones respecto del tipo de representación que pueden proporcionar. Los experimentos mentales no admiten la simulación de escenarios hipotéticos más allá de las condiciones generadas por el escenario específico porque el proceso de simulación mental es impulsado por el comportamiento de sus componentes. Dicho proceso está constreñido por un lado por la capacidad de procesamiento de los

experimentadores y por otro lado la plausibilidad que a su resultado le aportan los principios o leyes científicas y las intuiciones empíricas evocadas. Podría agregarse que incluso la retórica de la narrativa del experimento cumple un rol determinante en la plausibilidad con la cual el resultado aparece a los lectores. Esta limitación ha sido interpretada por algunos autores como un rasgo ligado a la inmutabilidad de esta clase de experimentos (véase Hacking, 1993: 307). Algunos autores han argumentado que el alcance de las simulaciones computacionales es más amplio que el de los experimentos mentales porque emplean vehículos de representación diferentes, más versátiles y eficaces que los usados en los experimentos mentales

Los modelos computacionales, particularmente aquellos con visualizaciones, van más allá de los experimentos mentales, ya que estos modelos pueden generar un comportamiento complejo similar al de los fenómenos naturales, proporcionando explicaciones de por qué existen estos fenómenos y no otros. También permiten interrogar y "profundizar" en el comportamiento (de los fenómenos simulados) para obtener detalles. Tales modelos permiten la externalización del pensamiento creativo en la ciencia y, por lo tanto, extienden la capacidad cognitiva para la simulación mental, de forma similar a la forma en que los telescopios han ampliado la capacidad perceptiva de ver. [...] Así como ningún científico estudia más las estrellas a simple vista, nadie usaría un experimento mental para investigar los fenómenos complejos estudiados por la ciencia contemporánea. (Chandrasekaran y Nersessian, 2012: 272)

Otra diferencia entre experimentos mentales y simulaciones computacionales se relaciona con la inteligibilidad u opacidad epistémica de estas prácticas (Di Paolo *et al*, 2000: 497). Para que un experimento mental sea exitoso debe satisfacer altos estándares de inteligibilidad porque todo el proceso tiene lugar en la imaginación. Si de la narrativa no puede anticiparse con facilidad el estado del sistema, es decir, qué cambios ocurrirán en el estado inicial dadas ciertas condiciones, si no se comprende por qué cierto resultado sucede, el experimento mental falla. La transparencia es una condición para la viabilidad del experimento mental. En muchos casos, los experimentos mentales parten de supuestos teóricos formulados explícitamente, razón por la cual se ha señalado

que su función principal es la de cimentar el conocimiento existente antes que la de crearlo. La narrativa del experimento debe fluir de manera continua para no crear “huecos” en la representación mental. La representación del experimento puede requerir de varias iteraciones hasta dar lugar a una conclusión clara.

Las simulaciones computacionales son epistemológicamente opacas, lo que significa que exploran hipótesis, pero de una manera computacional que los seres humanos no pueden seguir. Los procesos algorítmicos son muy diferentes de los procesos de intuición y representación visual a los que apelan los experimentos mentales. Son epistemológicamente opacas, no porque no estén claras las sucesiones de pasos que constituyen los algoritmos, sino porque la multitud de pasos interrelacionados, debido a su complejidad, hace al proceso opaco en sí mismo. Aunque las interacciones en un sistema simple pueden ser modeladas de una manera relativamente simple, un gran número de interacciones da lugar a modelos dinámicos complejos. En esto consiste, fundamentalmente, la opacidad de las simulaciones.

Esta diferencia ha sido tematizada por Lenhard (2018). De acuerdo con este análisis, experimentos mentales y simulaciones se asemejan en que ambas estrategias hacen uso de las iteraciones, sin embargo, difieren en los tipos de iteraciones que emplean y en las funciones que estas desempeñan en cada caso. Lenhard distingue entre dos formas de iteración: por convergencia y del tipo “atlas”. Las iteraciones por convergencia permiten explorar un problema en base a numerosos avances y retrocesos. Este tipo de proceso es similar al de explorar una nueva ruta hasta que se vuelve tu camino rutinario; al inicio tiene muchos avances y retrocesos pero finalmente se vuelve rutina, después de muchas iteraciones el camino “sedimenta”. Este tipo de iteración se parece a la repetición rutinaria de un ejercicio. Dado que los experimentos mentales se apoyan en la intuición y el conocimiento de fondo de los experimentadores, funcionan en base a intuición “entrenada” de manera que sus resultados aparecen de manera perspicua, como los movimientos de un pianista son el resultado de la práctica

sostenida y rutinaria. El segundo tipo de iteración consiste en agotar todas las posibilidades creando un compendio o atlas de ellas.

Los experimentos mentales emplean el tipo de iteración por convergencia. En ellos, la repetición levemente variada de un algoritmo elimina la opacidad. Las simulaciones compensan la opacidad epistemológica con iteraciones del tipo atlas. Los algoritmos de iteración utilizan un elevado poder computacional y pueden funcionar de maneras que los experimentos mentales no son capaces de ejecutar. El elemento persuasivo de los experimentos mentales (que Lenhard denomina el *efecto "aha"*) se debe a la transparencia epistémica que, en contraste con lo que sucede en las simulaciones, resulta en un instrumento basado en una colección de cálculos simples. Los resultados de las simulaciones, en cambio, crean cierta distancia cognitiva respecto del experimentador (en esto consiste su opacidad) incluso cuando las visualizaciones ayudan a reducir esta distancia. De acuerdo con Lenhard en las simulaciones computacionales un modelo se "deja correr" en numerosas ocasiones con ligeras variaciones, lo que a diferencia de la ejecución de un experimento mental no elimina la opacidad. Las iteraciones automáticas que son posibles de realizar en la computadora exploran el espacio posible del comportamiento de un modelo. Por el extenso conjunto de iteraciones realizadas, las simulaciones generan un "atlas" que compila muchos "mapas" únicos. En conclusión la iteración en los experimentos mentales eventualmente hace a las iteraciones superfluas cristalizando un enfoque más intuitivo, en las simulaciones las iteraciones son estructuralmente necesarias (Lenhard, 2018: 565).

En síntesis, aunque experimentos mentales y simulaciones computacionales tienen características comunes, fundamentalmente que pueden ser considerados formas de razonamiento simulativo, los experimentos mentales, a diferencia de las simulaciones computacionales no emplean algoritmos formales. Los recursos empleados en las simulaciones computacionales y los experimentos mentales presentan una diferencia en el grado de detalle con el que se puede explorar una

situación hipotética. Los experimentos mentales emplean un mecanismo de iteración diferente del utilizado en las simulaciones computacionales y, por esa razón, los criterios de confiabilidad de estos experimentos difieren de los criterios empleados para evaluar simulaciones computacionales.

5.5 Experimentos mentales y estrategias de idealización

Todos los tratamientos filosóficos de los experimentos mentales coinciden en señalar que estos experimentos emplean idealizaciones, abstracciones y contrafácticos de manera esencial. Las descripciones de los escenarios imaginarios y de los aparatos experimentales e instrumentales invocan, por ejemplo, superficies sin fricción o esferas perfectas. En otros casos, describen la forma y la composición de los objetos pero no sus colores, olores o sabores, y en ocasiones hipotetizan circunstancias que no pueden darse en la realidad.

Algunas de las interpretaciones disponibles sobre los poderes epistémicos de esta clase de experimentos les otorgan a las idealizaciones un papel constitutivo de su poder heurístico. Estas estrategias favorecen, desde este punto de vista, la instantaneidad con la que es posible derivar conclusiones aplicables a situaciones en el mundo natural, dejando de lado factores irrelevantes o eliminando la contribución de algunas causas.

Esta sección se propone proporcionar una elucidación de la noción de idealización científica y mostrar que los experimentos mentales emplean el mismo tipo de estrategias que utilizan otros modelos científicos. Se muestra en base a un caso de estudio que estas estrategias son las que facilitan la generalización de las conclusiones y la articulación de los principios teóricos. Se argumenta que la narrativa de los experimentos mentales permite elaborar modelos idealizados de los fenómenos, lo que permitiría atribuirles a esta clase de experimentos un papel en el contexto de justificación de las teorías físicas,

específicamente en la articulación de principios teóricos. En este sentido los experimentos mentales desempeñan la función de mediadores (tal como la interpretan Cartwright, 1999 y Morgan y Morrison, 1999) ilustrando las consecuencias de una teoría que de otro modo resultarían altamente contraintuitivas, proponiendo una explicación hipotética de un fenómeno, postulando fenómenos cuya existencia no resulta problemática o proporcionando una interpretación teórica para el caso (en ocasiones también presentar una nueva interpretación teórica que da sentido al caso). Desde esta perspectiva, el empleo de simplificaciones, abstracciones e hipótesis contrafácticas permite construir modelos simplificados de los fenómenos de los que se pueden extraer conclusiones sobre el comportamiento de los sistemas físicos reales.

Las estrategias de idealización constituyen un recurso ampliamente usado en la práctica científica para aproximar las descripciones teóricas a los fenómenos (Ladyman, 2008: 385). Su uso no es privativo de los modelos teóricos. En los experimentos reales los sistemas físicos son simplificados, aislando o reduciendo la influencia de una serie de factores, haciéndolos de este modo más manejables en vistas a indagar de manera más eficiente las correlaciones entre las variables. Por otra parte, las idealizaciones matemáticas son ubicuas en física. Estas permiten tratar problemas complejos reduciendo el tiempo de cálculo y empleando los recursos computacionales disponibles. En síntesis, la distorsión deliberada es admitida en los experimentos reales y, de modo general, es una práctica habitual en física porque posibilita la manipulación de un sistema para la evaluación de hipótesis, permite construir casos ejemplares para la aplicación de teorías e incluso hace posible la exploración de sistemas para los cuales no existe un marco teórico interpretativo.

Los experimentos mentales emplean idealizaciones de manera constitutiva. Si bien esta clase de experimentos no hace uso de idealizaciones matemáticas, ya que en ellos no tiene lugar no el recurso a la matematización o

computación, las narrativas de esta clase de experimentos se caracterizan por el uso de simplificaciones, distorsiones o abstracciones.

De manera muy general la noción de idealización puede caracterizarse como: “la introducción intencional de distorsión en las teorías o en los modelos científicos” (Weisberg, 2007: 240). La necesidad de introducir distorsiones³⁵ en las explicaciones teóricas se debe a que en ocasiones las leyes contradicen la descripción del fenómeno al que pretenden aplicarse. Las simplificaciones se emplean en los modelos científicos para aproximar las descripciones teóricas a los fenómenos, entendiendo que éstos son entidades complejas que no pueden conocerse en su totalidad. Por ejemplo, las leyes del movimiento planetario de Kepler describen las propiedades cinemáticas de las trayectorias de los planetas en un modelo heliocéntrico del Sistema Solar. Las mismas resultan explicables por medio de la ley de gravitación universal de Newton: todos los cuerpos se atraen con una fuerza que es directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia que los separa. No obstante, las órbitas elípticas de Kepler son imposibles si se toma en cuenta en la aplicación de la ley los efectos gravitacionales de los planetas en el sol y entre los planetas, debido a que las órbitas resultan perturbadas por la composición de las atracciones gravitatorias de los diferentes planetas entre sí. Por lo que en sentido estricto se aplican solo a un sistema de dos cuerpos: Planeta-Sol.³⁶

Es posible distinguir diferentes estrategias de idealización, tales como *simplificaciones*, *aproximaciones*, *abstracciones* y *distorsiones*. Si bien no todos estos términos tienen una definición precisa, es posible clarificar algunas diferencias entre las estrategias que operan suprimiendo o ignorando algún elemento del fenómeno para hacer más sencilla su representación y las que

³⁵ Se entiende aquí *distorsión* de una manera cercana a su significado lexicográfico, como una alteración de los elementos del sistema real que se emplean en la representación.

³⁶ El ejemplo está tomado de Ladyman (2008: 359).

operan distorsionando algún aspecto de los fenómenos. Las abstracciones se emplean en las descripciones de los fenómenos con el propósito de dejar de lado determinadas propiedades que pertenecen al fenómeno modelado por ser consideradas irrelevantes, por ejemplo, en la descripción de la trayectoria de un móvil no se toma en cuenta como variable relevante el color del mismo. La introducción deliberada de variables o la modificación de los parámetros en los modelos tiene el propósito de facilitar los cálculos o ampliar la aplicabilidad de las conclusiones a un número mayor de fenómenos (véase Cassini, 2013: 356).

Uno de los principales escollos en el tratamiento filosófico de las idealizaciones en ciencia es el de explicar en qué casos las distorsiones pueden introducirse de manera legítima. Algunas perspectivas han clasificado a las idealizaciones con el propósito de mostrar las diferentes características que estas estrategias adquieren dependiendo de los propósitos con los que son empleadas y de explicar cómo sus diferentes usos se justifican en virtud de ciertos criterios pragmáticos. Aunque dentro en la filosofía de la ciencia no existen acuerdos extendidos sobre cómo definir la noción de idealización, las clasificaciones disponibles coinciden en algunos puntos importantes. Acuerdan fundamentalmente en que la idealización consiste en la supresión, mitigación, modificación o adulteración de algún elemento que contribuye causalmente al fenómeno o que perturba el comportamiento de un sistema. Coinciden, además, en que las idealizaciones se introducen con el propósito de construir un modelo que permita subsumir (en muchos casos, explicar) un conjunto de fenómenos particulares y en que es posible aproximar los modelos al comportamiento real de los fenómenos a través de procesos de des-idealización.

En un texto ya clásico de la filosofía de la ciencia, Cartwright (1983) examinó particularmente la función de los modelos como mediadores. Desde su punto de vista, los modelos dan contenido empírico a las teorías, haciendo posible su aplicación a los fenómenos. Los modelos permiten aplicar una teoría de alto nivel a un efecto particular o fenómeno específico. Cartwright distinguió entre las

idealizaciones que denominó *concretas* y las que llamó *ficcionales*. Las primeras se caracterizan por la manipulación teórica o experimental de circunstancias concretas para minimizar o eliminar ciertos aspectos. Este tipo de idealizaciones permite arribar a leyes aproximadamente verdaderas, de carácter empírico o fenomenológico. La idealización ficcional o abstracción, en cambio, procede eliminando las causas que interfieren en la producción de un fenómeno. Las leyes que se alcanzan por este procedimiento son leyes fundamentales.

Otra perspectiva ampliamente difundida sobre la idealización científica es la elaborada por McMullin (1985). Desde su punto de vista, las teorías se aplican a circunstancias muy especiales. Las idealizaciones involucradas en las descripciones teóricas proporcionan enunciados aproximadamente verdaderos que representan sistemas físicos. McMullin identifica en las estrategias empleadas por Galileo el origen de los métodos contemporáneos de idealización y distingue entre idealizaciones *constructivas* y *causales* (McMullin, 1985: 255). La diferencia entre estos dos tipos de estrategias depende de la manera en que procede la idealización: cuando se realiza sobre la representación conceptual se trata de una *idealización constructiva*, cuando la idealización se realiza sobre el problema mismo se trata de una *idealización causal*. La *idealización constructiva* procede introduciendo correcciones en la descripción teórica para acercarla al problema real, aunque la descripción de este permanece inalterada. La *idealización causal* simplifica las líneas causales presentes en la situación real. En esta clase de idealización el problema o la situación real es distorsionado para acercarlo a la descripción teórica. Dentro de este tipo de idealizaciones, se encuentran las descripciones de las situaciones o sistemas que se realizan solo en la imaginación, como en los experimentos mentales. En estos casos, los sistemas representados son idealizados y como resultado se obtiene una descripción simplificada del problema. Dado que esta simplificación se da solo en la imaginación, se trata para McMullin de un tipo de idealización subjuntiva. Las idealizaciones causales también se emplean en los experimentos reales, en los

cuales es posible minimizar los factores perturbadores o separar las líneas causales a través del aislamiento del sistema experimental.

Tanto la propuesta de McMullin, como la de Cartwright, suponen que el tipo de aproximación a los fenómenos que los modelos proporcionan puede ser completado mediante un proceso de des-idealización. Si bien los modelos se construyen con diferentes fines, por ejemplo, explicar un conjunto de fenómenos, la adición de nuevas variables presentes en el fenómeno real permite una mayor precisión en la representación. Asimismo estos procesos de ajuste pueden mejorar las predicciones sobre el comportamiento de los fenómenos.

Weisberg (2007), por su parte, extiende la caracterización de McMullin. La clasificación de las idealizaciones de Weisberg está centrada en la manera en que las idealizaciones están justificadas. Distingue entre tres clases de idealización que desempeñan distintos papeles en las tradiciones de investigación científica. Las *idealizaciones galileanas* son simplificaciones que se realizan con el propósito de hacer más tratables a los problemas. Desde este punto de vista, el avance en el poder computacional y en las técnicas matemáticas debe permitir des-idealizar, removiendo la distorsión y añadiendo detalles a los modelos (véase Weisberg, 2007: 643). Las *idealizaciones minimalistas* proceden representando solo los factores que marcan una diferencia en la ocurrencia y el carácter esencial del fenómeno en cuestión. Este tipo de idealización se justifica en virtud de su poder explicativo: permiten construir modelos mínimos de los aspectos repetibles y universalizables de un sistema. Por último, las *idealizaciones de modelos múltiples* consisten en la práctica de construir diferentes modelos de un mismo fenómeno, relacionados pero incompatibles entre sí, cada uno de los cuales hace afirmaciones distintas sobre la naturaleza y la estructura causal que origina el fenómeno en cuestión (Weisberg, 2007: 647). Esta clase de idealizaciones no se justifican a partir de la posibilidad de des-idealizar los modelos, como en el caso de la idealización galileana. Al abordar modelos altamente complejos la construcción de modelos múltiples se asocia a los diferentes objetivos

perseguidos por cada una de las representaciones: precisión, generalidad, simplicidad, entre otros.

Los tratamientos filosóficos sobre el papel que las idealizaciones desempeñan en los experimentos mentales refieren a las clasificaciones mencionadas. Reiss (2012) propone que en los experimentos mentales es posible identificar las idealizaciones *constructivas* y *causales*. Entre las *idealizaciones constructivas*, identifica supuestos falsos en la descripción teórica o aspectos intencionalmente no especificados en esta, como las superficies completamente lisas en la descripción del movimiento en un plano inclinado o la redondez perfecta de las esferas que se desplazan a través de éste. Las *idealizaciones causales* se emplean en estos experimentos con el propósito de traducir un problema complejo a uno más simple, tomando en consideración solo un pequeño número de las causas y de los factores perturbadores de la ocurrencia de los fenómenos. Este tipo de idealizaciones son de carácter subjuntivo en los experimentos mentales y, en algunos casos, involucran la descripción de situaciones altamente idealizadas o de contrafácticos en sentido fuerte. Este análisis comparte la intuición realista que los tratamientos anteriores tienen en común.

Desde esta perspectiva, las idealizaciones son legítimas cuando es posible des-idealizar el modelo en cuestión para obtener una mejor aproximación a los fenómenos reales. Según Reiss, al identificar un principio que posibilita la predicción y explicación del fenómeno simplificado, finalmente es posible la explicación (o predicción) del fenómeno original añadiendo las causas omitidas. La legitimidad del empleo de idealizaciones en los experimentos mentales está condicionada por la satisfacción de dos condiciones. Una condición *epistémica* que consiste en disponer de métodos confiables para aislar las líneas causales y una condición *empírica* que establece que las líneas causales identificadas en el experimento realicen una contribución sistemática al fenómeno en cuestión incluso cuando se añadan las causas omitidas.

Otra perspectiva con un enfoque similar en lo que refiere al papel de las idealizaciones en los experimentos mentales es la de Laymon (1991). En su análisis propone que un criterio relevante para distinguir usos legítimos e ilegítimos de las idealizaciones en los experimentos mentales es intentar responder a la cuestión acerca de si la naturaleza daría la misma respuesta que la situación planteada en el experimento mental. Hay dos maneras de justificar los supuestos que involucran idealizaciones en los experimentos mentales (entiéndase por ello, las simplificaciones, distorsiones, abstracciones que tienen lugar en la narrativa del experimento). Por un lado, se puede mostrar que existe una serie de refinamientos en los experimentos reales tales que, al aplicarlos, estos puedan aproximarse a los postulados por el experimento mental (Laymon 1991: 174). Por otro lado, la posibilidad de justificar las distorsiones introducidas en los experimentos mentales puede apoyarse en buenas teorías (ampliamente aceptadas) sobre las causas capaces de perturbar el resultado experimental. En otras palabras, la legitimidad de estos procedimientos se basa en el supuesto de que existe una serie convergente de experimentos que pueden acercarse asintóticamente al límite ideal.

El caso de estudio corresponde al experimento mental que Newton emplea para mostrar que la fuerza atractiva que gobierna el movimiento de los objetos en las proximidades de la Tierra, también explica el movimiento de la Luna. En su descripción, Newton imaginó el vuelo de una bala de cañón en torno a una Tierra que no rota para ilustrar que la fuerza de la gravedad era la fuerza clave para explicar el movimiento planetario. En ausencia de gravedad y de resistencia del aire, la bala de cañón se mueve en la línea en que fue disparada. Cuando la gravedad de la Tierra actúa sobre la bala de cañón, sigue un camino que depende de la altitud y de la velocidad inicial con que fue disparada. Si la velocidad es muy baja, simplemente caerá de nuevo a la Tierra. Según la velocidad inicial que se le imparta seguirá las curvas *A, B, C, D, E, F, G* (Figura 1) y si la velocidad inicial

alcanza un valor crítico el proyectil describirá un círculo o una elipse y volverá a la montaña desde la cual se disparó.³⁷

El experimento mental de Newton puede entenderse como una interpretación descriptiva de un fenómeno y, como tal, como un modelo que explota un trasfondo teórico relevante y que facilita el acceso perceptual e intelectual al fenómeno. De la representación idealizada de un fenómeno (el movimiento de los proyectiles en el vacío con diferentes velocidades iniciales), Newton deriva una serie de proposiciones aproximadamente verdaderas sobre el comportamiento de los fenómenos. El tipo de razonamiento que propone el modelo es del tipo analógico, supone que las conclusiones sobre el movimiento de los proyectiles son aplicables al movimiento de la Luna y otros objetos astronómicos. La idea fundamental de Newton en materia de mecánica celeste es la de la composición de dos fuerzas: la fuerza de gravedad que impulsa el planeta hacia el Sol y la fuerza centrífuga que se opone a ella. La razón por la cual los planetas se mueven en elipses es resultado de la interacción entre la gravedad y la fuerza centrífuga. La inercia de los cuerpos, junto con una sola fuerza atractiva, la gravedad, gobiernan a un mismo tiempo los movimientos celestes y los de los proyectiles. Newton identificó la órbita kepleriana de la Luna con la órbita galileana de un proyectil que constantemente cae hacia la Tierra pero que no puede llegar a ella. La representación que Newton propone en el experimento mental se propone integrar dos fenómenos (el movimiento de los proyectiles y el desplazamiento orbital de la Luna) a partir de una representación idealizada. En el experimento mental o la bala de cañón sigue un camino que depende de la altitud y de la velocidad inicial con la que fue disparada. Al aumentar la velocidad inicial eventualmente orbitará la tierra en una órbita circular que regresará al cañón. Una vez que la velocidad inicial excede la velocidad de escape la bala de cañón se moverá a lo largo de una parábola y finalmente una hipérbola lejos de la tierra. De

³⁷ *Principia*, Libro III, *El sistema del mundo*, 1728.

manera similar el movimiento orbital de la luna es explicado por la acción de la fuerza de atracción que en otras condiciones (si la velocidad de la luna fuese diferente) haría que cayese hacia la tierra o que se escapase. Siguiendo la clasificación de McMullin, pueden identificarse en este caso idealizaciones constructivas (superficies totalmente planas, esferas perfectas) e idealizaciones causales (como la hipótesis del vacío para eliminar la influencia del aire). El objetivo de la representación no son los fenómenos en sí mismos sino abstracciones realizadas a partir de los fenómenos.

5.6 Conclusiones

La interpretación de los experimentos mentales como modelos presenta una serie de ventajas explicativas respecto de las que los equiparan a los experimentos reales y a los argumentos teóricos. Fundamentalmente, proporciona una explicación plausible de cómo los experimentos mentales producen nuevo conocimiento y logra integrar esta exégesis a la teorización filosófica sobre otras formas de razonamiento científico. Permite elaborar, además, una interpretación consistente de su eficacia heurística, un elemento que ha caracterizado a esta clase de experimentos y que ha motivado explicaciones esotéricas de su funcionamiento.

La falta de intervención en el mundo natural es un rasgo que tanto la explicación “experimental” como la “argumental” reconocen como rasgo propio de los experimentos mentales. La manipulación de variables en la imaginación difiere de la intervención material en que la ocurrencia del resultado de dicha manipulación está determinada enteramente, en el experimento mental, por la descripción del aparato experimental, por las hipótesis a las que se apela en la narrativa y por el conocimiento de fondo de los experimentadores. Esta es una diferencia importante con los experimentos reales. Los resultados de los experimentos mentales están limitados por los aspectos representados en el

sistema idealizado. El tipo de razonamiento que los experimentos mentales propician hace posible inferir ciertas conclusiones acerca del comportamiento de los sistemas reales y bloquea otras. La descripción de un escenario imaginario, de un aparato experimental o de un mecanismo, establece las condiciones iniciales e identifica las variables relevantes. La pregunta planteada en el experimento mental “¿Qué pasaría si..?” implica que el experimentador se represente la situación descrita y responda cómo se comportaría el sistema representado en tales condiciones dado cierto marco conceptual. Este ejercicio permite derivar una “moraleja” o en términos de Bailer-Jones (2003), un “mensaje”, que es el resultado de interpretar la situación en cuestión empleando los elementos provistos para ello. Hemos mostrado que esta es típicamente una forma de razonamiento sustitutivo, diferente de las inferencias lógicas que pueden realizarse a partir de un conjunto de enunciados.

Los experimentos mentales en algunos casos se apoyan en la visualización de una secuencia espaciotemporal de eventos que puede ser representada en la mente del experimentador. Esta representación usualmente contribuye a captar con mayor celeridad las conclusiones que el experimento se propone extraer. Sin embargo, hemos señalado que el aspecto fundamental de esta descripción no es la representación visual sino el tipo de razonamiento analógico con otros fenómenos conocidos por el experimentador que dicha descripción debe propiciar. La narrativa del experimento debe permitir al lector recuperar los principios matemáticos y físicos relevantes para anticipar cómo evolucionará el sistema en las condiciones descritas por el experimento. Asimismo, los principios teóricos empleados en la narrativa deben posibilitar la realización de ciertas inferencias, por ejemplo, la aplicación de los mismos a fenómenos nuevos o que no existen en circunstancias reales.

Por otra parte, el poder heurístico de los experimentos mentales puede explicarse identificando las estrategias empleadas para su formulación. Las idealizaciones, cuyo uso es extendido en la construcción de modelos científicos,

permiten explicar la celeridad con la que es posible derivar conclusiones a partir de esta clase de ejercicio de la imaginación. Las idealizaciones introducidas en los experimentos mentales pueden justificarse, desde una perspectiva realista, vinculando las situaciones descritas en los escenarios hipotéticos con los fenómenos. En muchos experimentos mentales es posible aproximar los sistemas idealizados a los sistemas reales mediante progresivos procesos de des-idealización, o bien apelando a teorías ampliamente aceptadas sobre las causas que perturban, por ejemplo, la producción de un fenómeno. Comparar los modelos (o vehículos de la representación, en este caso, los experimentos mentales) con los fenómenos puede propiciar procesos de corrección y delinear criterios de legitimidad para establecer una distinción entre usos apropiados y espurios de los experimentos mentales.

A diferencia de las simulaciones computacionales, los experimentos mentales no emplean algoritmos formales. Aunque se asemejan a éstas en tanto se basan en la construcción de modelos dinámicos que muestran la evolución de un sistema físico, los experimentos mentales son epistémicamente transparentes.

En síntesis, la interpretación de los experimentos mentales en ciencia como un tipo de modelos permite especificar de qué manera estos se relacionan con los fenómenos. El tipo de relación que establecen con los sistemas reales es el de una *representación selectiva* de ciertos aspectos de los fenómenos. La noción de representación empleada no implica que los experimentos mentales reflejan o representan exhaustivamente los sistemas físicos sino que, a partir de la selección de algunos aspectos, permiten realizar inferencias cuyas conclusiones son aproximadamente verdaderas respecto del mundo natural.

CAPÍTULO 6

La epistemología de los experimentos mentales. Condiciones de adecuación

6.1 Introducción

No existe un consenso acerca de qué criterios podrían emplearse para distinguir los usos exitosos de los experimentos mentales de los usos espurios. Algunas perspectivas han especificado un conjunto de estrategias que pueden emplearse para legitimar los resultados de esta clase de experimentos. Estas estrategias están asociadas a la interpretación de los experimentos mentales como argumentos o como experimentos. Las posiciones argumentales (por ejemplo, Norton, 1996 y 2003) han sostenido que, dado que los experimentos mentales son asimilables a los argumentos, los criterios para identificar un experimento mental exitoso dependen de la validez³⁸ de la inferencia en cuestión y del valor de verdad de sus premisas. Las perspectivas experimentales (como la de Sorensen, 1992) han sostenido que es posible aplicar a los experimentos mentales algunos criterios tomados de la epistemología de los experimentos reales, como la replicabilidad o la confirmación independiente de los resultados.

Por otra parte, en el contexto del análisis filosófico de las simulaciones computacionales, Winsberg (2010) ha propuesto que la validación de las simulaciones es un proceso que no puede ser separado de otros procesos como la verificación (o comprobación matemática de los algoritmos), la ratificación de

³⁸ En el caso de los argumentos deductivos, es posible reconocer el uso de razonamientos válidos. En el caso de los argumentos inductivos, y en general de los razonamientos ampliativos, Norton se refiere vagamente a principios vinculados a inferencias inductivas, como la "preservación de la probabilidad" pero no especifica criterios para identificar inferencias ampliativas válidas en el caso de los experimentos mentales.

la fidelidad del modelo respecto de la teoría y la comparación con los resultados experimentales disponibles. Es la confluencia simultánea de todas estas estrategias lo que legitima la confianza en los resultados de las simulaciones. La propuesta de Winsberg podría ser útil para establecer condiciones de adecuación para los experimentos mentales. En el intento de establecer la eficacia de un experimento mental parece necesario evaluar conjuntamente los múltiples elementos de cuya adecuación depende la confiabilidad de sus resultados.

El objetivo de este capítulo es reunir y evaluar contribuciones sobre la epistemología de las prácticas experimentales, las simulaciones computacionales y los experimentos mentales que permitan elaborar condiciones de adecuación para estos últimos. Dado que, como se ha intentado mostrar, los experimentos mentales no son reducibles a argumentos teóricos y no son completamente asimilables a los experimentos reales, se argumenta que los criterios de validación desarrollados para estas prácticas no son directamente trasladables a la evaluación de los experimentos mentales. Procuraremos determinar un conjunto de condiciones de adecuación que pueden aplicarse a la evaluación de los experimentos mentales. Estas condiciones responden a la naturaleza *sui generis* de estos y hacen posible establecer criterios de confiabilidad para la descripción del escenario imaginario, su vinculación con el conocimiento empírico, su interpretación teórica y sus resultados.

6.2 La descripción de los escenarios imaginarios: condiciones iniciales y marco teórico

Los experimentos mentales comparten algunas características con los experimentos reales. Como los experimentos reales se formulan para dar respuesta a algún problema bien formulado sobre el comportamiento del mundo natural o sobre el alcance de un principio teórico. En ellos tiene lugar la manipulación planeada y controlada de determinadas variables que permite

mostrar cómo, en una situación imaginada, algunas variables son funcionalmente dependientes de otras. Al igual que los experimentos reales, los experimentos mentales dependen de hipótesis de fondo, de conocimiento tácito y de las teorías a partir de las cuales se evalúa el experimento. Estas características sugieren que tanto el escenario imaginario como el aparato experimental descrito deben satisfacer ciertos requisitos para no conducir a conclusiones espurias.

El primero de los requisitos que los experimentos mentales exitosos deben satisfacer es proporcionar una respuesta a las preguntas o problemas que se proponen responder. Se admite de manera más o menos generalizada que los experimentos se diseñan para evaluar el alcance o el valor de verdad de alguna hipótesis de manera independiente de las expectativas del experimentador. Dada estas características específicas de los experimentos, Alan Janis (1991: 113-118) distingue tres causas por las cuales los experimentos reales pueden fallar en dar una respuesta concluyente al interrogante a partir del cual se formularon. En primer lugar, es posible que los experimentos no puedan ser llevados a cabo exitosamente debido a alguna falla en el equipamiento o a causa de alguna influencia externa que interfiere con el funcionamiento apropiado del dispositivo experimental. En este sentido, los experimentos fallan porque no pueden producir ningún resultado. En segundo lugar, los experimentos pueden fallar porque arrojan un resultado incorrecto. Esto puede deberse también al funcionamiento inadecuado del equipamiento involucrado en el aparato experimental, ya sea porque el resultado está contaminado por influencias externas o porque no es una representación correcta de lo que el experimento se proponía revelar. En tercer lugar, los experimentos pueden fallar porque, aunque logren producir resultados correctos, éstos no constituyen respuestas a las preguntas que motivaron el su ejecución. Según Janis, los experimentos mentales pueden fallar de maneras similares a como los experimentos reales fallan.

Ampliando la sugerencia de Janis (1991), es posible identificar tres causas por las cuales los experimentos mentales pueden fallar en dar respuesta al

problema que motiva su realización. En primer lugar, los experimentos mentales pueden fallar a causa de la (in)competencia del experimentador. Un experimento mental puede no arrojar un resultado (en absoluto) si el experimentador no es capaz de llevar a cabo el experimento propuesto por la narrativa. Esto puede deberse a que carece del conocimiento de fondo requerido o a que no es capaz de inferir conclusiones desde el escenario imaginario y dado un marco nomológico. Es decir, un primer sentido en que el experimento mental puede fallar se debe a la incapacidad del experimentador de llevarlo a cabo y con ello, de dar alguna respuesta al problema que motiva la ejecución del experimento. A diferencia de lo que ocurre con los experimentos reales, esta falla no se atribuye al instrumental o a las condiciones de ejecución del experimento, y con ello a factores externos, sino al aparato conceptual, el conocimiento de fondo o las habilidades inferenciales del experimentador (dado que el experimento se lleva a cabo en el “laboratorio de la mente”, según la expresión de Brown, 2011: 1).

En segundo lugar, un experimento mental puede fracasar debido a que el experimento está descrito de manera incompleta. Esta clase de falla está asociada a la producción de resultados incorrectos. Esto sucede, por ejemplo, cuando el diseño del experimento no considera todos los supuestos teóricos relevantes o no toma en cuenta alguna variable que tendría alguna influencia en el efecto o fenómeno invocado. Aquí es posible diferenciar entre una descripción incompleta y una imprecisa. Una descripción incompleta no toma en cuenta todas las condiciones iniciales relevantes. Una descripción imprecisa puede ser ambigua u oscura pero no falla en considerar las variables o supuestos relevantes. Dentro de estas precisiones, se plantea la cuestión de en qué sentido una descripción del experimento es satisfactoria y qué tan exhaustiva debería ser la narrativa del caso. Aunque no es posible ser demasiado explícito respecto de estas condiciones, debido a que en cada caso dependerá del contexto en que se formule el experimento, del objetivo cognitivo del mismo y de la complejidad de su objeto, es deseable que la descripción del caso evite estos vicios, es decir, que

sea lo más completa posible y que no propicie interpretaciones múltiples o ambiguas.

Finalmente, un experimento mental puede fallar porque produce resultados que no son respuestas a las preguntas que inicialmente motivaron el experimento mental. En estos casos, la hipótesis que el experimento se proponía poner a prueba o explorar permanece indeterminada, en tanto no puede ser rechazada o confirmada en algún grado, aunque el escenario imaginario promueva la exploración de alguna hipótesis alternativa.

Una segunda condición que los experimentos mentales exitosos deben satisfacer es presentar claramente el escenario o aparato experimental a partir del cual se propone poner a prueba una hipótesis. Dado que los experimentos ordinarios consisten en el examen de una hipótesis a partir de la variación de una situación controlada en el contexto de una teoría específica, puede suponerse que, en tanto los experimentos mentales se basan en el método de la variación concomitante, deberían proceder esencialmente del mismo modo. Esto implica que el escenario imaginario en cuestión no debe estar *indeterminado* (*under-determined*), es decir que, por ejemplo, en el caso en el que el experimento mental intente socavar una hipótesis teórica, los experimentadores deben entender cómo la manipulación de un escenario se relaciona con la hipótesis bajo escrutinio. Elke Brendel (2004) sostiene que un experimento mental es un proceso de razonamiento llevado a cabo en el contexto de un escenario bien articulado para responder a una pregunta específica acerca de una situación no imaginaria. La confiabilidad de los experimentos mentales, como la de los reales, depende de que en una situación planeada y controlada sea posible mostrar cómo algunas variables son funcionalmente dependientes de otras sobre la base de algunas teorías o hipótesis de fondo (Brendel, 2004: 97). La condición de no indeterminación de Brendel (2004: 96) y la exigencia de que el conjunto las creencias jerarquizadas que son objeto del experimento sea aceptada por ambos interlocutores, son casi equivalentes a la condición impuesta por Popper:

Creo que el uso argumentativo de experimentos imaginarios es legítimo solo si las opiniones del oponente en el argumento se expresan con claridad, y si se cumple la regla de que las idealizaciones hechas deben ser concedidas por el oponente, o al menos aceptables para el oponente. (Popper [1959] 2002: 466)

Para Brendel (2005: 97) las intuiciones, como la percepción y la memoria, son fuentes inevitables pero falibles de evidencia. Es posible diferenciar entre experimentos mentales que evocan intuiciones legítimas y experimentos mentales que evocan intuiciones engañosas. Para garantizar que los experimentos mentales apelan a las intuiciones “correctas”, Brendel propone un criterio de acuerdo con el cual todas las variables relevantes para el experimento y sus correlaciones deben ser especificadas, de este modo se evita que el experimento funcione como una “caja negra”:

En particular, esto significa que si inventamos un escenario en el que manipulamos o cambiamos datos de manera desconocida, los efectos de estas manipulaciones o cambios siempre deben estar bajo control, es decir, debemos entender cómo pueden afectar a otros supuestos implícitos en el experimento mental y estos efectos aún pueden justificar la conclusión prevista por el experimento mental. (Brendel, 2005: 98)

De acuerdo con este criterio, los experimentos mentales están *indeterminados* cuando no es posible dar una explicación plausible de cómo evolucionaría el sistema en el escenario imaginario descrito. Esto puede ocurrir porque la descripción del experimento mental no especifica de qué manera ciertos aspectos del escenario se vinculan con otros. Brendel sugiere, además, que, en ciertos casos, la descripción de los experimentos mentales abunda en detalles que distraen la atención del experimentador y encubren la manera en que las conclusiones se siguen de las condiciones iniciales propuestas. La *indedeterminación* de los escenarios resulta, entonces, un criterio que permite detectar casos espurios.

Brendel considera que el experimento del rayo de luz de Einstein (1949: 53), que muestra que si fuera posible correr a la velocidad de la luz paralelamente a un

haz de luz ,el observador vería un campo oscilatorio estacionario, es un caso que especifica suficientemente la correlación de las variables y es por lo tanto un ejemplo exitoso. La hipótesis del éter estacionario como referencial absoluto para todo movimiento contra la que se dirige el experimento, no excluye la posibilidad del contrafáctico invocado por Einstein. La situación planteada podría ocurrir en el marco teórico de Maxwell. Por el contrario, el experimento de la caja de fotones de Einstein (Bohr [1930] 1949: 199-242) es un caso espurio. Esto se debe que la descripción del escenario desatiende el hecho de que la teoría de la relatividad supuesta en su formulación tiene como resultado una variación en el escenario que modificaría el resultado del experimento supuesto por Einstein. Más específicamente cuando el fotón sale del mecanismo, el resorte provoca un cambio en el campo gravitacional del reloj, por lo que se trataría de un caso de indeterminación del escenario imaginario. Finalmente, Brendel señala que otra razón para desestimar un experimento mental es que el escenario propuesto sea demasiado lejano a nuestras intuiciones empíricas y que, por lo tanto, estas no sean de utilidad para determinar, por ejemplo, si un caso cae bajo el alcance de un concepto o si un fenómeno es subsumido o es una excepción de una ley física.

Otra cuestión que se vincula con las características de la descripción del escenario imaginario es la relación de cercanía que el experimentador tiene con la situación hipotética. Es posible inferir que si el escenario hipotético resulta demasiado lejano a situaciones familiares, las respuestas a las preguntas planteadas por el experimento pueden resultar poco confiables. Algunas de las posiciones escépticas respecto del uso de experimentos mentales para apoyar hipótesis realizan observaciones similares. Según Ludwig Wittgenstein: "Consultar una tabla en la imaginación es tan poco consultar una tabla como la imagen del resultado de un experimento imaginado es el resultado de un experimento." (Wittgenstein, 1951: 293). Aunque él mismo se valió de experimentos mentales en la argumentación de algunas de sus tesis más célebres (considérese por ejemplo el uso que el autor hace en la obra citada del

experimento del escarabajo para argumentar la imposibilidad de concebir un lenguaje privado), no consideró que esta metodología pudiera resultar esclarecedora en el abordaje de problemas filosóficos. El escepticismo de Wittgenstein se funda en el hecho de que cuando los supuestos contrafácticos involucrados en la descripción del experimento nos llevan a mundos demasiado extraños, no es posible evaluar la plausibilidad de las hipótesis que resultan explicativas de fenómenos con los que no estamos familiarizados. Por lo cual, una manera de evadir el escepticismo respecto del funcionamiento de los experimentos mentales, consiste en establecer límites para los supuestos involucrados.

De manera similar, es preciso establecer alguna delimitación para el tipo de posibilidad que los experimentos mentales exitosos pueden explorar. Brendel sugiere que las hipótesis formuladas en el experimento mental no deben ser excluidas como imposibles por los supuestos teóricos en los que se fundamenta el análisis (Brendel, 2004: 105-106). Por su parte, Katherine Wilkes (1988) afirma que los experimentos mentales se caracterizan por postular una situación o un estado de cosas imaginario y preguntarse: “¿qué sucedería?” dadas ciertas condiciones impuestas por el marco teórico establecido. De la presentación de un escenario imaginario se sigue un argumento que propone una interpretación de la situación imaginada. Wilkes argumenta que para que un experimento mental sea epistemológicamente significativo todos los supuestos relevantes en la descripción del escenario imaginario deben ser *posibles* en un sentido *fuerte*. La que denomina *posibilidad teórica fuerte* (Wilkes, 1988: 18) se define en función de lo que podría o no suceder dado el estado actual del conocimiento científico. Aunque estas consideraciones pueden resultar útiles para descartar ciertos casos empleados en la argumentación filosófica, algunos experimentos mentales en física desafían la posibilidad física con el fin de evaluar la aplicabilidad de un principio teórico o mostrar la plausibilidad de una hipótesis. Incluso en los casos en los cuales es posible realizar un experimento real que se acerca a las

condiciones ideales propuestas por el experimento mental, de su calidad de contrafáctico depende su poder heurístico (por ejemplo, en los casos en los que el experimento supone superficies completamente lisas o movimiento sin fricción). Teniendo en cuenta estas razones, no resulta evidente en qué grado sería fructífero imponer una restricción a los contrafácticos apoyados por el escenario descrito en el experimento mental.

El Skaf (2017: 19) sostiene que la mayoría de los análisis epistemológicos de los experimentos mentales asumen la noción de posibilidad nomológica como criterio de confiabilidad. Esto quiere decir que cualquier experimento mental exitoso debe ser nomológicamente posible. No obstante, esta restricción parece ser demasiado estrecha ya que algunos experimentos mentales exitosos lo son a pesar de transgredir esta condición. De acuerdo con su análisis, un experimento mental científico es nomológicamente posible si su descripción especifica los particulares involucrados, su dinámica y los procesos de medición que deben realizarse bajo alguna teoría aplicable al escenario. Pero en algunos casos resulta de utilidad dejar abierta la posibilidad nomológica de los procesos involucrados en el escenario imaginario. En tanto la función cognitiva (qué es lo que se propone determinar el experimento) y la relevancia de su conclusión respecto de una hipótesis o principio teórico estén adecuadamente determinados, la descripción de los procesos (particulares y su dinámica) pueden estar nomológicamente indeterminados. Puede resultar de utilidad evaluar las consecuencias de procesos cuya posibilidad nomológica no esté claramente establecida para considerar las consecuencias de sus efectos respecto de un principio teórico. Este es para Skaf el caso del experimento demonio de Maxwell en el cual los detalles de cómo se llevaría a cabo el proceso de selección de las moléculas no es claramente posible desde un punto de vista nomológico. En la descripción del experimento, Maxwell deja indeterminado el proceso atribuyéndolo a un ser inteligente que deja pasar las moléculas más rápidas de aire frío dentro del compartimiento de aire caliente y las moléculas más lentas del compartimiento de aire caliente dentro del

compartimiento lleno de aire frío, pero explicita las consecuencias de este proceso para la segunda ley de la termodinámica (El Skaf, 2017: 28).

Aunque no es posible establecer una condición universal aplicable a la evaluación de todos los casos sobre los límites de la posibilidad nomológica en los experimentos mentales físicos legítimos, es plausible establecer un criterio general para evaluar la plausibilidad de los escenarios involucrados. Para determinar si el caso hace un uso legítimo de los contrafácticos sería preciso detallar el experimento en términos de su objetivo cognitivo, los procesos involucrados y su resultado de manera que sea evidente si las suposiciones contrafácticas que violan las leyes físicas son relevantes para la evaluación de una hipótesis en el sentido señalado por El Skaf (2017).

En síntesis, de estas consideraciones pueden derivarse una serie de criterios que la descripción de un experimento mental exitoso debería satisfacer. En primer lugar, el experimento mental debe ser capaz de dar respuesta a las preguntas que motivan su formulación. Si la narrativa del experimento no ofrece una respuesta definida o no es capaz de proporcionar una respuesta en absoluto a la pregunta que origina su realización, esta es una razón para considerarlo como ilegítimo. Debido a que existen varias razones por las cuales un experimento mental puede fallar (algunas de ellas involucran la destreza del experimentador que interpreta la narrativa y ejecuta el experimento), la posibilidad de que la narrativa de un experimento mental proporcione una respuesta satisfactoria a una pregunta está sujeta al cumplimiento de otras condiciones. No puede determinarse si un caso cumple con este criterio sin indagar las características de la descripción del escenario. Si el escenario está descrito de manera incompleta, porque falla en considerar las variables o supuestos teóricos relevantes, esta puede ser una razón por la cual el experimento en cuestión no arroje un resultado. Si el escenario está descrito de manera imprecisa, entonces, puede que dependa de la destreza del experimentador completar la información omitida o introducir los supuestos teóricos relevantes.

En segundo lugar, todas las variables relevantes para el experimento mental deben ser especificadas y debe estar claro en la narrativa cómo dichas variables se correlacionan y de qué modo las distorsiones, abstracciones e idealizaciones afectan esta correlación.

En tercer lugar, la narrativa del experimento debe evitar la extrañeza ya que es más difícil evaluar la corrección de las operaciones (funcionamiento adecuado de un mecanismo, introducción de modificaciones en la situación) que se llevan a cabo en contextos respecto de los cuales los experimentadores no tienen las intuiciones adecuadas (o no tienen ninguna intuición en absoluto). Esta restricción puede ser moderada, o relativizada, considerando el contexto en el que se formula el experimento mental. Un escenario hipotético puede resultar sugestivo aunque produzca respuestas incorrectas si se le presenta a un lego, mientras que un experto, por ejemplo, en física de partículas, es capaz de representarse correctamente lo que ocurriría en circunstancias extrañas para un no iniciado en esta disciplina.

En cuarto lugar, aunque no es posible delimitar el tipo de posibilidad que es lícito explorar en los experimentos mentales en física, es deseable que la transgresión de las leyes de la naturaleza esté justificada en función de algún objetivo cognitivo bien establecido. Algunos experimentos mentales aceptan un marco nomológico (relevante para la realización del experimento) pero suponen pequeñas variaciones en algunas leyes, por ejemplo, en el valor de la constante de la gravedad. En otros casos, se suponen escenarios en los cuales la conexión nómica entre los eventos no está suficientemente detallada y la evolución del escenario implicaría la violación de un principio aceptado, como ocurre en el caso del demonio de Maxwell. Por consiguiente, la condición sobre la restricción de la noción de posibilidad podría limitarse a señalar que, si se introducen hipótesis contrafácticas que violan leyes físicas, debe estar claro en qué sentido dichos supuestos son relevantes para la evaluación de una hipótesis o teoría. Una última condición, que podría resultar relevante dada la especificidad con la que se

propusieron las condiciones anteriores, es que, en caso de que el experimento mental se proponga como un contraejemplo a una hipótesis, la posición rival debe estar descrita de manera aceptable para quienes abogan por ella, de manera que se evite que el experimento mental se apoye en simplificaciones o caricaturizaciones (y que funcione de manera similar a la falacia del hombre de paja).

6.3 Extractores de intuiciones

Algunas de las posiciones que consideran que los poderes epistémicos de los experimentos mentales no se reducen a los de los argumentos teóricos, sostienen que estos experimentos justifican sus conclusiones apelando a algún tipo de intuición (empírica o racional) que subyace al sistema de creencias de los experimentadores. La conocida crítica de Daniel Dennett y Douglas Hofstadter al uso de los experimentos mentales en el contexto de la argumentación filosófica señala que estos en ocasiones funcionan como “extractores de intuiciones” (Hofstadter y Dennett, 1981: 375). De acuerdo con esta interpretación, algunos experimentos mentales están diseñados para centrar la atención de los experimentadores en algunos aspectos de la narrativa e ignorar otros y de este modo influirlos para que deriven ciertas conclusiones. Para Dennett (1995: 451) esta clase de experimento funciona de manera defectuosa cuando desorienta la imaginación en lugar de conducirla adecuadamente, por apoyarse en prejuicios o emplear el sentido común en contextos en los que no resulta adecuado hacerlo. El examen de algunos experimentos mentales que tratan acerca de dominios del mundo físico de los cuales no tenemos experiencias perceptivas, sugiere que en estos contextos la confiabilidad de los experimentos no puede descansar en el sentido común. La extrapolación de situaciones con las que estamos familiarizados a la interpretación del mundo microscópico puede resultar en la proliferación de representaciones inadecuadas que funcionan como “extractores

de intuiciones” y socavan la legitimidad de esta clase de experimentos mentales. Un ejemplo típico es el uso de los conceptos de partícula y de onda, tomados de la física clásica, que resultan inapropiados para describir los fenómenos cuánticos. Para evaluar esta tesis, es necesario precisar la crítica de Dennett y Hofstadter y especificar su aplicación al caso de los experimentos mentales en física.

Dennett y Hofstadter (1981) elaboran su objeción a la legitimidad de los experimentos mentales en el marco de la crítica a la noción de comprensión defendida por Searle (1980) en su conocido experimento mental de la Habitación China. Para estos autores, lo que Searle hace es introducir a los lectores en una ilusión que vuelve implausible la hipótesis de la inteligencia artificial fuerte y que al mismo tiempo los induce a creer que la intencionalidad es un rasgo fundamental de lo mental que no puede existir en máquinas programadas porque carecen de los poderes causales que tiene el cerebro. Desde esta perspectiva, el experimento de la Habitación China es un experimento mental filosófico espurio que tiene un efecto persuasivo que induce a los lectores a aceptar la conclusión de Searle por ser un extractor *de intuiciones*. La peculiaridad de los experimentos de este tipo es que son capaces de hacer proliferar representaciones que no son relevantes para contrastar tesis filosóficas.

Esta crítica podría trasladarse, con algunos matices, al caso de los experimentos mentales científicos. En algunos experimentos mentales, aparentemente efectivos, se introducen subrepticamente supuestos de cuya aceptación depende la interpretación de la situación experimental. En la descripción de un contrafáctico puede incluirse de manera solapada una interpretación teórica que hace más factible la aceptación de un resultado o la interpretación del mismo. Dichas interpretaciones subrepticias “gatillan” ciertas intuiciones e “inhiben” otras. Esta observación podría emplearse para objetar la utilidad de experimentos mentales que extrapolan consecuencias del mundo microscópico al mundo de los objetos macroscópicos de tamaño promedio. Por

ejemplo, el experimento del gato de Schrödinger (1935: 157) pretende mostrar las extrañas consecuencias que tiene el principio de superposición de estados si se aplica a un gato encerrado en una caja con un dispositivo que depende de la desintegración de un átomo radioactivo. El experimento muestra que en este escenario el gato está simultáneamente vivo y muerto como resultado de un evento que puede ocurrir o no. Esta conclusión, sugerida por la narrativa del caso, resulta de la extrapolación de las consecuencias del principio de superposición al mundo macroscópico. Aunque nunca podríamos “ver” a un gato en este estado, el experimento de Schrödinger supone que el estado de algunos objetos macroscópicos serán superposiciones de estados correspondientes a dos situaciones físicas diferentes.

Brendel (2005: 106) propone que una manera de evaluar si las conclusiones que propicia la narrativa de un experimento mental son aceptables es preguntarse si un análogo estructural del escenario representado tiene la misma plausibilidad intuitiva y arroja las mismas conclusiones. Resulta difícil establecer en casos como el mencionado en el párrafo anterior, en qué consistiría un análogo estructural del experimento mental. Dada la imposibilidad de realizarlo como experimento real, podría suponerse que se trataría de otro experimento mental que extrapole las características o el comportamiento del mundo microscópico a las características del mundo macroscópico. Aunque tampoco es claro de qué manera esta comparación podría arrojar luz sobre una situación que no tiene un análogo en el mundo macroscópico, ya que nunca se ha observado un objeto en superposición de estados.

En síntesis, algunos experimentos mentales funcionan como extractores de intuiciones cuando sus descripciones introducen supuestos teóricos que hacen plausible una determinada interpretación y socavan otras. Podría asimilarse estos casos a las falacias de petición de principio. Una manera de detectar si un experimento en cuestión funciona como un extractor de intuiciones es considerar el alcance y la plausibilidad de otras hipótesis relevantes (por

ejemplo hipótesis rivales a las que son objeto del experimento) en el escenario descrito por el experimento mental.

Otra cuestión que es relevante indagar es en qué consisten las intuiciones que los experimentos mentales explotan y cuál es su función. Una de las perspectivas inaugurales sobre esta cuestión es la de Mach (1897). Desde su punto de vista, las intuiciones empíricas constituyen una forma de conocimiento tácito impreso de forma inarticulada e inanalizada en los sujetos pero disponible a la inspección interna. Las intuiciones empíricas nos compelen a concebir al mundo físico de cierto modo, no dependen de nuestra voluntad de formarnos juicios sobre los fenómenos y constituyen una forma de conocimiento objetivo. La procedencia de las intuiciones es doble. Por un lado, se forman a partir de las experiencias pasadas del individuo; por otro lado, proceden filogenéticamente de la experiencia acumulada de la especie y como tales configuran la capacidad cognoscitiva de los individuos. Aunque no queda totalmente claro en la teoría de Mach, las intuiciones podrían interpretarse como contenidos mentales acompañados de un sentimiento de certeza que es compulsivo, ya que, debido a la fuerza persuasiva de las intuiciones, “captamos” un estado de cosas como posible o imposible. En los experimentos mentales el conocimiento tácito del mundo físico opera negativamente permitiendo “captar” de manera intuitiva qué estados de cosas serían imposibles porque entran en contradicción con la masa de experiencias acumuladas.

Varias perspectivas contemporáneas comparten el punto de vista de Mach sobre la naturaleza y función de las intuiciones en los experimentos mentales. Brendel (2005: 96) entiende a las intuiciones como parte del conocimiento de fondo en los experimentos mentales (junto con algunos principios teóricos) y considera que determinan inconscientemente nuestra interpretación. De acuerdo con esta interpretación, las intuiciones son actitudes proposicionales acompañadas de un sentimiento de certeza. Algunas intuiciones son relativamente estables y compartidas, lo que se debe al hecho de que

pertenece a la misma especie biológica y a comunidades culturales y científicas con algún conocimiento en común. No obstante estas especificaciones, la perspectiva de Brendel es ambigua. En algunos casos las intuiciones suscitadas por los experimentos mentales se asimilan a idealizaciones. Por ejemplo, en el experimento de Galileo de los cuerpos en caída libre, la abstracción de la forma y el color de los objetos y la atención solo en su peso es un elemento que la autora caracteriza como de carácter *intuitivo* en la descripción del experimento. En otros casos el carácter intuitivo se relaciona con la atribución de obviedad a supuestos implícitos. En el mismo ejemplo de Galileo, para la autora es obvio que los cuerpos en caída libre son arrojados en el vacío. En otros casos, la intuición es interpretada como un proceso de reorganización de experiencias ya presentes en el sistema de creencias del experimentador. Por ejemplo, la imaginar cuerpos de distinto peso en caída libre y la combinación de cuerpos de distinto peso es para la autora una operación que realizamos intuitivamente. En consecuencia, no está claro si la noción de intuición se refiere a ciertos contenidos mentales o a las propiedades de ciertos procesos mentales que se caracterizan por ser instantáneos y aparentemente no conscientes.

De acuerdo con Rowbottom (2014: 119-134), las intuiciones desempeñan un papel en los experimentos mentales científicos, donde algunos juicios son producto de un proceso que no consiste en un razonamiento consciente. La "materia prima" de estos procedimientos se deriva de la experiencia. Algunos escenarios imaginarios propician juicios intuitivos basados en la experiencia pasada. Los escenarios que propician estas intuiciones pueden ser descritos de una manera "neutral" o independiente de cualquier interpretación teórica, lo que sugiere que existe una base de intuiciones más o menos estable. A menudo, los mismos escenarios permiten construir argumentos diferentes dependiendo de principios teóricos distintos que dan un marco interpretativo al fenómeno en cuestión. Si bien Rowbottom examina el caso del plano inclinado de Stevin y el de la caja de fotones que aparece en la disputa entre Einstein y Bohr, no queda claro

que sea posible describir un escenario de una manera teóricamente neutral y que esta descripción origine dos interpretaciones contrapuestas. Si las intuiciones son juicios acerca de lo que ocurriría en ciertas circunstancias dado el conocimiento previo del experimentador, no resulta evidente que dichos juicios puedan formularse de manera independiente de todo supuesto teórico o incluso que sea posible describir dichos escenarios en un lenguaje puramente observacional, aceptable para representantes de teorías rivales como en el mencionado caso de la disputa Einstein-Bohr.

Aunque las perspectivas examinadas definen una noción general de intuición, no especifican qué diferencia a las intuiciones que desempeñan un papel en los experimentos mentales en física, de otro tipo de intuiciones, como las intuiciones morales. Resulta difícil determinar si las intuiciones físicas son más universales y estables que, por ejemplo, las intuiciones morales y en qué medida están determinadas por el conocimiento teórico de los experimentadores. Una contribución que se remite a la práctica científica para dar contenido a la noción de intuición es la de Tallant (2012). De acuerdo con este autor, es posible distinguir entre dos acepciones de la noción de intuición en el contexto de la física.³⁹ La posibilidad de que una teoría sea “visualizable” en el espacio y el tiempo formó parte de la discusión acerca de la legitimidad de la teoría cuántica. Este es para Tallant el sentido más común en el que la noción de intuición se emplea en física. Las acepciones que el autor señala aparecen en las publicaciones actuales de física con mayor factor de impacto, reflejan que los usos que pueden atribuirse a este término en la práctica científica actual se refieren a una familia de sentidos asociados. Un resultado experimental descrito como *intuitivo* es un resultado esperable y, por lo tanto, valorado positivamente. Una teoría *intuitiva* es una teoría que proporciona o permite derivar una explicación de un fenómeno. Una imagen

³⁹ Tallant (2012: 2960) menciona un sentido general en el que puede entenderse la noción de intuición, que opera tanto en geometría como en matemática y física, de acuerdo con el cual se consideran intuitivos u obviamente verdaderos a los puntos de partida de una teoría (proposiciones o supuestos).

intuitiva, remite a una representación adecuada de los fenómenos, adecuada respecto de los objetivos del usuario. Dada la ubicuidad de las representaciones visuales en la física, la noción de intuición aparece regularmente asociada con la capacidad de proporcionar modelos visualizables de los fenómenos representados. Una teoría cuya exposición propicia un entendimiento *intuitivo* es una teoría virtuosa. En el mismo sentido, los físicos sostienen que las intuiciones se construyen o se suscitan en física a través de la presentación de imágenes y modelos visuales. Este proceso consiste en generar un sentimiento de familiaridad con la representación a partir de la presentación de diagramas y otras representaciones visuales. Aunque Tallant encuentra un aire de familia en los usos de la noción de intuición en las publicaciones actuales en física, parece subsistir una diversidad de significados. En algunas ocasiones, se adjetiva a una teoría como intuitiva debido a su poder explicativo. En otros casos, el adjetivo se asocia a la cualidad de ser visualizable de una teoría. Quizás el segundo sentido se vincule más directamente con las intuiciones sobre las que algunos autores afirman que descansan los experimentos mentales, pero esta aproximación es un tanto incierta.

Para examinar del funcionamiento de los experimentos mentales no resulta imprescindible contar con una definición precisa de la noción de intuición. Este término podría sustituirse por otras nociones como *conocimiento de fondo* o *supuestos* que a fines prácticos permiten señalar los mismos elementos. No obstante, es posible realizar algunas aclaraciones sobre este concepto, muy usado en la literatura de los experimentos mentales. El uso del término “intuición” en la epistemología de estos experimentos parece estar asociado a dos rasgos salientes. En primer lugar, a su origen. Las intuiciones que tienen lugar en los experimentos mentales se originan en la experiencia y se asocian a actividades de visualización. Asimismo se vinculan a la vivencia personal de los sujetos, a la manera en que efectivamente estos han experimentado el mundo, a la extensión de sus conceptos empíricos. En general, la noción de intuición se vincula con el

acervo de conocimiento empírico perceptivo. En segundo lugar, las intuiciones han sido asociadas a juicios instantáneos y compulsivos. Dado que las intuiciones constriñen la estructura a partir de la cual los sujetos conocen el mundo, en algún sentido también restringen aquello que éstos pueden considerar como físicamente posible. De manera que, además de aportar el contenido empírico de las premisas de los experimentos mentales, las intuiciones proporcionan el trasfondo con el que se evalúan las hipótesis consideradas.

Otro rasgo de las intuiciones empleadas en los experimentos mentales es que, aunque parecen gozar de cierta estabilidad, se construyen históricamente y, por lo tanto, son falibles y pueden ser modificadas, corregidas e incluso creadas. Esto da noticia de su dependencia de los conceptos teóricos. En este sentido, Kuhn señala que la interpretación galileana del movimiento del péndulo, que implica una transformación de la visión aristotélica de este fenómeno, fue posible debido a que Galileo fue educado para interpretar los movimientos de acuerdo al paradigma tardomedieval del *impetus*. En este punto de vista se basó la interpretación de Galileo e hizo posible la transición de paradigma:

Hasta que no se inventó dicho paradigma escolástico, el científico no podía ver péndulos sino piedras oscilantes. Los péndulos fueron engendrados por algo muy similar a un cambio de *Gestalt* introducido por el paradigma. (Kuhn [1962] 2004: 188-189).

Aunque a partir de estas consideraciones es posible descartar que las intuiciones sean entidades o procesos substantivamente diferentes de las creencias que tienen origen en la experiencia, subsiste la cuestión sobre el papel que pueden desempeñar legítimamente en los experimentos mentales. Ante todo, podría pensarse que los experimentos mentales serán exitosos en contextos específicos en los que los experimentadores comparten un conjunto suficiente de valores epistémicos y creencias. Dentro de estas tramas, las intuiciones pueden entenderse como sesgos que pueden funcionar desestimando posibilidades empíricas. Si las tesis se evalúan solo a partir de ciertos consensos, entonces, las

mismas pueden resultar aceptables o no en virtud de las disposiciones previas a aceptarlas o rechazarlas.

Si bien no es posible dar una respuesta definitiva a la cuestión de cuándo las intuiciones son invocadas de manera legítima, es factible establecer algunos criterios de corrección para los experimentos mentales y, con ello, fijar condiciones para la evaluación de una hipótesis. La interpretación que a partir de estas consideraciones parece razonable asumir es que las intuiciones o el conocimiento empírico de fondo solo tienen un valor evidencial relativo a contextos específicos. Las intuiciones solo pueden ser desafiadas o modificadas en casos en los cuales se comparta conocimiento de fondo y a partir de su presentación se descubran alternativas inesperadas.

No obstante, el conocimiento de fondo y las intuiciones desempeñan roles precisos en la epistemología de los experimentos mentales. De la contemplación del escenario hipotético y la realización de las operaciones indicadas en la narrativa, los experimentadores formulan un juicio sobre lo que ocurriría en ciertas circunstancias. Las intuiciones activadas por la narrativa pueden hacer plausibles algunas de las premisas de la interpretación que quien formula un experimento mental se propone sugerir. En este sentido, la interpretación apoyada por el autor se “filtra” desde las intuiciones. De manera análoga las intuiciones pueden contribuir a exhibir contradicciones: el escenario descrito en la narrativa del experimento suscita ciertas intuiciones que entran en contradicción con la hipótesis. Las intuiciones pueden, además, facilitar la presentación de entimemas: en la narrativa de un experimento mental puede esconderse un argumento en el cual se omiten premisas altamente plausibles. En estos casos, es necesario evaluar la aplicabilidad de las hipótesis a los escenarios descritos por los experimentos mentales. Por otra parte la presentación narrativa de un caso ilustrativo de una teoría puede activar esquemas conceptuales que permanecen

ocultos.⁴⁰ La rareza de algunos escenarios hipotéticos permite tomar distancia de posicionamientos e interpretaciones automatizadas por los lectores.

6.4 Modelado y representación

Winsberg (2010) desarrolla una serie de criterios epistemológicos para evaluar las simulaciones computacionales basándose en una similitud que considera fundamental entre estas y los experimentos reales: la relativa autonomía de ambas prácticas respecto de las teorías. Sostiene que, como sucede con los experimentos ordinarios, se han desarrollado una serie de estrategias para la validación de los resultados de las simulaciones computacionales. Algunas de estas estrategias son directamente análogas a las propuestas por Franklin (1986) para los experimentos reales. Entre las estrategias de validación para las simulaciones se encuentran: la eliminación de fuentes de error, la confirmación independiente y el empleo de los resultados mismos para justificar su validez.

Para Winsberg, las simulaciones computacionales comienzan con un modelo matemático que representa la evolución espaciotemporal de un sistema estudiado en términos de ecuaciones o reglas de evolución de las variables del modelo. La construcción de un modelo, o al menos una versión “idealizada” del proceso de su construcción, comprende, según Winsberg, (2010: 11), a) una teoría bien confirmada y un modelo sugerido por la teoría para interpretar un sistema, donde el modelo consiste, a su vez, en una descripción abstracta del sistema y algunas ecuaciones diferenciales; b) un tratamiento del modelo que consiste en la asignación de valores iniciales a las variables y parámetros del modelo; c) un “solucionador” que combina el modelo con el tratamiento para crear un algoritmo computable diseñado para ser un sustituto aproximado, discreto, de las

⁴⁰ Esta función se sigue parcialmente de los desarrollos de Gendler (2007).

ecuaciones diferenciales continuas del modelo; y d) los resultados que son producidos cuando el “solucionador” se corre en una computadora. Las técnicas de discretización de las ecuaciones del modelo se emplean para transformar funciones continuas en sus contrapartes discretas. Las expresiones algebraicas resultado de este proceso pueden ser calculadas paso a paso por una computadora. Este es un proceso de transformación que constituye una aproximación a la que se asocian estrategias que permiten reducir el error, por ejemplo, empleando intervalos de tiempo y espacio muy pequeños.

La construcción de algoritmos computacionales es compleja y no es posible ni conveniente para nuestros propósitos abordarla en este apartado. Basta con señalar que en vistas a la tratabilidad matemática y computacional del modelo de simulación, en la construcción de los algoritmos se realizan simplificaciones de los modelos derivados de la teoría (como ignorar factores), se reducen grados de libertad, se realizan suposiciones de simetría a veces poco realistas y se incorporan ecuaciones diseñadas para computar algún efecto del mundo natural dejado de lado (Winsberg, 2010: 12). La introducción de estas ecuaciones, que no tienen su contraparte en el modelo teórico, permite que la simulación produzca resultados más fieles que los que se hubieran obtenido de la consideración del efecto físico. Las observaciones de Winsberg sobre la complejidad de los procesos involucrados en la construcción de una simulación computacional y sobre las estrategias empleadas para validarla, pueden iluminar el estudio de los criterios de confiabilidad de los experimentos mentales.

La simulación computacional es una metodología extendida en la investigación científica desde finales siglo XX. Aunque estas estrategias han tenido un papel fundamental en el contexto de aplicación de las teorías científicas, se han empleado también para explorar nuevos dominios fenoménicos y controlar otros experimentos. Como en el caso de los experimentos mentales, los filósofos han sido reticentes a asignarles un papel en el contexto de justificación y se ha

puesto en duda en qué sentido las simulaciones pueden ser usadas para extender el conocimiento del mundo físico.

El problema epistemológico de las simulaciones computacionales puede ponerse en paralelo al de los experimentos mentales. El examen filosófico de su funcionamiento se ha centrado en su relación con la experimentación real, sus relaciones con las teorías, la función de los supuestos falsos en la construcción de simulaciones y el establecimiento de criterios de confiabilidad para su uso. Algunos autores han afirmado que los experimentos mentales son los ancestros de las simulaciones computacionales. Esta es la sugerencia de Chandrasekaran y Nerssesian (2012). Aunque subsisten diferencias importantes, especialmente en la ausencia de algoritmos formales en los experimentos mentales, hay analogías positivas importantes entre estas prácticas. Ambas dependen de la representación de la evolución de un sistema físico y emplean para ello estrategias de modelización que, aunque difieren cualitativamente, desempeñan el mismo papel: hacer tratable un problema en el que la consideración de todas las variables relevantes sería inasequible.

En el caso de algunos experimentos mentales, las idealizaciones pueden interpretarse como abstracciones de los sistemas reales. Las predicciones que los experimentos mentales permiten realizar pueden “ajustarse” al comportamiento de los sistemas reales mediante un proceso de desidealización. Las maneras de vincular las estrategias de idealización con los sistemas reales pueden considerarse en algún sentido análogas a las estrategias empleadas para evaluar la legitimidad de las transformaciones del modelo físico introducidas para construir los algoritmos computables. Aunque en estos casos, para hacer tratables a las ecuaciones, la construcción del modelo no solo implica realizar simplificaciones y abstracciones, sino también incorporar ecuaciones diseñadas para computar algún efecto del mundo natural que no está presente en el modelo teórico.

Pueden encontrarse otras analogías positivas entre las estrategias de validación de las simulaciones computacionales y los criterios de confiabilidad de los experimentos mentales. Algunos mecanismos para validar los resultados de las simulaciones computacionales consisten en la comparación de esos resultados con los datos disponibles provenientes de la experiencia, es decir, de la observación y medición de los fenómenos. El mismo procedimiento puede emplearse para la corroboración de las predicciones de los experimentos mentales.

Cuando las simulaciones se emplean en dominios en los cuales la información empírica es escasa, surge la cuestión de cómo determinar si sus resultados proporcionan información confiable acerca de los fenómenos que se proponen investigar. En estos casos, la validación de las simulaciones computacionales consiste en determinar si el modelo elegido representa adecuadamente el fenómeno que se propone simular. Este procedimiento difiere de los involucrados en la verificación, en los que se intenta determinar si el modelo computacional proporciona una solución aproximada de las ecuaciones matemáticas del modelo teórico inicial (Oberkampf y Roy, 2010: 25). De manera análoga a lo que ocurre con la experimentación real, donde los científicos emplean diferentes técnicas (como la calibración y la confirmación independiente) para justificar racionalmente sus resultados, se emplean una variedad de estrategias para justificar la confiabilidad de las simulaciones. Entre éstas pueden señalarse: el empleo de métodos de análisis y teorías matemáticas confiables, la comparación de los resultados de la simulación con los de otras simulaciones del mismo fenómeno que emplean diferentes algoritmos (si los hubiera) y la comparación con los datos disponibles sobre el dominio de fenómenos simulado.

Es razonable suponer que en dominios en los cuales la experimentación real es imposible o extremadamente difícil, los resultados de los experimentos mentales deben ser comparados con los resultados de otros experimentos mentales sobre el mismo fenómeno, cotejados con datos empíricos relevantes, e

incluso vinculados con los resultados de simulaciones computacionales concernientes a estos fenómenos.

Para Winsberg (2010), la validación de las simulaciones es un proceso que no puede ser separado de la verificación, la ratificación de la fidelidad del modelo respecto de la teoría, el rigor matemático, la intuición física y la comparación con los resultados experimentales. Es la confluencia simultánea de todos estos factores lo que legitima la confianza en los resultados de las simulaciones. Por lo que el éxito de las simulaciones es alcanzado mediante un proceso de avance y retroceso, una especie de proceso de ensayo y error. Las fallas en el modelo no pueden evaluarse aisladamente ya que no pueden atribuirse con facilidad a un factor identificable. Pueden deberse a un error en la selección de las ecuaciones, en las operaciones de discretización o en la construcción del modelo computacional. Por esta razón, Winsberg propone que las fallas en las simulaciones deben ser evaluadas holísticamente. En otras palabras, la naturaleza abigarrada y compleja de las simulaciones sugiere una epistemología *sui generis* con criterios de confiabilidad que deriven de una comprensión apropiada de su naturaleza.

En suma, consideramos que, como ocurre en el caso de las simulaciones, la evaluación de la confiabilidad de los experimentos mentales también implica evaluar de manera holística las diferentes operaciones que se presentan en el experimento mental. Además de poner en relación el resultado del experimento con resultados experimentales y observaciones relevantes, es deseable considerar a la luz del conocimiento presupuesto la validez de las idealizaciones y las representaciones visuales (gráficos y diagramas que en ocasiones acompañan el experimento). De manera análoga, el fracaso de un experimento mental no puede adjudicarse de manera simple y directa a uno de sus elementos componentes. Tanto la descripción del escenario como las hipótesis teóricas empleadas en su interpretación deberían revisarse, pero no hay una manera única

de hacerlo. El error podría encontrarse en presupuestos ocultos que no se tuvieron en cuenta en el momento de diseñar el experimento.

6.5 Interpretaciones teóricas

Una de las razones que Bishop (1999) presenta para argüir que los experimentos mentales no son argumentos es el hecho de que el mismo experimento mental puede dar lugar a diferentes interpretaciones. Mientras que las instancias del mismo tipo de argumento deben tener conclusiones idénticas, diferentes pensadores pueden sacar conclusiones diferentes del mismo experimento mental en diferentes ocasiones. Esta lectura apela a varios episodios de la historia de la ciencia en los cuales el mismo escenario, o uno muy similar, fue empleado para ilustrar, justificar o incluso refutar teorías rivales. El experimento de la caja de fotones, propuesto por Einstein y replicado por Bohr (1949: 199-242), es el ejemplo clásico de interpretaciones teóricas incompatibles para el mismo experimento mental.

Esta característica de los experimentos mentales representa una dificultad para considerar su uso como experimentos cruciales. Entre los casos de experimentos mentales que originan a interpretaciones diferentes podrían considerarse dos posibilidades: a) que asumiendo las mismas condiciones iniciales dos puntos de vista teóricos diferentes den lugar a interpretaciones diferentes sobre la evolución del sistema y b) que el mismo escenario imaginario se desarrolle en un único curso o dé un resultado unívoco pero que este resultado sea interpretado como evidencia a favor de puntos de vista rivales. El primer caso podría considerarse análogo a lo que sucede con la interpretación estándar de la mecánica cuántica y la teoría de Bohm cuando desde sendos presupuestos teóricos se describe el comportamiento de electrones en el experimento de la doble rendija. Mientras que la teoría de Bohm predice trayectorias definidas por lo que posición final de una partícula está determinada por su posición inicial, para

la interpretación de Copenhague las partículas no poseen una localización en el espacio, ni, por tanto, trayectorias definidas, antes de ser detectadas. En el segundo caso, el experimento tiene una única solución o resultado pero interpretaciones teóricas diferentes lo consideran evidencia a favor de hipótesis diferentes. Algunos autores han expresado este hecho señalando que es posible derivar diferentes “moralejas” del mismo experimento. El segundo caso podría ilustrarse con las interpretaciones ptolemaica y copernicana del experimento de la torre. El mismo resultado, que un objeto arrojado desde lo alto de una torre cae al pie de esta, es interpretado por Ptolomeo, como evidencia de que la Tierra no se mueve (de lo contrario caería más allá de éste debido a que en el desplazamiento de la Tierra la torre cambia su posición) y por Galileo como evidencia de que la Tierra se mueve (porque la caída es un movimiento compuesto en línea recta y circular, que acompaña el movimiento de la Tierra).

Una de las dificultades para evaluar las diferentes interpretaciones rivales de un mismo experimento mental reside en distinguir adecuadamente los presupuestos del experimento. Si bien es posible formular algunas de las hipótesis presupuestas que operan en el análisis de un experimento, siempre es posible “excavar” más profundo y explicitar premisas de orden teórico, ontológico o metodológico que subyacen a dichas hipótesis. Por consiguiente, en lo que refiere al trasfondo teórico de un experimento mental, no es evidente en cuál de las “capas” del conocimiento de fondo es necesario detenerse para hacer conspicua una interpretación. Por otra parte, aunque la descripción del dispositivo experimental o de la situación hipotética se formule en un vocabulario observacional, esta nunca es teóricamente neutral. En suma, la cuestión de cuanta especificación de los supuestos teóricos sea necesaria para evaluar la confiabilidad de un experimento mental no puede establecerse *a priori* de manera unívoca sino que debe ser considerada caso a caso.

Por otra parte, algunas perspectivas afirman que la confiabilidad de los experimentos mentales depende también del tipo de inferencias involucradas en

la “moraleja” que se deriva del experimento. Este es otro criterio para evaluar interpretaciones rivales del mismo experimento mental. Las contribuciones de Sorensen (1992) y Häggqvist (1996) aportan algunos elementos para evaluar los argumentos involucrados en un experimento mental. Para estos puntos de vista, es preciso reconstruir las premisas que se infieren de la ejecución del experimento mental y ponerlas en relación con los principios teóricos y con un trasfondo de presupuestos que constituyen el conocimiento de fondo de los experimentos. Dichas reconstrucciones involucran la introducción de operadores modales. Por lo general, los argumentos vinculados a los experimentos mentales contienen una premisa que implica que, dados ciertos presupuestos que pueden ser especificados como leyes científicas o creencias relativas al comportamiento de determinados sistemas físicos consideradas verdaderas por los experimentadores, se concluye que un enunciado considerado necesariamente verdadero es o bien contingente o bien imposible.

Si bien estos aportes permitirían formalizar (o elaborar una reconstrucción semi-formal de) un experimento mental, suponen que todos los casos funcionan como contraejemplos que ponen en duda la verdad o el alcance de un supuesto teórico. Ambas perspectivas consideran que los experimentos mentales eficaces son los que permiten inferir, dado el conjunto de creencias del experimentador expresable en un conjunto de enunciados, que ciertos escenarios (cuya plausibilidad depende también del acervo de creencias del sujeto que lleva a cabo el experimento) constituyen anomalías para una hipótesis. Sin embargo, no todos los experimentos mentales funcionan como falsadores de supuestos empíricos o teóricos, por lo que la reconstrucción propuesta solo puede ser efectiva para evaluar interpretaciones rivales en estos casos específicos.

En síntesis, la formulación y especificación de los supuestos teóricos y empíricos de los que depende una interpretación teórica facilitan la evaluación de las interpretaciones rivales de los experimentos mentales. Por otra parte si bien no puede establecerse un esquema único de reconstrucción de los argumentos

que permiten arribar a diferentes interpretaciones, este trabajo de reformulación permite esclarecer de qué manera se infiere determinada interpretación. En algunos casos, los experimentos mentales pueden ser reconstruidos como instancias del *modus tollens*, específicamente cuando el escenario descrito en la narrativa se propone como contraejemplo de una hipótesis. En esas ocasiones, es de utilidad tomar en cuenta las sugerencias de Sorensen y Häggqvist para explicitar si el caso se propone minar una suposición teórica mostrando su falsedad o su alcance, es decir, si se propone refutarla (mostrando que es falsa) o si se propone limitar su alcance (mostrando, por ejemplo, que no es necesariamente verdadera para todos los casos que se propone subsumir). En otros casos, las interpretaciones rivales podrían ser evaluadas formulado los argumentos como instancias de inferencias a la mejor explicación en los que la interpretación teórica resulta de un argumento abductivo⁴¹ en el cual una interpretación es sugerida como la mejor explicación posible para lo que ocurre en un escenario hipotético. En otros ejemplos, puede considerarse si alguna generalización puede sostenerse con cierto grado de probabilidad. En estos casos, deberá evaluarse si el experimento proporciona evidencia suficiente para concluir que la interpretación puede aplicarse a un conjunto de instancias similares.

6.6 Los resultados experimentales

Dado que algunos experimentos mentales fueron posteriormente realizados como experimentos reales y que en ocasiones su falta de ejecución se

⁴¹ Sobre esta sugerencia, hemos reconstruido el experimento mental de los cuerpos en caída libre de Galileo: “En el caso de Galileo, por ejemplo, el salto hacia la proposición que sostiene que *la velocidad de la caída es una constante, y no depende del peso y la composición de los cuerpos* no puede darse inductivamente. Deductivamente, en cambio, cualquier proposición puede ser inferida. Pero, ¿por qué inferir esa en particular? Afirmamos que el principio de equivalencia constituye la hipótesis que mejor explica tanto la falsedad de los principios aristotélicos destruidos por la *reductio* como los fenómenos visualizados en el experimento.” (Borge y Mettini, 2018: 355).

debe a imposibilidades técnicas, es preciso evaluar qué papel pueden desempeñar las confirmaciones experimentales en la constatación de las predicciones realizadas por los experimentos mentales. En vistas a este propósito, es preciso reexaminar la relación que hemos afirmado que existe entre experimentos mentales y reales. Una cuestión que se plantea aquí es la de distinguir claramente el criterio de identidad del criterio de confiabilidad de los experimentos mentales.

Se ha argumentado que los experimentos mentales y los experimentos reales no son el mismo tipo de práctica pero que, no obstante, pueden cumplir las mismas funciones. Los experimentos mentales son más parecidos a los modelos que a los experimentos reales porque comparten más rasgos con éstos y se distingue de la experimentación real el elemento que es crucial a estas prácticas, esto es, la intervención en el mundo natural. Como ya señalamos, en los experimentos mentales se describe un escenario imaginario que representa el comportamiento de un sistema físico y, en ocasiones, también algún dispositivo o mecanismo con las instrucciones que especifican las operaciones que el experimentador puede realizar. Esta descripción, enmarcada en un conjunto de supuestos teóricos, da lugar a la representación en la imaginación de la evolución del sistema y a la obtención de un resultado que posteriormente es interpretado teóricamente dentro de la narrativa del experimento. La descripción del escenario imaginario a veces implica la transgresión de leyes naturales, pero esta no es una característica esencial de todos los experimentos mentales.

Una diferencia importante puede señalarse entre la descripción narrativa de un experimento mental que posibilita una representación o realización imaginaria del mismo y un experimento real. La diferencia relevante consiste en que la realización material no forma parte del diseño de un experimento mental, algo que necesariamente sucede con el diseño de un experimento real. El experimento mental es concebido como tal y, por lo tanto, se narra para ser llevado a cabo enteramente en la imaginación. Por consiguiente, la descripción de los escenarios

hipotéticos y la realización puramente hipotética del experimento son criterios que permiten identificar a los experimentos mentales y distinguirlos de los reales. Esta caracterización de los experimentos mentales nos permite distinguir aquellos realizados como experimentos reales de sus contrapartes diseñadas para llevarse a cabo en la mente de los experimentadores. Si bien los resultados sacan a la luz la misma conexión causal que constituye una respuesta específica a la misma pregunta teórica, lo hacen utilizando estrategias diferentes y empleando mecanismos de conocimiento diferentes.

Por otra parte, aunque algunos casos son realizables como experimentos reales esto no implica que los experimentos mentales sean epistemológicamente superfluos. Los resultados de algunos experimentos pueden derivarse igualmente de las leyes pertinentes y las condiciones iniciales por lo que algunos autores consideran que en estos casos la ejecución de estos experimentos es redundante. Pero esto no implica necesariamente que su valor epistémico sea nulo. Consideremos nuevamente el experimento que Galileo ([1632] 1967: 186-188) formula para mostrar cómo se comportan los marcos de referencia inerciales. La ventaja epistémica del experimento consiste en proporcionar una interpretación consistente de un conjunto de fenómenos y mostrar su correlación con otros fenómenos comúnmente experimentados. Aunque la ejecución del experimento mental no comporta ninguna ganancia si se compara con una inferencia que puede realizarse tomando las leyes y la descripción de ciertos hechos como premisas, el hecho de que el escenario apele a experiencias ordinarias para apoyar una hipótesis y para negar que el fenómeno en cuestión sea evidencia a favor de la hipótesis aristotélica es un elemento persuasivo adicional.

Algunas perspectivas sostienen que la realización material no es un elemento que permita diferenciar entre experimentos mentales y reales si estos son considerados en función de sus objetivos cognitivos. Desde este punto de vista, si se considera que ambas prácticas coinciden en plantear el mismo tipo de pregunta, no existe una diferencia sustancial entre ambos, por lo que podrían

emplearse los mismos criterios para evaluar la confiabilidad de ambas prácticas. Buzzoni (2010: 13) propone que lo que denomina “viabilidad técnico práctica” es un criterio decisivo para admitir a un experimento mental como genuino. Un experimento mental puede ser criticado si se cree que su realización experimental es imposible, ya sea porque contradice observaciones o leyes científicas sobre las cuales existe una certeza razonable, o bien porque contiene contradicciones que hacen imposible cualquier realización *a priori*.

El tipo de viabilidad a la que refiere Buzzoni, puede considerarse como un criterio de *realizabilidad en principio*. En consecuencia, no es necesario adherir a la idea de que los experimentos mentales y reales son un mismo tipo de prácticas para adoptar la *realizabilidad en principio* como criterio de confiabilidad para los experimentos mentales. Algunos experimentos mentales cuya realizabilidad empírica es imposible contribuyen a tomar en cuenta algunas alternativas relevantes para establecer un principio teórico. Este es el caso del experimento mental del ascensor que Einstein (Einstein e Infeld, 1938: 53) emplea para mostrar que si no hay diferencias observables, no se puede distinguir entre un marco de referencia inercial en reposo en un campo gravitatorio y uno acelerado en ausencia de gravedad. Este constituye un escenario legítimo para evaluar supuestos teóricos porque si la secuencia de eventos que describe pudiera ocurrir, entonces, se producirían las consecuencias que el experimento describe. Si la condición de *realizabilidad en principio* está satisfecha en principio, no importa cuán remota pueda ser la realización de ciertos experimentos mentales como experimentos reales. En síntesis, aunque los experimentos mentales y reales son prácticas diferentes, que emplean recursos diferentes, la viabilidad de un experimento mental puede evaluarse en conexión con la experimentación real considerando si la secuencia de eventos que narra el experimento mental se produce de acuerdo con las conexiones causales que suponemos operativas en el mundo real.

Ahora bien, incluso si es posible adoptar este criterio para valorar instancias de experimentos mentales irrealizables empíricamente, queda por examinar la cuestión de si la coincidencia de sus resultados con los de experimentos reales debe aceptarse como criterio de confiabilidad de los experimentos mentales. Según algunas perspectivas, los resultados de los experimentos mentales realizables deberían coincidir completamente con los resultados de los experimentos reales correspondientes. No obstante, dado que los experimentos mentales carecen de la intervención directa en la realidad física, sus resultados dependen del estado del conocimiento empírico y de la aplicación correcta de los supuestos teóricos. En palabras de Hacking, los experimentos mentales no pueden “torcer la cola del león” (Hacking, 1983: 149). Su proyección sobre la realidad está limitada por el conocimiento a partir del cual se los formula y por la adecuación de los supuestos teóricos que se ponen en juego. Esto no implica que sean incapaces de proporcionar nuevo conocimiento o que no puedan emplearse para apoyar hipótesis, pero, dado que son entidades diferentes de los experimentos reales, su mecanismo de obtención de conocimiento es también diferente. No obstante, en algún sentido los experimentos mentales pueden ser *traducidos* a experimentos reales, en tanto los primeros establecen conexiones causales y con ello anticipan las relaciones entre objetos o elementos del sistema que se proponen representar. Pero los experimentos reales en los cuales podría concretarse un interrogante planteado por un experimento mental no necesariamente coinciden en sus resultados con los experimentos mentales originales. Los resultados de un experimento real que consiste en la realización material de un experimento mental podrían corregir e incluso refutar el resultado del experimento mental. La realización de experimentos reales derivados de experimentos mentales, aportaría un elemento de evidencia independiente a las teorías en cuestión (en el caso de que se trate de experimentos confirmatorios) o un elemento para poner en duda las hipótesis en cuestión. Pero su relación con los experimentos mentales debería asimilarse a la que guardan los modelos con los experimentos reales que están vinculados a estos.

Por otra parte, al considerar los vínculos entre los resultados de los experimentos mentales y reales, es relevante evaluar si las estrategias de validación de la confiabilidad de los resultados de los experimentos reales pueden emplearse también para validar experimentos mentales. La experimentación real es considerada generalmente como la principal fuente de conocimiento nuevo sobre el mundo natural. No obstante, los medios para acceder a este conocimiento son también falibles por lo que los filósofos de la experimentación han consignado una serie de estrategias utilizadas para validar las observaciones y los resultados experimentales. Hacking (1983: 91) argumentó que los experimentadores cuentan con varios elementos para argumentar a favor de la estabilidad de los resultados experimentales. Señaló que aunque los experimentos sean dependientes, como mínimo, de la teoría del aparato experimental, las observaciones permanecen robustas a pesar de los cambios en las teorías de los aparatos y de los fenómenos. Por otra parte, la posibilidad de manipular e intervenir en los objetos bajo observación es una prueba de su independencia respecto de las teorías que sostiene el observador. Asimismo, la confirmación independiente de los resultados experimentales (u observacionales) empleando diferentes instrumentos, cuyo funcionamiento es explicado por teorías físicas diferentes es otro elemento para argumentar a favor de la estabilidad de dichos resultados. La propuesta inicial de Hacking fue largamente debatida. Otras perspectivas señalaron que el conocimiento experimental es influido por factores de índole sociológico, psicológico y pragmático y que la objetividad de dicho conocimiento podría ser puesta en duda debido a estas interferencias. Autores como Galison (1987 y 1997), postularon que la noción de evidencia está determinada por las tradiciones a las que los experimentadores pertenezcan, por la relación entre la teoría y la experimentación (cuyos cambios no ocurren siempre simultáneamente) y por varias otras formas en las cuales la teoría influye sobre las decisiones de los experimentadores.

Una de las posiciones más escépticas respecto de la confiabilidad del conocimiento experimental está representada por Collins (1985), para quien no hay criterios formales que puedan aplicarse para decidir sobre el correcto funcionamiento de un aparato experimental y, por lo tanto, el resultado correcto de un experimento es establecido discrecionalmente por una comunidad científica en función de factores sociales, diversos intereses cognitivos y consideraciones de orden pragmático. Otras contribuciones hacen sugerencias similares, dirigidas fundamentalmente a señalar que el acuerdo con los compromisos teóricos y las prácticas preexistentes desempeñan un papel determinante en la aceptación o rechazo de los resultados experimentales. Una posición intermedia entre las perspectivas que defienden la existencia de motivos racionales para la validación de los resultados experimentales y las que rechazan esta idea porque sostienen que la práctica científica está determinada por factores sociológicos, históricos y pragmáticos, es el punto de vista de acuerdo con el cual tanto el procedimiento material, como la teoría del aparato y del fenómeno en cuestión son recursos plásticos. Hacking (1992: 56) sugiere que los resultados, las teorías y el equipo de laboratorio evolucionan y se acoplan mutuamente. Cuando adquieren estabilidad, producen el carácter auto reivindicativo de la ciencia de laboratorio. Las preguntas, el conocimiento de fondo, el instrumental, los detectores y generadores de datos, los datos, su análisis y representación se ajustan mutuamente para que podamos concebir a los datos como fenómenos y a las hipótesis como verdaderas respecto de estos fenómenos. Aunque no sea posible evaluar aquí estas posiciones, es indudable que los científicos emplean estrategias diversas para validar los resultados experimentales, estén estas teñidas o no de los intereses progresistas o conservadores de las comunidades a las que pertenecen.

Franklin (2002) desarrolla una “epistemología de la experimentación” y detalla un conjunto de estrategias que permiten adjudicar un grado de creencia razonable a los resultados experimentales. Estas estrategias se derivan de la observación de las prácticas experimentales y no constituyen reglas racionales

excluyentes o exhaustivas. Además de la confirmación independiente y la estabilidad de las observaciones, señala la calibración del aparato experimental a partir de la reproducción de fenómenos conocidos, la reproducción de artificios conocidos de antemano, la eliminación de fuentes de error, el uso de los propios resultados experimentales para argüir a favor de su validez, el empleo de teorías bien conocidas de los fenómenos para explicar los resultados experimentales, el empleo de aparatos experimentales basados en teorías bien corroboradas y el empleo de argumentos estadísticos para apoyar los resultados (Franklin, 2002: 224-225).

Si bien estas estrategias no pueden ser directamente trasladadas a los experimentos mentales, son susceptibles de emplearse, con algunas modificaciones, en la validación de esta clase de experimentos. Evocar fenómenos bien conocidos en la descripción de los experimentos mentales es una estrategia que en algunos casos se emplea para exhibir la perspicuidad de la hipótesis que se propone defender. Estos fenómenos se producen en circunstancias habituales y lo que propone el experimento mental es una interpretación novedosa de su comportamiento. Tal es el caso de varios experimentos mentales de Galileo, que, además, pueden ser reproducidos en muchas circunstancias similares. Otra estrategia para apoyar la confiabilidad de los experimentos mentales es la inclusión en la descripción del experimento de algún aparato experimental que produce fenómenos bien conocidos, como sistemas de poleas, interferómetros, contadores Geiger, o péndulos.

Si se considera a la evaluación de los resultados experimentales como una búsqueda de posibles errores, encontrar el responsable del “error” parece ser más bien un ejercicio detectivesco que consiste en examinar un conjunto de mecanismos y teorías. En el caso de los experimentos mentales, la ausencia de ejecución material permite reducir las fuentes posibles de error en sus resultados. El resto de los componentes de un experimento mental, a saber, las hipótesis que se pretende apoyar o socavar con el caso, los supuestos de fondo en la

descripción del escenario y el aparato experimental, el conocimiento empírico y físico que se supone que los experimentadores poseen para poder realizar el experimento, pueden ser puestos bajo escrutinio de una manera similar a como se examinan en los experimentos reales.

Aunque los experimentos mentales pueden fallar en ocasiones de maneras similares a las de los experimentos, también pueden errar de maneras en las que los experimentos ordinarios no pueden (Brown, 2004: 29). A diferencia de los experimentos reales, los experimentos mentales pueden fallar en considerar todas las resoluciones posibles de una situación. Un ejemplo de este tipo de error puede encontrarse en el experimento de Lucrecio de acuerdo con el cuál no es posible concebir un universo limitado. Aunque el caso propone un escenario en el cual sería absurdo concebir un límite para el universo, falla en cuanto es posible concebir un espacio finito pero no limitado, por ejemplo, esférico. Esta es una razón para considerar que, aunque es posible aplicar algunos de los criterios de confiabilidad epistémica de los experimentos reales a los experimentos mentales, las particularidades de estas prácticas demandan considerar posibles fuentes de error que no están presentes en los experimentos reales. Otra posible fuente de error en los experimentos mentales se encuentra en la introducción de idealizaciones y abstracciones. Una manera de evaluar la confiabilidad de estas distorsiones es considerar en qué sentido es posible desidealizar los escenarios y acercarlos a los fenómenos reales. En este sentido, considerar a los experimentos mentales como límites ideales a los que los experimentos reales se acercan puede ser de utilidad para valorar comparativamente sus resultados e identificar idealizaciones ilegítimas.

6.7 Conclusiones

Si bien no es posible establecer un conjunto de condiciones aplicables a la evaluación de todos los experimentos mentales, un repertorio de aspectos

característicos de estos experimentos debe ser tenido en cuenta con el fin de determinar si el caso tiene éxito en concretar el objetivo cognitivo para el cual se formula. Un primer conjunto de condiciones se vincula al escenario hipotético descrito en la narrativa de los experimentos mentales. Un segundo grupo de criterios se establece a partir de la consideración una clase de experimento mental ilegítimo conocido como *extractor de intuiciones*. Un tercer conjunto de condiciones puede derivarse de las similitudes entre los experimentos mentales y las simulaciones computacionales. Un cuarto grupo de criterios se sigue de las restricciones que es posible aplicar para garantizar la legitimidad de la interpretación teórica de un experimento mental. Por último, la vinculación entre los resultados de los experimentos reales y los experimentos mentales proporciona algunas pistas para evaluar las semejanzas y diferencias entre estos resultados.

En relación con el escenario hipotético descrito en la narrativa del experimento mental, varios factores deben tenerse en cuenta para la evaluación de la eficacia del experimento. En primer lugar, es preciso evaluar si los experimentos mentales proporcionan una respuesta a la pregunta que motiva su formulación. Se pueden identificar razones diferentes por las cuales los experimentos fallan en este sentido. Por una parte, los experimentos mentales pueden fallar por razones vinculadas a la competencia del experimentador. Un experimento mental puede no arrojar un resultado (en absoluto) si el experimentador no es capaz de llevar a cabo el experimento propuesto por la narrativa. Por otra parte, el experimento mental puede fallar en dar una respuesta debido a que el experimento está descrito de manera incompleta o imprecisa, lo que imposibilita que el experimento arroje un resultado unívoco. Finalmente, un experimento mental puede fallar porque produce resultados que no son respuestas a las preguntas que inicialmente motivaron el experimento mental. En estos casos, la hipótesis que el experimento se proponía poner a prueba o explorar

permanece indeterminada aunque el escenario imaginario promueva la exploración de alguna hipótesis alternativa.

En segundo lugar, es necesario analizar en qué grado el experimento proporciona evidencia a favor o en contra de alguna hipótesis. Dado que los experimentos ordinarios se proponen contrastar una hipótesis a partir de la variación de una situación controlada en el contexto de una teoría específica, puede suponerse que, en tanto los experimentos mentales se basan en el método de la variación concomitante, deberían proceder esencialmente del mismo modo. En este sentido, una condición mínima para que el experimento proporcione evidencia relevante para una hipótesis es que en la narrativa del experimento todas las variables y parámetros relevantes para su ejecución sean adecuadamente identificadas y su correlación apropiadamente establecida.

La cercanía que el experimentador tiene con la situación planteada en el escenario hipotético es otro elemento a considerar para valorar su efectividad. Si el escenario hipotético resulta demasiado lejano a los contextos respecto de los cuales el experimentador posee representaciones, las respuestas a las preguntas planteadas por el experimento pueden resultar poco confiables. Cuando los supuestos contrafácticos involucrados en la descripción del experimento nos llevan a mundos demasiado extraños, no es posible evaluar si las hipótesis resultan explicativas de los fenómenos postulados.

Otra cuestión, íntimamente vinculada con la naturaleza de los escenarios hipotéticos, es la de establecer los grados de contrafacticidad aceptables en la descripción de un experimento mental. Aunque no es posible delimitar el tipo de posibilidad que los experimentos mentales físicos legítimos podrían explorar para producir resultados relevantes, es plausible establecer un criterio general para evaluar la plausibilidad de los escenarios involucrados. Para determinar si el caso hace un uso legítimo de los contrafácticos es preciso detallar el experimento en términos de su objetivo cognitivo, los procesos detallados en su descripción y su

resultado, de manera que sea evidente si los supuestos contrafácticos que violan determinadas leyes físicas son relevantes para la evaluación de una hipótesis en cuestión.

Una última condición relativa al escenario hipotético propuesto en el experimento mental, es que en caso de que el experimento mental se proponga como un contraejemplo a una hipótesis, la posición rival sea descrita de manera aceptable para quienes abogan por dicha teoría rival, de manera que el experimento mental evite apoyarse en simplificaciones o caricaturizaciones.

Un caso particular de experimento mental fallido identificado por Dennet y Hofstadter es el de los *extractores de intuiciones*. Aunque no es claro cómo detectar si el experimento en cuestión es un extractor de intuiciones, esta noción llama la atención sobre algunos supuestos implícitos en la descripción del experimento que minan la objetividad del resultado. Si la descripción del experimento introduce supuestos teóricos que hacen plausible una interpretación y socavan otra, entonces, podría asimilarse estos casos a las falacias de petición de principio. Una manera de detectar si un experimento en cuestión funciona como extractor de intuiciones es considerar el alcance y la plausibilidad de otras hipótesis (emparentadas con las que son el objeto del experimento) en los escenarios en los cuales se intenta evaluarlas.

La posibilidad de detectar casos que funcionen como extractores de intuiciones remite a la cuestión acerca de qué intuiciones podrían ser legítimamente invocadas en un experimento mental. En la literatura reciente que tiene por objeto tanto el análisis epistemológico de esta clase de experimentos en ciencias fácticas como la naturaleza general de las intuiciones en física, no existe un acuerdo sobre la naturaleza de estos contenidos mentales. Más allá de esta controversia, no parece esencial delimitar filosóficamente la noción para reconocer que las intuiciones, comoquiera que se las conceptualice, se vinculan al conocimiento empírico de fondo de los experimentadores y a su sistema

conceptual. Dado que las intuiciones constriñen la estructura a partir de la cual los sujetos conocen el mundo, en algún sentido también restringen aquello que pueden considerar como físicamente posible. Podría pensarse que los experimentos mentales serán exitosos en contextos específicos en los cuales los experimentadores comparten un conjunto suficientemente amplio de valores y creencias. Dentro de estas tramas, los sesgos, los prejuicios, la parcialidad de los intereses operan introduciendo una circularidad y desestimando posibilidades empíricas. Si las tesis se valoran solo a partir de ciertos consensos, entonces, las mismas pueden resultar aceptables o no en virtud de las disposiciones previas a aceptarlas o rechazarlas. La interpretación que a partir de estas observaciones nos parece razonable asumir es que las intuiciones o el conocimiento empírico de fondo solo tienen un valor evidencial relativo a contextos específicos. Las intuiciones solo pueden ser desafiadas o modificadas en casos en los cuales se comparta conocimiento de fondo y se presenten alternativas inesperadas dado cierto marco teórico compartido.

Las analogías positivas que pueden encontrarse entre simulaciones computacionales y experimentos mentales, permiten identificar las estrategias de validación de las simulaciones y los criterios de confiabilidad que podrían aplicarse a los experimentos mentales. Ambas prácticas dependen de la representación de la evolución de un sistema físico y emplean para ello estrategias de modelización que, aunque varían cualitativamente, desempeñan el mismo papel: hacer tratable un problema en el que la consideración de todas las variables resulta inasequible. Las maneras de vincular los resultados idealizados de los experimentos mentales con los sistemas reales puede considerarse como un reflejo de las estrategias empleadas para evaluar la legitimidad de las transformaciones del modelo físico introducidas para construir de los algoritmos computables. En estos casos, para hacer tratables a las ecuaciones, la construcción del modelo no solo implica realizar simplificaciones y remover grados de libertad o hacer suposiciones de simetría irrealistas (Winsberg, 2010:

12) sino también incorporar ecuaciones diseñadas para computar algún efecto del mundo natural dejado de lado.

Otros mecanismos para validar las simulaciones computacionales son trasladables a la evaluación de la confiabilidad de los experimentos mentales. Particularmente, las estrategias que consisten en la comparación de sus resultados con datos disponibles provenientes de la observación y medición de los fenómenos. En dominios en los cuales la experimentación real es imposible o extremadamente difícil, los resultados de los experimentos mentales pueden ser comparados con los resultados de otros experimentos mentales sobre el mismo fenómeno, cotejados con datos empíricos relevantes, e incluso vinculados con los resultados de simulaciones computacionales concernientes a estos fenómenos. Estrategias similares se emplean habitualmente en la validación de las simulaciones.

La comparación entre simulaciones y experimentos mentales permite hallar una estrategia más para la evaluación de éstos. Como sugiere Winsberg (2010), la naturaleza abigarrada y compleja de las simulaciones demanda una epistemología *sui generis* con criterios de confiabilidad que provengan de una comprensión apropiada de su naturaleza. La evaluación de la confiabilidad de los experimentos mentales también implica valorar, en función de estrategias usuales en la epistemología de la experimentación, las diferentes operaciones que se presentan en el experimento mental.

La interpretación teórica de los resultados de los experimentos mentales es una cuestión que ha originado una controversia relativa a sus poderes probatorios. Existen varios casos de experimentos mentales que dan lugar a interpretaciones diferentes a partir del mismo escenario. El desacuerdo en las interpretaciones del escenario hipotético podría deberse a que, dadas las mismas condiciones iniciales, dos puntos de vista teóricos diferentes dan lugar a interpretaciones diferentes sobre la evolución del sistema, o bien a que el mismo

escenario imaginario se desarrolla en un único curso pero este resultado es interpretado como evidencia a favor de puntos de vista antagónicos. En cualquier caso, para esclarecer cuál de las interpretaciones es la más adecuada, sería preciso especificar los supuestos relevantes en cada caso y describir el experimento en un lenguaje teóricamente neutral, es decir, en términos que los partidarios de dos teorías rivales acepten. Pero en lo que refiere al trasfondo teórico de un experimento mental, no es evidente en cuál de las capas del conocimiento de fondo es necesario detenerse para hacer conspicua una interpretación. En cuanto a la descripción del aparato experimental o de la situación hipotética, no es claro que pueda de manera que permita en última instancia decidir qué interpretación teórica constituye una mejor explicación de los fenómenos descriptos. En consecuencia, aparte de la sugerencia de explicitar los supuestos relevantes, no es posible establecer un criterio claro acerca de cómo proceder en el caso de interpretaciones rivales de un mismo experimento.

La reconstrucción de las inferencias que dan lugar a una interpretación teórica puede favorecer una evaluación de su validez. Aunque no existe una única manera de formalizar este procedimiento, es útil considerar si el escenario hipotético se propone como un contraejemplo, un caso para el que un supuesto teórico propone una interpretación o un caso que puede considerarse canónico y con ello realizar una inferencia ampliativa a otros casos. Estas funciones pueden facilitar la lectura de los argumentos como *modus tollens*, inferencias a la mejor explicación o razonamientos inductivos.

La comparación entre experimentos mentales y sus contrapartes reales hace posible reexaminar y esclarecer algunas cuestiones vinculadas a los mecanismos empleados en uno y otro caso para la obtención de resultados. La coincidencia entre predicciones realizadas por experimentos mentales y resultados empíricos ofrece respaldo epistémico a los experimentos mentales. Pero la falta de coincidencia perfecta entre los resultados de los experimentos mentales y reales no es una razón suficiente para dudar de la legitimidad de los

primeros, dado que los mecanismos que emplean para obtener conocimiento del mundo natural son diferentes. El conocimiento que los experimentos mentales pueden proporcionar está condicionado por varios factores como el conocimiento del experimentador, los supuestos teóricos, las idealizaciones y abstracciones, que pueden restringir de manera más directa los resultados que puedan obtenerse del ejercicio imaginario, mientras que los experimentos reales intervienen directamente en el mundo natural. Si algún criterio general puede considerarse como condición de adecuación de los experimentos mentales a los reales, es el de *realizabilidad en principio* del experimento mental. Según este criterio, si la secuencia de eventos que el escenario imaginario describe pudiera ocurrir, entonces, se producirían las consecuencias o el fenómeno que el experimento describe. En otras palabras, el criterio señala que un experimento mental aceptable debe capturar adecuadamente las relaciones causales, de manera que si las situaciones invocadas por la narrativa pudieran tener lugar, sería posible, dado nuestro conocimiento del mundo, decir qué ocurría.

Por otra parte algunas estrategias de validación de los experimentos reales pueden trasladarse a los experimentos mentales. Los experimentos mentales son más confiables cuando evocan fenómenos bien conocidos o se incluye en su narrativa la descripción de algún aparato experimental que produce fenómenos bien conocidos. Estas estrategias pueden ser valoradas positivamente en la evaluación de la confiabilidad de un experimento mental. Sin embargo, es necesario no perder de vista que los experimentos mentales pueden fallar de maneras diferentes a como los experimentos reales pueden hacerlo. Fundamentalmente, los experimentos mentales, a diferencia de los reales pueden fallar en concebir todas las alternativas empíricas posibles y en llevar demasiado lejos las distorsiones empleadas en el escenario idealizado.

TERCERA PARTE

Las funciones de los experimentos mentales

CAPÍTULO 7

Las funciones de los experimentos mentales

7.1 Introducción

En el contexto de la filosofía de la experimentación, las prácticas experimentales han sido clasificadas de acuerdo con tres criterios diferentes: a) sus rasgos estructurales, b) los resultados que son capaces de producir y c) las funciones que pueden desempeñar respecto de las teorías. El primer criterio permite ubicar a los experimentos mentales dentro de la clase de experimentos no materiales (Morgan, 2003). Debido a sus características estructurales, los resultados de los experimentos mentales son siempre cualitativos. Las funciones que los experimentos mentales pueden desempeñar respecto de hipótesis o principios teóricos han sido materia de una ardua discusión. El propósito de este capítulo es revisar los criterios taxonómicos empleados en las teorizaciones de los experimentos reales y examinar las clasificaciones existentes de los experimentos mentales para elaborar una tipificación que tome en consideración el papel que han desempeñado históricamente y que sea coherente con la epistemología propuesta.

La perspectiva dominante en la filosofía analítica de la ciencia del siglo XX ha relegado a los experimentos mentales al contexto de descubrimiento, arguyendo que solo pueden desempeñar funciones heurísticas e ilustrativas. Desde la perspectiva de Hacking, los experimentos reales tienen una *vida propia* (Hacking, 1983: 150) porque pueden madurar, evolucionar, adaptarse o reestructurarse. Son capaces de producir adelantos en la tecnología de los instrumentos con el fin proporcionar mejores resultados y desarrollar las

habilidades de los experimentadores. Incluso pueden ser replicados a la luz de nuevas teorías. Pero los experimentos mentales son fijos y, en gran parte, inmutables (Hacking, 1992: 307). Esto significa que están limitados a un contexto en particular y no pueden interpretarse en otros. Al depender de una serie de supuestos básicos tomados de un marco nomológico, los poderes epistémicos de los experimentos mentales se limitan a su capacidad de exhibir tensiones en el marco conceptual de los experimentadores. En síntesis, la dependencia teórica y la ausencia de intervención fueron las principales razones para negar que los experimentos mentales puedan desempeñar las mismas funciones que los experimentos reales y para restringir su uso legítimo al contexto de descubrimiento.

Dentro de la epistemología de los experimentos mentales, las clasificaciones existentes coinciden en que pueden cumplir satisfactoriamente funciones heurísticas, ilustrativas y también críticas, refutando una hipótesis o presentando un contraejemplo que mina su alcance. Pero pocas perspectivas les atribuyen algún papel en la justificación de hipótesis y teorías.

No obstante las observaciones de Hacking, los experimentos mentales han cumplido un importante papel en la historia de la ciencia. Muchos de ellos han sido decisivos en la generación de nuevos conceptos, en la refutación de hipótesis y, en algunos casos, han sido relevantes para la formulación y justificación de nuevas teorías. Por esta razón, la relegación de los experimentos mentales a las funciones heurísticas no hace justicia a su importancia histórica. Atendiendo a ciertas restricciones, es posible defender la legitimidad de su uso apologético (como evidencia a favor de hipótesis teóricas).

La revisión de la práctica científica en lo que concierne a la construcción y usos de las simulaciones y modelos computacionales, puso de manifiesto que en muchos casos estas pueden desempeñar las mismas funciones que los experimentos reales. Dado que, como se ha argumentado, las simulaciones y los

experimentos mentales comparten diversos rasgos estructurales, es posible extender algunas de estas conclusiones a los experimentos mentales. Los experimentos reales y las simulaciones computacionales han desempeñado funciones exploratorias cuando anteceden a la configuración de un marco teórico (Steinle, 2002: 423). Esta es una importante función que ha sido desatendida en la epistemología de los experimentos mentales.

En este capítulo, se propone un criterio clasificatorio para los experimentos mentales. Se introduce la noción de *uso* en contraposición a las funciones. Dado que los experimentos mentales se caracterizan por presentar una narrativa similar a la descripción de un experimento y están acompañados por un argumento que provee la interpretación teórica, se argumenta que el primer elemento puede ser separado del segundo dando lugar a diferentes empleos del experimento relacionados con diferentes objetivos cognitivos. Esta interpretación es sugerida por los casos en los cuales el mismo experimento ha desempeñado funciones opuestas respecto de una misma hipótesis. Este hecho permite sostener, por un lado, que los experimentos mentales gozan de cierta independencia teórica y no están anclados a un contexto teórico en particular, como supuso Hacking, y, por otro, que su función está definida por el objetivo cognitivo que tienen dentro de un determinado marco contextual.

Sobre la base de la noción de uso de los experimentos mentales, se elabora una tipificación que se propone valorizar su relevancia histórica y dar cuenta del papel que efectivamente desempeñaron en los contextos en los que fueron formulados y usados. La distinción propuesta contempla el hecho de que en algunos casos el mismo experimento mental ha sido usado con diferentes propósitos. Por tanto, no supone que las distinciones sean excluyentes. La tipificación distingue entre usos heurísticos y usos críticos. Entre los primeros se diferencia entre los usos destinados a plantear un problema, los usos ilustrativos y los exploratorios. Entre los segundos se distingue entre usos refutatorios y confirmatorios.

7.2 Las tipificaciones existentes en la epistemología de los experimentos mentales

Las tipificaciones existentes de los experimentos mentales los diferencian a partir de las funciones que pueden desempeñar respecto de hipótesis teóricas y en virtud del mecanismo empleado para inferir conclusiones acerca de los fenómenos. La mayoría de las clasificaciones sigue algún criterio inspirado en la clasificación de los experimentos reales. De acuerdo con el grado de semejanza o diferencia que la perspectiva en cuestión atribuya a los experimentos mentales respecto de los reales, se restringen las funciones legítimas de los primeros.

El catálogo de Brown (2011) supone que los experimentos mentales funcionan de manera análoga a los reales y que, en principio al menos, pueden cumplir las mismas funciones. Los experimentos mentales tienen la capacidad de: ilustrar un principio teórico, presentar un fenómeno para el que no existe una teoría bien desarrollada y apoyar o refutar una hipótesis. Su clasificación distingue entre los experimentos mentales *constructivos*, que proporcionan algún tipo de apoyo a una hipótesis teórica, y los *destructivos*, cuyo propósito es socavar algún principio.

Dentro de la categoría de experimentos constructivos, Brown identifica los experimentos *mediadores* que facilitan la derivación de conclusiones a partir de una teoría bien establecida. En este mismo grupo, los experimentos constructivos *conjeturales* presentan un fenómeno ideal para el cual no existe una teoría bien desarrollada. Los experimentos constructivos *directos* proporcionan evidencia a favor de una hipótesis a partir de la presentación de un fenómeno mental o ideal. Por su parte, los experimentos *destructivos* funcionan de manera análoga a un argumento por reducción al absurdo de un principio teórico, derivando una consecuencia absurda de una hipótesis. Brown postula, además, que existe una clase de experimentos cuyas características son transversales a los experimentos

constructivos directos y destructivos: los experimentos mentales *platónicos* a los que atribuye poderes epistémicos extraordinarios.

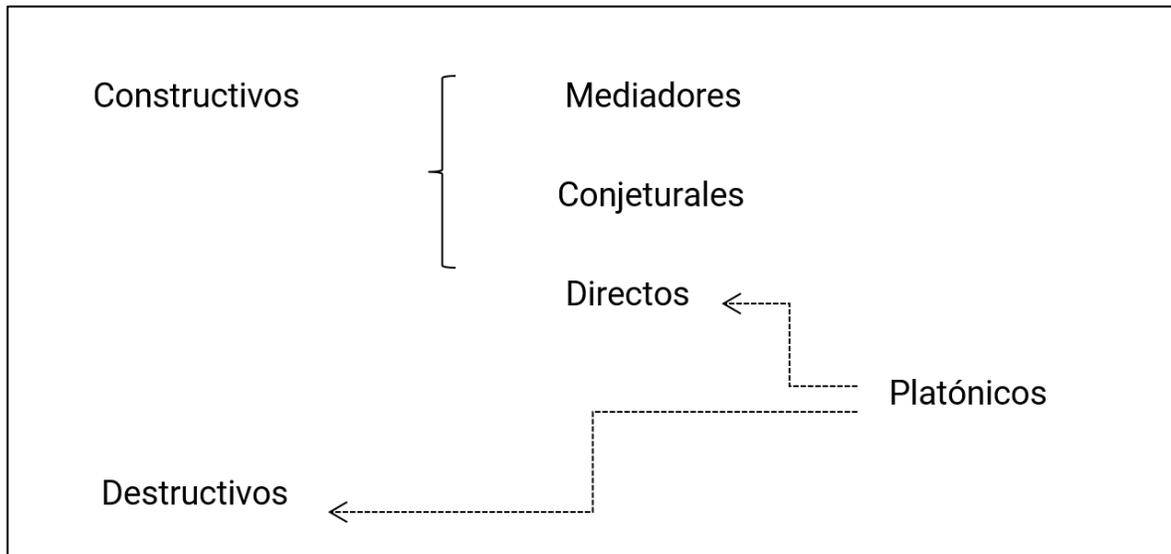


Figura 7.1: clasificación de Brown (2011: 33).

El ejemplo paradigmático de la clase de experimentos mentales mediadores es para Brown el experimento del *demonio de Maxwell* (Maxwell, 1871: 308-309). Los experimentos mentales *mediadores* ilustran algún aspecto de la teoría que podría parecer contradictorio o absurdo, haciéndola más aceptable. Esta clase de experimentos puede actuar de manera semejante a un diagrama en una prueba geométrica, colaborando en la comprensión la demostración formal. La descripción del experimento incluye dos cámaras que contienen gas, uno más caliente que el otro, separadas por una puerta operada por un ser inteligente que deja pasar las moléculas más rápidas del gas más frío al compartimiento donde está el gas más caliente y viceversa. El objetivo de todo el ejercicio es mostrar que lo que parecía impensable no es tan impensable después de todo: la disminución de la entropía del sistema. Para Brown, el experimento de Maxwell no es una objeción a la segunda ley de la termodinámica, sino una evidencia de su carácter estadístico. El demonio de Maxwell ayuda a hacer más plausibles algunas de las

consecuencias de la teoría, eliminando de ese modo una barrera para su aceptación (Brown, 2011: 37).

El célebre experimento del balde giratorio de Newton ([1687] 1974: 10-12) es para Brown un ejemplo de un *experimento conjetural*. El fenómeno presentado por el experimento mental, a saber, la curva que se produce en la superficie del agua cuando el balde y la cuerda han dejado de girar, es un fenómeno ideal que se produciría en la circunstancia en la cual el balde pudiera rotar en un universo vacío (Brown, 2011: 39). La existencia del espacio absoluto postulada por Newton proporciona una explicación de la ocurrencia del fenómeno.

Los experimentos mentales *destructivos* generalmente están dirigidos contra hipótesis bien definidas. En algunos casos, se proponen socavar otros experimentos mentales que apoyan hipótesis (Brown, 2011: 34). Un caso ejemplar de esta clase de experimentos es el empleado por Galileo para refutar la hipótesis aristotélica de la inmovilidad de la Tierra. Para los aristotélicos, si la Tierra rotara, una bala de cañón arrojada desde la cima de una torre no caería al pie de la misma, sino que quedaría muy atrás (ya que la torre que está unida a la Tierra giratoria, se alejaría de la bala de cañón por el movimiento de rotación). En respuesta este experimento, y contra la hipótesis aristotélica de la inmovilidad de la tierra Galileo elabora en el *Diálogo sobre los dos máximos sistemas del mundo*, el conocido experimento del barco (Galileo, [1632] 1967: 186-188). El experimento muestra que no hay diferencia en el comportamiento de pequeños animales encerrados en la cabina de un barco (como mariposas, aves, peces y moscas), ni en el comportamiento de un sistema como una botella agua que gotea, cuando el barco se encuentra en reposo y cuando se encuentra en movimiento rectilíneo uniforme. La postulación de un escenario posible de estas características se propone socavar la hipótesis aristotélica, describiendo un escenario posible en el que no se produce el fenómeno invocado por el experimento mental de la bala de cañón.

Según Brown, el experimento mental del plano inclinado ideado por Stevin (Mach, [1893]1960: 25-33) establece, a partir de un fenómeno que puede representarse fácilmente en la imaginación, una teoría acerca de las propiedades de los planos inclinados. En este caso se trata de un experimento que cumple funciones probatorias y para Brown se clasifica como un *experimento mental directo*. Si bien Brown reconoce que el experimento se desarrolla sobre la base de hipótesis auxiliares⁴² que permiten derivar ciertas consecuencias de la presentación del fenómeno, sostiene que esta clase de experimentos procede *directamente*, sin la mediación de una inferencia, desde la presentación de un fenómeno al establecimiento de una teoría que lo explica.

La categoría de *experimentos mentales platónicos* es la más controversial de la clasificación elaborada por Brown. Esta clase de experimentos tiene la particularidad de trascender el conocimiento empírico y proporcionar *a priori* conocimiento del mundo natural. El tipo de fenómeno que tiene lugar en los experimentos mentales platónicos proporciona evidencia a favor de una hipótesis con independencia de toda interacción con los fenómenos físicos. Brown suscribe a una particular noción de fenómeno y a una definición del conocimiento *a priori* que asocia ciertas instancias del conocimiento de las leyes físicas con la intuición racional. Las particularidades de esta propuesta fueron ya examinadas en el capítulo 4. Para recapitular brevemente, el experimento mental empleado por Galileo ([1638] 1963: 62-65) para establecer que todos los cuerpos caen con la misma aceleración es el caso predilecto de Brown. El experimento de Galileo es un experimento mental platónico, en primer lugar, porque su resultado no depende de la introducción de ningún dato empírico nuevo. En segundo lugar, porque la inferencia que se realiza desde las condiciones iniciales establecidas no es ningún tipo de verdad lógica. En tercer lugar, el experimento aumenta el grado de creencia racional en la teoría postulada por Galileo luego de su ejecución (Brown, 2011: 98).

⁴² En particular, Brown sostiene que el principio de acuerdo con el cual no es posible generar una máquina de movimiento perpetuo a partir de un impulso inicial es una hipótesis auxiliar que nos permite concluir que los pesos permanecen en equilibrio.

Aunque en lo que podría identificarse como *la parte negativa* (o destructiva) del experimento, Galileo deriva una contradicción empleando la teoría aristotélica para explicar el fenómeno propuesto, de esta no puede derivarse la hipótesis que Galileo propone para reemplazar a la aristotélica. En este excedente reside el poder epistémico especial de los experimentos mentales platónicos.

Además del experimento de Galileo, Brown menciona dos casos de experimentos mentales platónicos: el experimento EPR y el experimento de la *Vis Viva* de Leibniz.⁴³

La clasificación de Brown comparte algunos rasgos esenciales con la de Popper (1959, Apéndice XI) por lo que parece plausible suponer se inspiró en ella. Popper, por su parte, no elaboró una taxonomía sobre la base de las funciones que los experimentos puedan desempeñar, sino de los usos o empleos que, desde su punto de vista se pueden identificar en algunos casos canónicos. El uso que Popper denomina *crítico* coincide con la función atribuida por Brown a los experimentos mentales destructivos. El experimento de Galileo de los cuerpos en caída libre es un ejemplar del uso crítico (Popper, 1959: 515-516). El uso *heurístico*, admisible en el contexto de descubrimiento de las teorías, coincide con los experimentos conjeturales de la clasificación de Brown. Según Popper, un caso representativo del uso heurístico de los experimentos mentales es el experimento concebido para defender el atomismo: “imaginemos que tomamos un trozo de oro o de otra substancia y que lo partimos en trozos cada vez más pequeños hasta que no puede subdividirse más” (Popper, 1959: 516). Si bien el uso crítico de los experimentos mentales es legítimo porque, en el marco de una controversia entre puntos de vista rivales, equivale al intento de poner de manifiesto que el autor de una teoría ha pasado por alto ciertas posibilidades, el *uso apologético* de los experimentos mentales es totalmente inadmisibles. Los experimentos mentales para Popper pueden refutar hipótesis pero no confirmarlas. Dentro de los

⁴³ Debido a que un tratamiento adecuado del experimento EPR excedería los límites de este trabajo, no lo analizaremos aquí. En la sección 5 examinaremos el experimento de la *Vis Viva* de Leibniz.

experimentos apologeticos ilegítimos categoría se pueden incluir varios casos de experimentos mentales en mecánica cuántica, como el experimento del microscopio de rayos *Gamma* de Heisenberg (1930: 20-21).

Contexto de descubrimiento	Heurísticos
Contexto de justificación	Críticos
	Apologeticos

Figura 7.2: Clasificación de Popper (1959: 516- 518).

Norton (1991), por su parte, clasifica a los experimentos mentales de acuerdo con la inferencia involucrada en la obtención de conclusiones. Dado que desde esta perspectiva los experimentos mentales son argumentos encubiertos, los detalles descritos en la narrativa de los mismos no son esenciales para la derivación de conclusiones. Los experimentos mentales exitosos son aquellos en los cuales es posible identificar una inferencia válida, cuyas premisas son (aproximadamente) verdaderas. Los detalles particulares descritos en la presentación del escenario imaginario son eliminados en la conclusión del argumento por vía inductiva o deductiva. Aunque Norton no realiza esta aclaración en la presentación de su clasificación, dado que desde su perspectiva los poderes epistémicos de los experimentos mentales son equivalentes a los de los argumentos, los usos en el contexto de justificación para apoyar o refutar una hipótesis tienen el mismo alcance que el de los argumentos. En el caso de los experimentos deductivos o del tipo 1, que estructuralmente son todos equivalentes a reducciones al absurdo, su alcance es el de mostrar una inconsistencia que se sigue de la teoría, es decir, ofrecer un contraejemplo imaginario. Los experimentos mentales que proceden por vía inductiva, o del tipo 2, no proporcionan el mismo tipo de apoyo evidencial que podría seguirse de la ejecución de un experimento real. En su examen de los experimentos mentales en el trabajo de Einstein, Norton (1991), sugiere que estos ejemplares (que podrían

considerarse casos “exitosos”) presuponen una posición verificacionista, de acuerdo con la cual, una teoría no debería incluir términos que no sean observacionales y que las teorías deben evitar distinguir estados de cosas que son observacionalmente indiscernibles. Norton, por otra parte, argumenta que los experimentos empleados en el contexto de descubrimiento se caracterizan por su opacidad. Aunque en muchos casos contienen algunos de los elementos seminales de nuevas perspectivas teóricas, su funcionamiento no es claro. No obstante, suelen presentar imágenes persuasivas que contribuyen al planteo de un problema.

Los experimentos mentales de tipo 1 son para Norton típicamente argumentos por reducción al absurdo. Los particulares pueden estar involucrados en un contraejemplo de una afirmación universalmente cuantificada de cuya contradicción se sigue la conclusión. Según Norton, la termodinámica presenta algunos de los experimentos mentales más efectivos y esto se debe a que las tres leyes de la termodinámica pueden formularse como aserciones acerca de imposibilidades. Una forma simple de derivar consecuencias de estas aserciones es a través de argumentos por reducción al absurdo.

Para derivar el teorema X se toma alguna de las leyes de la termodinámica como premisa y se supone la negación $\neg X$. Se muestra luego, por ejemplo, que $\neg X$ permite diseñar una máquina de movimiento perpetuo prohibida por las premisas. El argumento se convierte casi automáticamente en un experimento mental, ya que el diseño de una máquina como la del ejemplo implica la postulación de estados de cosas hipotéticos e incluye particulares que dan al argumento su carácter experimental. (Norton, 1991: 132)

La máquina de Carnot (1897: 52-56) puede ser interpretada como un experimento mental de este tipo. Los detalles invocados en la descripción del experimento involucran idealizaciones, como el supuesto de un gas ideal y la eliminación de la ficción en el movimiento de los pistones. El experimento establece el límite superior de eficiencia de todas las máquinas térmicas y considera que solo el calor consumido puede convertirse en trabajo aunque no

haya fricción (Hewitt, 2007: 354). El experimento funciona suponiendo la negación de la segunda ley de la termodinámica: es imposible diseñar una máquina de movimiento perpetuo del segundo tipo, cuyo único efecto es transferir calor de un reservorio de calor más frío a uno más caliente.

En los experimentos mentales de tipo 2 la conclusión es librada de los particulares por un salto inductivo. Este tipo de experimentos mentales suelen incluir la premisa implícita de que el caso aludido es un caso típico y que los particulares referidos son inesenciales para apoyar la conclusión (Norton, 1991: 131). El experimento del ascensor de Einstein (Einstein e Infeld, 1938: 53), que establece que todos los marcos de referencia inerciales son indistinguibles entre sí, es un experimento de esta clase. La narrativa del experimento involucra suposiciones ideales: un ascensor (o algún compartimiento opaco que no permita ver hacia el exterior) en alguna región remota del espacio lejos de la influencia de fuentes de atracción gravitacional, uniformemente acelerado, y un observador situado en su interior que desconoce las circunstancias. El punto central del experimento consiste en señalar que un observador situado dentro de una cámara opaca observa que los objetos se comportan del mismo modo en que lo harían en un campo gravitacional homogéneo. El salto inductivo del argumento consiste en afirmar que este es un caso típico y que, por lo tanto, es posible generalizar la equivalencia entre un marco de referencia uniformemente acelerado y uno en reposo (Norton 1991: 138).

Tipo 1	<i>Deductivos</i>
Tipo 2	<i>Inductivos</i>

Figura 7.3: Clasificación de Norton (1991: 131).

Sorensen (1992) y Haggqvist (1996), por su parte, ofrecen una clasificación que parte de la premisa de que los experimentos mentales desempeñan

esencialmente funciones negativas (refutatorias) en relación con los principios teóricos.

Desde la perspectiva de Sorensen (1992), la estructura lógica de este tipo de experimentos es la de una paradoja, por lo que sus funciones legítimas corresponden a las de los *refutadores aléticos*. Funcionan mostrando inconsistencias en la aplicación de un principio o hipótesis a la explicación de un fenómeno. Su estructura lógica puede regimentarse empleando operadores modales. La inferencia en cuestión consiste en la derivación de conclusiones a partir de los principios teóricos denominados *enunciados fuente*, que se interpretan como enunciados necesarios. La forma de estos razonamientos es: si el enunciado X (una ley o principio físico) es necesariamente verdadero, entonces, es posible la aplicación de tal principio en una situación determinada; luego, si de la aplicación de dicho principio en un escenario posible se deriva una consecuencia que hace altamente implausible al principio en cuestión, el experimento mental justifica el rechazo del enunciado fuente o al menos una restricción respecto del tipo de modalidad que se le atribuye.

Si bien, para Sorensen, los experimentos mentales son esencialmente experimentos no ejecutados, proporcionan un tipo de conocimiento modal a través de un mecanismo que puede ser parcialmente explicitado mediante una reconstrucción argumental. Esto no implica que los experimentos mentales sean en sí mismos argumentos. Esta clase de experimentos es capaz de mostrar que existe algún tipo de inconsistencia local que es preciso resolver abandonando alguna de las creencias a partir de las cuales se genera. Consecuentemente, las reconstrucciones de Sorensen no describen argumentos sino que proporcionan una versión semi-formal de los compromisos modales que las narrativas de los experimentos asumen y muestran las maneras más sencillas de resolver las contradicciones que se producen entre estas premisas (estas posibilidades están para Sorensen sugeridas en la descripción misma del experimento). Si bien no la propone como una clasificación, diferencia entre los experimentos que funcionan

como *refutadores de necesidad* y los que funcionan como *refutadores de posibilidad*.

Refutadores <u>aléticos</u> (Funciones negativas)	Refutadores de necesidad
	Refutadores de posibilidad

Figura 7.4: clasificación de Sorensen (1992: 135).

El experimento mental del demonio de Maxwell (Sorensen, 1992: 142) se dirige a mostrar que una “certeza” estadística fue tomada como necesidad física. Se trata de un tipo particular de *refutador de necesidad*, un caso en el que no se pretende mostrar la falsedad del enunciado fuente sino que se ha incurrido en algún error interpretativo en el tipo de necesidad que este enunciado tiene (o en el alcance de un operador modal). En este caso, el enunciado fuente es la segunda ley de la termodinámica. El escenario imaginario no refuta el axioma de acuerdo con el cual es imposible que el calor pase de un cuerpo más frío a uno más caliente a menos que algún otro cambio acompañe al proceso,⁴⁴ sino que propone un escenario en el cual un fenómeno prohibido por la ley resulta posible, mostrando con ello que el principio es de carácter puramente estadístico.

La interpretación de Mach sobre el experimento del balde giratorio de Newton es para Sorensen otro caso de *refutadores de necesidad* (Sorensen, 1992: 144-147). Los argumentos de Mach ([1893] 1960: 226-237), en los que reelabora varios de los experimentos de Newton para mostrar la existencia del espacio absoluto, si bien no prueban la falsedad de un enunciado fuente, muestran que un fenómeno que es evidencia de un principio puede ser explicado de una manera alternativa. Según Mach, hay un “contrafáctico erróneo” empleado para ilustrar el

⁴⁴ En enunciado fuente que Sorensen toma como segunda ley de la termodinámica, es la formulación de Clausius de acuerdo con la cual: “El paso de calor de un cuerpo más frío a uno más caliente no puede tener lugar sin compensación” (Clausius, 1879: 80).

principio que afirma la existencia del espacio absoluto, ya que el fenómeno que tiene lugar en el escenario descrito por Newton puede ser relativizado a un marco de referencia inercial diferente. Según Newton, la curva que se forma en la superficie del agua que contiene el balde es evidencia del espacio absoluto porque el movimiento del agua no puede ser considerado como relativo respecto de otros objetos debido a la omnipresencia de la fuerza centrífuga. Según Mach, las estrellas fijas son el marco inercial siempre que estemos dispuestos a postular una fuerza universal cuya magnitud depende de la aceleración y masa de los objetos que interactúan y que puede operar instantáneamente a distancia; entonces, podemos decir que el agua que rota en relación con las estrellas activa esta fuerza: cuanto más rápido gira el agua, más la enorme masa de las estrellas distorsiona la forma del agua (Sorensen, 1992: 146).

Los *refutadores de posibilidad* constituyen el segundo tipo de experimentos mentales. Estos funcionan mostrando que de un enunciado fuente es posible extraer consecuencias espurias. Los escenarios imaginarios evocan fenómenos o situaciones que se siguen de la irrestricción de un principio y que sugieren, por su apariencia absurda, el abandono del enunciado fuente en cuestión. Un ejemplo de este tipo de experimentos es el empleado por Mach ([1893] 1960: 219-220) para mostrar la transitividad de la relación de igual masa:

Los (...) refutadores de posibilidades a menudo se dedican a la exageración; hacen que una consecuencia ya absurda sea más absurda. Mach es responsable de un ejemplo encantador de esta ampliación. Su experimento mental demuestra la transitividad física de la relación de igual masa, es decir, que dos masas que son iguales a una tercera masa son iguales entre sí. Supongamos que *A*, *B* y *C* son cuerpos elásticos que se mueven libremente a lo largo de un anillo rígido y liso. Supongamos, además, que el principio no se cumple para estos tres objetos. En particular, supongamos que *A* tiene un valor de masa igual a *B* en el momento de la colisión y que *B* tiene un valor de masa igual a *C* pero que *C* tiene un valor de masa mayor que *A*. Si luego le impartimos una velocidad a *A*, *A* pasará la velocidad a *B*, y luego *B* a *C*. Pero como *C* tiene un valor de masa mayor que *A*, su colisión dará a *A* más velocidad de la que *A* tenía inicialmente. Este excedente se llevará luego a *B*, luego a *C*, después de lo cual se le dará un impulso adicional a *A*. Entonces, en cada ciclo, los cuerpos van más y más rápido sin límite. Esta

amplificación sin fin viola la primera ley de la termodinámica, que dice que la energía no se crea ni se destruye. Entonces, Mach concluye que "igual masa" debe ser transitiva. Pero observe que incluso un ciclo alrededor del anillo está excluido por la conservación de energía. (Sorensen, 1992: 154)

Podría interpretarse que la paradoja planteada por la negación del principio que Mach se propone demostrar funciona como una especie de reducción al absurdo y compele a aceptar el principio en cuestión. Es una clase de experimento que produce una "amplificación" (Sorensen, 1992: 155) de las consecuencias absurdas que pueden seguirse de un principio.

Sorensen sintetiza la distinción entre las dos clases de experimentos indicando que los primeros muestran que un principio es demasiado restrictivo, mientras que los segundos muestran que un principio es demasiado vago, estableciendo escenarios posibles de los que se siguen consecuencias paradójicas.

Häggqvist (1996), por su parte, propone cuatro formas de regimentar los experimentos mentales. No ofrece con ello una clasificación propiamente dicha, pero señala que es posible tipificar mediante la formalización del argumento asociado al experimento la forma de resolver la paradoja "inducida" por el experimento mental. En palabras de Häggqvist:

No pretendo exponer una estructura central "verdadera" e invariable compartida por todos los experimentos mentales. La ambición es la más modesta de encontrar un idioma factible para discutir las propiedades de los experimentos mentales que parecen importantes. (Häggqvist, 1996: 88)

Desde esta perspectiva, los experimentos mentales se diseñan típicamente para sugerir la conclusión de que una tesis es falsa. Los esquemas de regimentación propuestos tienen la ventaja de reflejar la estructura dialéctica que Häggqvist atribuye a los experimentos mentales. Primero, se ofrece a la contemplación un escenario contrafáctico presentado como posible; luego, se sostiene que la tesis o principio objetivo (del experimento) está relacionada con

una consecuencia determinada en ese escenario; finalmente, se afirma que esa consecuencia no se produce en ese escenario y, con ello, se sugiere que la tesis objetivo del experimento es falsa. Estos esquemas proponen una versión modalizada de la forma de los argumentos conectados con los experimentos reales. Si un experimento (real) conduce a la revisión o al rechazo de una teoría o hipótesis (T) previamente considerada, se puede decir que lo hace por medio de un condicional ($I \rightarrow O$) que predice un cierto resultado (O) dadas ciertas condiciones iniciales (I). Este condicional está conectado a través de un condicional anidado $T \rightarrow (I \rightarrow O)$; la falsedad de T se infiere de este condicional junto con una premisa que dice que I es verdadera y de otra premisa que dice que O es falsa (véase Häggqvist, 2009: 63).

Los esquemas posibles para reconstruir experimentos mentales (de los cuales podría derivarse una clasificación asociada a la forma del argumento en cada caso, aunque Häggqvist no los presenta así) no son excluyentes entre sí. Desde su perspectiva un mismo experimento mental podría ser regimentado como dos argumentos incompatibles entre sí. El esquema α permite derivar una inconsistencia de la teoría (T) o hipótesis en cuestión y la presentación del escenario (C) por medio de la derivación de un resultado (W) que se sigue de la teoría en las condiciones provistas por el escenario. El esquema α puede ser simbolizado con el siguiente conjunto: $\{T, \diamond C, T \rightarrow (C \Box \rightarrow W), C \Box \rightarrow \neg W\}$. El resto de los esquemas se diferencian del esquema α que prueba que T es falsa como falsa, por las diferentes maneras en las que “lidian” con la inconsistencia en cuestión y, por lo tanto, las conclusiones que se siguen de este conjunto de premisas.

El esquema β tiene los mismos elementos que el esquema α $\{T, \diamond C, T \rightarrow (C \Box \rightarrow W)\}$, pero la conclusión niega que afirmar que W sería verdadera en el escenario C es incompatible con T . La conclusión afirmada por β es: $\neg(C \Box \rightarrow \neg W)$.

El esquema γ afirma que aunque el escenario es posible y el experimentador mental tiene razón sobre lo la consecuencia que se seguiría de él, está demasiado lejos en el espacio lógico para ser relevante para la teoría bajo ataque. Este esquema también puede basarse en la afirmación de que la teoría, entendida correctamente, no se compromete con la afirmación falsa sobre el escenario, independientemente de tales consideraciones sobre la distancia modal. (Hägqvist, 2009: 66). La conclusión que se sigue del esquema γ es: $T \rightarrow (C \Box \rightarrow W)$. Lo que esta conclusión afirma es que la consecuencia extraña W no constituye una instancia de refutación para T .

El esquema δ deriva la conclusión de que un escenario es imposible: $\neg \Diamond C$, que se sigue de la contradicción entre la premisa que afirma la consecuencia que se seguiría del enunciado T en el escenario C ($T \rightarrow (C \Box \rightarrow W)$), y de la premisa que afirma que en tal escenario no se produciría esa consecuencia ($C \Box \rightarrow \neg W$).

Hägqvist revisa extensamente casos de experimentos mentales en filosofía analítica pero los experimentos mentales en física no son el principal objetivo de su estudio. No obstante, sus esquemas pueden aplicarse al análisis de casos científicos. Hägqvist (2009: 69-71) revisa la controversia entre Einstein y Bohr en torno al experimento de la caja de fotones (Bohr, 1949: 199-242), empleado sus esquemas de regimentación. De acuerdo con su interpretación, aunque las reconstrucciones no son excluyentes de otras lecturas argumentales, permiten dar cuenta de cómo en este célebre episodio el mismo experimento da lugar a dos argumentos contrapuestos. El argumento de Einstein puede estructurarse como un argumento α en el que la tesis objetivo es el principio de incertidumbre:

(E) Es posible (en principio) que un solo fotón salga de una caja equipada con un temporizador arbitrariamente exacto y una balanza de resorte arbitrariamente sensible [$\Diamond C$].

(E₂) Si se cumple el principio de incertidumbre, entonces, si un solo fotón saliera de una caja equipada con un temporizador arbitrariamente exacto y luego se pesara la caja, el tiempo y la energía de su paso no serían medibles simultáneamente con cualquier grado de precisión violando la desigualdad $\Delta E_x \Delta t > h/4\pi$ (donde h es la constante de Planck / 2π) [$T \rightarrow (C \square \rightarrow W)$].

(E₃) Pero si un solo fotón saliera de una caja equipada con un temporizador arbitrariamente exacto y luego se pesara la caja, el tiempo y la energía de su paso se podrían medir simultáneamente con cualquier grado de precisión [$C \square \rightarrow \rightarrow W$].

(E₄) Por tanto, el principio de incertidumbre no se cumple [$\neg \diamond T$].

Häggqvist interpreta la respuesta de Bohr como una reconstrucción que sigue el esquema (β), según el cual, no se puede sostener que W sería verdadera en C . Einstein afirma que es posible imaginar un dispositivo que permita la medición simultánea, arbitrariamente exacta, del tiempo y la energía de la salida del fotón, y que si un fotón saliera de dicho dispositivo, produciría lecturas arbitrariamente exactas para ambas magnitudes, aunque el principio de incertidumbre predice lo contrario. Bohr señaló atinadamente que esto no es posible porque la salida del fotón de la caja hace que se mueva en un campo gravitacional, lo que según la relatividad general afecta la velocidad del reloj, socavando así la exactitud deseada de la medición del tiempo de salida.

En síntesis, las clasificaciones existentes de los experimentos mentales les atribuyen diferentes funciones de acuerdo con su grado de similitud con los experimentos reales. Exceptuando la perspectiva de Brown, que supone que los experimentos mentales y reales son esencialmente similares, todas las perspectivas restringen las funciones de los experimentos mentales en el contexto de justificación. Así, por ejemplo, para Norton los experimentos mentales deductivos permiten derivar conclusiones absurdas y con ello, en algunos casos, mostrar la inconsistencia de un principio. Los experimentos que coinciden con

argumentos inductivos presentan un caso como ejemplar del que puede inferirse que la conclusión se aplica a todos o a muchos casos similares, pero no les atribuye funciones confirmatorias. Las taxonomías de Sorensen y Häggqvist, aunque destacan el aspecto experimental de estos experimentos centran su atención en las funciones refutatorias. En suma, para la gran mayoría de los filósofos de la ciencia los experimentos mentales, a diferencia de los experimentos reales, no tienen la capacidad de proporcionar evidencia confirmatoria para una hipótesis o teoría.

7.3 Problemas en torno a la clasificación de los experimentos mentales. Usos y funciones

En los diferentes intentos de clasificar los experimentos mentales, sus funciones se subordinan a la teoría. Todas las perspectivas asumen, con excepción a la de Brown, que los resultados que estos experimentos pueden arrojar son meramente cualitativos y, en general, afirman que dependen de una teoría bien desarrollada para realizar alguna contribución significativa. Lo que, según estos puntos de vista, podemos aprender de los experimentos mentales se limita al descubrimiento de la inadecuación o inaplicabilidad de un concepto teórico, o bien a la posibilidad de derivar consecuencias absurdas o contradictorias con algún principio teórico importante al intentar aplicarlo a escenarios imaginarios.

Dos supuestos comunes a las clasificaciones revisadas de los experimentos mentales son: 1) que en general admiten que experimentos mentales y reales pertenecen a un mismo tipo de práctica, y 2) que suponen que los únicos resultados capaces de proporcionar alguna evidencia a favor de una hipótesis son los de los experimentos reales. El primer supuesto se discutió en el Capítulo 2 de este trabajo, donde se mostró que aunque existen similitudes importantes entre experimentos reales y mentales, las relaciones de parentesco

entre estos dos tipos de experimentos son mucho más lejanas de lo que posiciones como la de Brown pretenden sostener. El segundo supuesto, de acuerdo con el cual solo los experimentos reales pueden proporcionar evidencia a favor de una hipótesis, se asocia a la idea de que la falta de intervención en el mundo natural hace a los experimentos mentales completamente dependientes de una teoría. Este punto requiere un examen más detenido. Hay buenas razones para considerar que los experimentos mentales no son completamente dependientes de las teorías y, con ello, para sostener que son capaces de cumplir funciones en el contexto de justificación, más allá de las de mostrar inconsistencias en una teoría. La posibilidad de distinguir en los experimentos mentales la descripción del escenario imaginario experimental de la interpretación del experimento favorece esta posición. El caso del experimento en la caja de fotones es uno que exhibe esta característica. Bishop (1999: 535-538) reconstruye el contexto de la disputa entre Einstein y Bohr y propone una interpretación de acuerdo con la cual el mismo experimento se emplea como contraejemplo y como justificación del principio de incertidumbre. De la revisión de este episodio, Bishop extrae la conclusión de que los experimentos mentales no son argumentos. Desde su punto de vista, Einstein y Bohr analizaron el mismo experimento desde perspectivas teóricas diferentes, lo cual dio lugar a dos interpretaciones contrapuestas de su resultado. Einstein concluyó que es posible medir el tiempo y la energía del fotón con cualquier grado arbitrario de precisión. Bohr concluyó que existe un límite irreductible para la determinación simultánea del valor de dos variables conjugadas, como la posición y el momento de una partícula.

De Mey (2003), por su parte, señala que los experimentos mentales tienen dos componentes: 1) la descripción de la situación imaginaria que coincide con la noción de realización material y 2) la descripción de su resolución, que coincide con la idea de interpretación teórica de los resultados. De Mey aplica la caracterización de Radder (1996) al examen de los experimentos mentales. De acuerdo con este análisis es posible identificar tres "fases" de la experimentación:

(1) preparación, (2) interacción y (3) detección. En primer lugar, durante la preparación, el objeto y el aparato se preparan de acuerdo con el plan del experimento. Posteriormente, la interacción da como resultado la transferencia de información del objeto al aparato. Finalmente, la detección implica obtener la información "mediante la medición u observación de la propiedad relevante del aparato (Radder 1996: 11) Asimismo, Radder distingue entre (a) la descripción o interpretación teórica de cada fase a partir de (b) su realización material. Aunque es posible identificar interpretaciones de cada fase tomadas por separado, la descripción general del proceso experimental consiste en un argumento. Radder asume que, aunque tal argumentación no siempre es explícita en la práctica experimental, "si se le pregunta al experimentador, podrá o debería ser capaz de presentar una historia plausible sobre cómo se produce el resultado experimental" (1996: 12). La noción de "realización material" es empleada por Radder para describir la igualdad de acción a pesar de la diferencia de interpretación, noción que le permite a su vez, definir la reproducibilidad de la realización material de un experimento bajo diferentes interpretaciones teóricas. De estas consideraciones De Mey concluye que en los experimentos mentales es posible diferenciar la realización material de las diferentes interpretaciones posibles (De Mey, 2003: 76). Normalmente, las acciones que de acuerdo con la narrativa del experimento mental debe realizar el experimentador se expresan en un lenguaje común, no técnico. Por otra parte un experimento mental solamente adquirirá algún valor justificatorio si al menos su descripción, en las fases que Radder denomina preparación e interacción, *puede* presentarse de forma independiente de la teoría o, por lo menos, en un lenguaje teóricamente neutral. Sin embargo, a diferencia de los experimentos reales los resultados de un experimento mental solo pueden describirse teóricamente en términos de un argumento.

Esto aporta que los experimentos mentales son independientes de la teoría en cierto grado, por lo que los mismos casos pueden emplearse en contextos teóricos diferentes y para objetivos diferentes. Por consiguiente, para

determinar la función que el experimento mental en cuestión desempeña respecto de una hipótesis es fundamental identificar el objetivo cognitivo a partir del cual se ha planteado el experimento y especificar, en cuanto sea posible, el conjunto de supuestos involucrados en la formulación del experimento en un contexto histórico particular.

Otra de las cuestiones que es necesario retomar para esclarecer la cuestión de si los experimentos mentales pueden desempeñar funciones en el contexto de justificación, es la de su significación evidencial. Ante todo, es importante señalar que en las perspectivas empiristas y aprioristas sobre el funcionamiento de los experimentos mentales subyacen concepciones diferentes de la naturaleza de la evidencia.

Según la perspectiva empirista representada por Norton, el valor epistémico de los experimentos mentales es el mismo que el de los argumentos ordinarios. Si desempeñan algún papel en la elección de teorías, este es idéntico al del argumento al que en cada caso el experimento mental es equivalente (o en virtud del cual puede ser reconstruido); su confiabilidad depende del valor de verdad de las premisas y del tipo de inferencia empleado. Los experimentos reales, en cambio, involucran intervención directa en el mundo natural y por ello son capaces de exhibir el comportamiento de los fenómenos independientemente de las expectativas generadas por la teoría. Desde esta perspectiva, solo los experimentos ordinarios son capaces de proporcionar información decisiva para la elección de teorías, mientras que los experimentos mentales solo son capaces de reorganizar información empírica ya disponible. Los experimentos mentales que dan lugar a la realización de un experimento real tienen, desde este punto de vista, un valor epistémico adicional. Los experimentos mentales que carecen de este rasgo son capaces de mostrar la coherencia de una teoría o el alcance de un principio, pero carecen de cualquier valor evidencial.

En general, la explicación empirista del funcionamiento de los experimentos mentales conserva algunas de las ideas centrales de la filosofía analítica de la experimentación. Desde este punto de vista, aunque la historia, la sociología y la tradición configuran la actividad experimental, y aunque el mecanismo de auto vindicación de las disciplinas experimentales es ubicuo, es la intervención en la naturaleza la que decide la adecuación de las representaciones. En consecuencia, para la perspectiva empirista de Norton, los experimentos mentales son incapaces de funcionar como experimentos cruciales o como experimentos decisivos, es decir, son incapaces de poner a prueba simultáneamente teorías rivales o proporcionar evidencia considerada decisiva para la aceptación o el rechazo de una teoría.⁴⁵

Recientemente, French y Murphy (2021) han sostenido que los experimentos mentales pueden generar predicciones e incluso pueden producir fenómenos. Proponen revisar las nociones empleadas para conceptualizar la novedad en ciencia y señalan que los experimentos mentales y las simulaciones computacionales pueden producir resultados de naturaleza disruptiva, que son valiosos en tanto violan las expectativas de la teoría y son difíciles de incorporar a las estructuras de conocimiento existentes.

Brown, como representante de la perspectiva apriorista, asume que la significación evidencial es una propiedad intrínseca del experimento (McAllister 1996: 234). Para Brown, los fenómenos son modos universales y estables en los cuales está articulado el universo y, por tanto, los experimentos mentales son capaces de proporcionar evidencia a favor de una hipótesis científica porque pueden exhibir dichos fenómenos con independencia de los datos sensibles. Debido a los supuestos metafísicos de la posición de Brown, la significación

⁴⁵ De acuerdo con la distinción introducida en Cassini (2015: 131).

evidencial de algunos experimentos mentales es independiente del contexto argumentativo en el cual fueron presentados.

En contraste con esta perspectiva puede situarse a la concepción *historicista* de la significación evidencial, de acuerdo con la cual los méritos de un experimento concreto son el resultado de logros históricos y locales. McAllister sugiere en esta línea que la fuerza persuasiva de los experimentos mentales es un logro conceptual e históricamente localizado. De acuerdo con McAllister:

Que un experimento dado es relevante para establecer y desacreditar una afirmación determinada, en otras palabras, que un experimento dado tenga significación evidencial para una afirmación determinada, depende en parte de los supuestos de fondo. (McAllister, 1996: 235)

Las nociones de evidencia experimental supuestas por las perspectivas empirista de Norton y apriorista de Brown implican que o bien los experimentos mentales son epistémicamente irrelevantes en la elección de teorías o bien algunos experimentos mentales tienen poderes epistémicos independientes del contexto de discusión en el que fueron propuestos. Ambas posiciones pueden ser matizadas si se atiende al contexto de formulación de algunos experimentos mentales y al hecho de que algunos supuestos teóricos fueron revisados a partir de su formulación. Siguiendo a Lakatos, es plausible sostener que no existe racionalidad instantánea en la historia de la ciencia y con ello afirmar que los experimentos cruciales (o decisivos) solo lo son históricamente. Para Lakatos: “experimento crucial es un título honorífico que puede conferirse a ciertas anomalías pero sólo mucho después del evento, sólo cuando un programa ha sido rechazado por otro” (Lakatos, 1971: 27). Las observaciones de Lakatos, y la prevalencia de la tesis Duhem-Quine, derivaron en un consenso amplio entre los filósofos de la ciencia de acuerdo con el cual no es posible explicar el abandono de una teoría científica y la aceptación de otra sólo en virtud de la verificación o refutación de sus respectivas consecuencias observacionales.

A partir de estas consideraciones, es posible limitar la tesis empirista de Norton sobre el papel de la experimentación real en la elección de teorías. La significación evidencial de los resultados experimentales debe ser enmarcada en un contexto de discusión, que incluye la especificación de los supuestos e hipótesis auxiliares admitidas, la enunciación de su objetivo cognitivo y la reconstrucción de una trama argumental e interpretativa en la que se inserta. Esta conclusión puede extenderse a los experimentos mentales.

Por su parte, la tesis apriorista sobre el valor evidencial de los experimentos mentales también puede ser revisada a partir de las consideraciones de Lakatos. Un resultado experimental se interpreta siempre a partir de un conjunto de supuestos que deben ser compartidos y accesibles a los representantes de las teorías rivales para que dicho resultado sea relevante como evidencia a favor o en contra de alguna hipótesis. En esta línea, Stöltzner (2004: 253) sostiene que el examen detallado de casos históricamente relevantes de experimentos mentales puede mostrar cómo estos casos lograron introducir transformaciones en los programas de investigación, incluso si se afirma que tales transformaciones no son tan radicales como las que los exámenes tradicionales les han atribuido. De acuerdo con su análisis, un experimento mental puede entrar en el núcleo duro de un programa de investigación después de nuevos pasos de refinamiento mediante pruebas y refutaciones. Brown interpreta el experimento de Galileo de los cuerpos en caída libre como un caso de refutación de la hipótesis aristotélica y vindicación de la tesis galileana (es decir, como un experimento crucial). Stöltzner, en cambio, señala que luego de hacer algunas precisiones sobre los supuestos involucrados en el clásico ejemplo de Galileo:

[...] El experimento mental arroja solo un contraejemplo heurístico que muestra que la teoría aristotélica no refleja la diferencia entre, ahora en términos newtonianos que Galileo dejó de lado deliberadamente, la causa de la aceleración y otros factores como la resistencia del medio. En esta dialéctica de contraejemplos encontramos en acción una reorganización conceptual. Galileo

reclasificó los modelos a los que se aplica la teoría y desarrolló nuevos conceptos básicos. (Stöltzner, 2004: 253)

En síntesis, la significación evidencial de los experimentos mentales no puede ser evaluada independientemente del contexto teórico en el que fueron formulados. Su capacidad de impulsar una revisión de principios teóricos, o incluso de proporcionar apoyo a una hipótesis, está indisolublemente ligada al objetivo cognitivo con el que fueron diseñados. Esta clase de experimentos puede ampliar nuestro conocimiento sobre el mundo natural en los casos en los que se intenta aplicar algún concepto o hipótesis a situaciones no experimentadas, que involucran contrafácticos o que dan lugar a consecuencias altamente antiintuitivas.⁴⁶

Otra razón que se ha invocado para limitar las funciones de los experimentos mentales se centra en la diferencia en número que existe entre experimentos mentales y reales. Si bien algunos experimentos mentales han tenido un considerable impacto en la formulación, aceptación y crítica de las teorías científicas, es un hecho que en comparación con los experimentos ordinarios, la apelación a experimentos mentales es menos frecuente. Esta afirmación es discutible en sí misma por lo que es necesario hacer algunas aclaraciones.

En primer lugar, es importante reconocer que la evaluación comparativa entre el número de experimentos mentales y reales depende del criterio de identificación para los experimentos. En el caso de los experimentos mentales, subsiste un profundo desacuerdo en este punto y, como consecuencia, los catálogos de experimentos mentales existentes presentan variaciones

⁴⁶ Un desarrollo de esta idea puede encontrarse en Kühne (2019): “Kuhn y Lakatos elaboraron esta idea señalando que incluso las definiciones de conceptos tienen un contenido cuasi-empírico incorporado. Este contenido cuasi-empírico es el alcance de las aplicaciones pretendidas del concepto. Para aprender qué cuenta como una aplicación prevista, uno puede construir escenarios imaginarios y luego decidir si en este escenario el concepto todavía se puede aplicar o no (Kühne, 2009: 15)”.

considerables. Como ejemplos de estas variaciones pueden considerarse los listados elaborados por McAllister (2012: 18) y Ruiz Sosa (2012). De acuerdo con el primero, en los inicios de la ciencia moderna y hasta el siglo XIX encontramos un buen número de experimentos mentales. Entre los ejemplos canónicos se cuentan los llevados a cabo por Galileo, Descartes, Newton, Leibniz, Stevin y Huygens. McAllister (2012: 18) reconoce en el mecanicismo dominante del siglo XVII dos perspectivas opuestas, enraizadas en el papel que se le atribuyó a la imaginación en el razonamiento científico. Ambas produjeron un buen número de experimentos mentales pero, en opinión de McAllister, las funciones que se les atribuyeron fueron muy diferentes. En el siglo XIX la escuela molecular mecánica francesa conservó una actitud optimista respecto del papel de la imaginación en la generación de hipótesis científicas y dio lugar a algunos experimentos mentales importantes como el del demonio de Laplace (McAllister, 2012: 22). Una actitud similar puede reconocerse en la construcción de modelos por parte de Maxwell, quien atribuyó a la imaginación la virtud de completar o dar sentido a la observación. Por su parte, Ruiz Sosa (2012: 373-376) identifica 57 experimentos mentales relevantes en el desarrollo de la física cuántica entre 1927 y 1936. Dependiendo del criterio de identidad que se utilice, puede considerarse que algunos de los ejemplos mencionados en este catálogo son repeticiones de otros experimentos mentales ya mencionados o que se trata de razonamientos hipotéticos que no constituyen casos de experimentos mentales. No obstante estas diferencias, es posible afirmar sin demasiada controversia que los experimentos mentales son menos frecuentes que los reales en la práctica científica.

Por otra parte, el análisis epistemológico de los experimentos mentales en física suele recurrir a un número limitado de ejemplos: el experimento de Galileo de los cuerpos en caída libre, el experimento del plano inclinado de Stevin, el experimento EPR, entre otros casos recurrentes. Esto puede conducir a la conclusión de que los experimentos mentales son escasos y de que su

contribución al conocimiento no es significativa ni en términos absolutos, debido a la escasez de ejemplares, ni en términos relativos, al compararlos con experimentos reales (Reiss, 2002: 6).

Contra esta posición, es posible afirmar que los experimentos mentales han realizado contribuciones significativas al conocimiento científico y que son filosóficamente interesantes por derecho propio. Como se ha mostrado mediante la presentación de casos en los capítulos anteriores, los experimentos mentales han contribuido a la formulación de conceptos y al planteo de problemas científicos; además, han tenido consecuencias en la revisión de teorías científicas y, como se mostrará más adelante en este capítulo, han desempeñado funciones relacionadas con la justificación de hipótesis científicas. La diferencia comparativa en el número de experimentos mentales y reales no es una razón suficiente para descartar su relevancia en el razonamiento científico. Por otra parte, los experimentos mentales científicos presentan características epistémicas que, si bien no son extraordinarias porque podemos encontrar claras similitudes entre su funcionamiento y el de otras prácticas científicas, son intrigantes desde el punto de vista filosófico, fundamentalmente por su capacidad de producir conocimiento nuevo sin introducir datos nuevos.

7.4 Clasificación de los experimentos mentales

En el capítulo anterior se argumentó que los experimentos mentales pueden concebirse como una clase particular de modelos científicos. Algunas de las perspectivas que tomamos como base de nuestra hipótesis les asignan a los modelos científicos un papel particular respecto de las teorías. La función más importante que los modelos pueden desempeñar es la de *mediadores* entre las teorías y los fenómenos. Particularmente en el contexto de la física, donde las teorías son demasiado abstractas y alejadas de la experiencia para ser aplicadas o contrastadas, los modelos proporcionan ejemplificaciones de las relaciones

entre los conceptos abstractos (Cartwright, 1983). Desde este punto de vista, los modelos tienen una función representativa esencial que no depende de la teoría (al menos no de una teoría determinada) y posibilitan la intervención en los fenómenos. Hay, asimismo diferentes maneras en las que los modelos son capaces funcionar como mediadores: pueden mediar entre una teoría y el mundo, ser una representación abstracta de un sistema físico gobernada por una o más teorías, o puede ser una representación concreta de algún aspecto abstracto del mundo. Un modelo puede funcionar también como mediador en su rol como el objeto de una investigación. En algunas ocasiones el modelo mismo, más que el sistema físico, se vuelve objeto de investigación. El modelo sirve como fuente mediata de conocimiento en ocasiones en las que el sistema físico inaccesible o el acceso al mismo es limitado, por lo que solo sabemos cómo se comporta el modelo en ciertas circunstancias. El modelo de quarks y los modelos cosmológicos son casos de estos usos de los modelos (véase Morrison 2008: 68-88). Morrison retoma la distinción de Cartwright y la refina distinguiendo entre modelos representativos y mediadores. Los modelos representativos actúan como mediadores entre la teoría y el mundo y funcionan como la fuente del conocimiento mediato. Los modelos representativos suplantando al sistema físico como el objeto tratado por el modelo interpretativo de la teoría.

Se ha argumentado que los experimentos mentales gozan de autonomía respecto de las teorías al menos en dos aspectos: no son subsidiarios de la teoría en cuanto tienen la capacidad de desempeñar diversas funciones y pueden ser replicados en contextos teóricos diferentes. Cuando se los interpreta como modelos científicos, cumplen funciones análogas: pueden mediar entre un sistema físico y el mundo, representar algún aspecto concreto de un sistema, ser el objeto de la investigación científica en ocasiones en las cuales el acceso sistema es limitado o este es inaccesible. Se ha argumentado también que, al menos en un sentido limitado, los experimentos mentales son capaces de proporcionar conocimiento nuevo a partir de la presentación de casos no

experimentados (en ocasiones imposibles de experimentar por razones físicas o técnicas) a los cuales aplicar principios teóricos. El valor evidencial que estos puedan tener se define dentro de un contexto teórico particular en el cual el experimento activa ciertas intuiciones. Explicitar lo más exhaustivamente posible este contexto de supuestos permite actualizar su valor heurístico y evaluar, retrospectivamente, su valor epistémico.

La tipificación de los experimentos mentales que proponemos se centra en los usos que estos han tenido en el contexto del razonamiento científico, atendiendo a consideraciones relativas al estado del conocimiento en el momento de su formulación y al objetivo cognitivo proyectado en su formulación original o en alguna de sus versiones. La distinción entre usos *heurísticos* y *críticos* puede interpretarse en continuidad con la distinción entre los contextos de descubrimiento y justificación trazada por la filosofía tradicional de la ciencia, aunque no tiene la pretensión de establecer distinciones excluyentes. Entre los *usos heurísticos* encontramos los empleos orientados a *plantear un problema*, esto es, a delimitar y precisar un problema teórico (llamando la atención sobre una cuestión importante que, de lo contrario, podría no haber sido evidente), los *usos ilustrativos* que se dirigen a esclarecer, divulgar, mostrar la aplicación de principios teóricos y los *usos exploratorios* en los que se busca ofrecer explicaciones potenciales ideando escenarios que, si fueran reales, darían lugar al fenómeno que constituye el *explanandum* en cuestión. Entre los *usos críticos* se encuentran los *usos apologéticos* que se proponen dar apoyo evidencial a una teoría y los *usos refutatorios* que se proponen exhibir contradicciones, derivar inconsistencias o mostrar la inaplicabilidad de un principio teórico.

Usos heurísticos	Usos para plantear un problema
	Usos Ilustrativos
	Usos exploratorios
Usos críticos	Usos apologéticos
	Usos refutatorios

Figura 7.5: funciones de los experimentos mentales.

7.4.1 Usos heurísticos

La atribución de funciones heurísticas e ilustrativas a los experimentos mentales ha sido uno de los aspectos menos controvertidos en la discusión de sus poderes epistémicos. La función heurística se ha asociado principalmente con el descubrimiento de nuevas hipótesis. En este sentido, los experimentos mentales pueden desempeñar una gama de papeles como perfeccionar una hipótesis, mejorar conjeturas, plantear adecuadamente un problema o llamar la atención sobre algún fenómeno o propiedad de un sistema que ha pasado inadvertida hasta el momento. El uso ilustrativo de los experimentos mentales se vincula específicamente con los intentos de ejemplificar el alcance de una teoría presentando un escenario imaginario que muestra el comportamiento de algún fenómeno de acuerdo a las hipótesis estipuladas.

El uso heurístico de un experimento mental no limita su empleo con otras finalidades. El caso del experimento de Einstein, en el que se imagina corriendo a la velocidad c junto a un rayo de luz, puede ser interpretado como un experimento que pretende llamar la atención sobre un problema teórico:

Después de diez años de reflexión, tal principio resultó de una paradoja que ya había descubierto a la edad de dieciséis años: si persigo un rayo de luz con la velocidad c (velocidad de la luz en el vacío), debería observar tal rayo de luz como un campo electromagnético espacialmente oscilante en reposo. Sin embargo,

parece que no existe tal cosa, ya sea sobre la base de la experiencia o según las ecuaciones de Maxwell. Desde el principio me pareció intuitivamente claro que, juzgado desde el punto de vista de tal observador, todo tendría que suceder de acuerdo con las mismas leyes que para un observador que, en relación con la tierra, estaba en reposo. Porque, de lo contrario, ¿cómo debería saber el primer observador, es decir, ser capaz de determinar, que se encuentra en un estado de movimiento tan uniforme?

Se ve que en esta paradoja ya está contenido el germen de la teoría de la relatividad especial. Hoy todo el mundo sabe, por supuesto, que todos los intentos de esclarecer satisfactoriamente esta paradoja estaban condenados al fracaso mientras el axioma del carácter absoluto del tiempo, es decir, de la simultaneidad, estuviera anclado sin reconocerlo en el inconsciente. Es evidente que reconocer este axioma y su carácter arbitrario ya implica realmente la solución del problema. El tipo de razonamiento crítico que se requirió para el descubrimiento de este punto central fue impulsado decisivamente, en mi caso, especialmente por la lectura de los escritos filosóficos de David Hume y Ernst Mach. (Einstein, 1949: 53)

Aunque el experimento mental de Einstein, descrito en sus *Notas autobiográficas* (1949), suele considerarse como un contraejemplo dirigido a la teoría de Maxwell (1865), de acuerdo con la cual la luz es una onda electromagnética que se propaga con una velocidad constante, no es claro que el experimento fuera formulado para refutar esta tesis. Norton (2012: 124) propone que tomando en cuenta información histórica relevante (en el momento en el que Einstein sitúa la primera formulación del experimento, alrededor de 1896, no conocía aún la teoría electromagnética de Maxwell) y considerando las repeticiones (o nuevas narrativas) del experimento posteriores a la formulación de la Relatividad Especial, puede considerarse que el experimento motivó el abandono de las teorías de la emisión por parte de Einstein. Pero esta es una conclusión que Norton extrae de forma laboriosa recomponiendo supuestos adicionales que fueron introducidos en las sucesivas reformulaciones e interpretaciones del experimento. De acuerdo con esta lectura, el mérito del caso reside en que le permitió a Einstein derivar una serie de fallas fatales en la teoría de la emisión. Las tres objeciones que se pueden derivar de la observación de una onda estacionaria contra la teoría electrodinámica del éter son: a) que el

fenómeno que se observaría en tales condiciones no es posible en base a la experiencia; b) que el fenómeno no podría ser previsto por las ecuaciones de Maxwell; y c) que el fenómeno no es posible para la mecánica clásica (el fenómeno debería responder a las mismas leyes desde el punto de vista de un observador en reposo). De acuerdo con Norton (2012: 126), estas objeciones pueden responderse dentro del marco de la teoría de Maxwell, pero son certeras respecto de la teoría de la emisión dado que: a) implican que ninguna fuente luminosa en el universo se aleja de nosotros a la velocidad c y con ello, aparece como una onda de luz congelada; b) una teoría de la emisión que hace posible que el rayo de luz aparezca como una onda congelada contradice a las ecuaciones de Maxwell; y c) en una teoría de la emisión, el movimiento a la velocidad c es respecto de una fuente luminosa, pero no es posible determinar esta velocidad respecto del estado de la onda congelada (lo que implica que la teoría no puede establecer de qué manera el estado presente determina los estados futuros).

Para Brown (2011: 33) por su parte, el caso puede ser clasificado como un experimento mental destructivo porque claramente presenta un fenómeno que si pudiera obtenerse refutaría la teoría de Maxwell. Brown (2011: 7) presenta una versión simplificada del problema, reduciendo la cuestión a una hipótesis (de acuerdo con la cual la luz consiste en la oscilación de un campo electromagnético) y un fenómeno que refuta sus predicciones (la onda estacionaria que puede observar el sujeto que corre a la velocidad de la luz). Uno de los aportes de Norton permite contradecir la conclusión de Brown. De acuerdo con el detallado examen de Norton, el fenómeno descrito por Einstein no se produciría en las condiciones propuestas, dado que “un observador que se mueve rápidamente estaría rodeado por luz congelada.” (Norton, 2012: 126).

El experimento ha sido objeto de una larga controversia y no es posible afirmar de manera definitiva que fue usado con fines refutatorios sin adicionar un conjunto de supuestos que fueron asumidos por Einstein en diferentes momentos de su desarrollo teórico. No obstante, es claro que el caso llama la atención sobre

un problema teórico. De acuerdo con Einstein, si fuera posible para un observador viajar en paralelo a un rayo de luz, a la misma velocidad, no vería una onda propagándose sino una onda estacionaria. Sin embargo, esto no puede ser así, ya que el cambio es esencial para una onda de luz; si el campo eléctrico o magnético es estático, no dará lugar al otro y, por tanto, no habrá onda electromagnética. El valor heurístico del caso se puede apreciar en el relato del propio Einstein que identifica en este experimento el “germen de la relatividad”, particularmente de la hipótesis de acuerdo con la cual la velocidad de la luz debe ser una constante universal, para todos los observadores, independientemente de si se mueven respecto de la fuente luminosa.

Los experimentos mentales categorizados por Brown (2011) como experimentos mediadores responden a los usos que denominamos ilustrativos. El experimento de Newton sobre las agujas que se mantienen sobre sus puntas en equilibrio estático también tiene el propósito de ilustrar un principio teórico. Según la cosmología Newtoniana, las estrellas se encuentran distribuidas de manera uniforme por todo el universo a diferentes distancias. Suponiendo, como inicialmente lo hiciera Newton, que el universo es espacialmente finito y que los cuerpos experimentan una atracción gravitacional directamente proporcional a sus masas e inversamente proporcional al cuadrado de sus distancias, la materia esparcida por el universo tendería a colapsar hacia su centro. El problema del colapso gravitatorio fue planteado por primera vez por el reverendo Bentley en su correspondencia con Newton de 1692. La paradoja establecía que en un universo finito poblado de estrellas que se atraen mutuamente, éstas deberían acercarse cada vez más entre sí, por lo que el universo debería colapsar.

Asumiendo un universo infinito, Newton propuso una solución posible a este problema. Bajo este supuesto, si la atracción gravitatoria entre las diferentes estrellas decrece como el inverso del cuadrado de la distancia que las separa y el número de estrellas se incrementa proporcionalmente al cuadrado de dicha distancia, ambos efectos se compensan y el universo se mantiene en equilibrio.

La solución de Newton al problema propuesto por Bentley incluye, además, el supuesto de que las estrellas se encuentran distribuidas de manera uniforme por todo el espacio, de modo que la fuerza neta sobre cada estrella es exactamente nula, ya que las infinitas fuerzas atractivas se cancelan mutuamente.

La razón por la que la materia regularmente esparcida por un espacio finito se reuniría en el centro la concibe usted de la misma manera que yo. Pero que tuviese que haber una partícula central tan exactamente colocada en el centro como para ser siempre atraída por igual por todos sus lados, y por tanto en un continuo reposo, me parece una suposición al menos tan difícil como hacer que la aguja más puntiaguda se sostenga con su punta sobre la superficie de un espejo. Ya que si el centro matemático preciso de la partícula central no estuviese exactamente en el mismo centro matemático del poder de atracción de la masa entera, la partícula no sería atraída por todos sus lados por igual. Y mucho más difícil que esto es suponer que todas las partículas en un espacio infinito estuviesen tan exactamente equilibradas entre sí como para permanecer inmóviles en perfecto equilibrio. Creo que esto es tan difícil como hacer, no que una aguja, sino que un número infinito de ellas (tantas como partículas hay en un espacio infinito) permaneciesen perfectamente en equilibrio sobre sus puntas. Sin embargo, reconozco que esto es posible, al menos para un Poder Divino. Y si hubiesen sido colocadas así alguna vez, estoy de acuerdo con usted en que continuarían sin movimiento alguno en esa posición, a menos que el mismo poder les imprimiese un nuevo movimiento. (Newton a Bentley, January 17, 1693; en Janiak, 2014: 124).

La solución al problema del colapso gravitatorio, en principio motivado por la sugerencia de Bentley, incluye la hipótesis de que las estrellas se encuentran en reposo en un delicado equilibrio, tan preciso que las fuerzas atractivas se cancelan mutuamente. Para ilustrar la altísima improbabilidad de esta situación, Newton sugiere que imaginemos la dificultad de parar a una aguja sobre un espejo, en equilibrio sobre su punta y que amplifiquemos esa dificultad para lograr el equilibrio de un número infinito de agujas (Fig. 7.6). El uso ilustrativo de este experimento es evidente. La dificultad de obtener una situación como la que la narrativa del experimento ilustra se vincula con la observación de Newton al final del texto, de acuerdo con la cual, solo un poder divino podría producir este fenómeno. No obstante, aunque altamente improbable, la situación descrita en

el experimento mental resulta físicamente posible de acuerdo con la mecánica clásica.

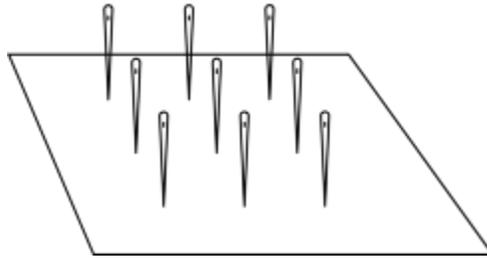


Figura 7.6

Una de las características más sobresalientes de los usos heurísticos de los experimentos mentales, presente en los dos casos revisados, es la pregnancia de las imágenes que sus relatos comunican. La vivacidad de las representaciones que los lectores pueden formarse a partir de estos relatos no es directamente proporcional a la perspicuidad de las hipótesis que se proponen ilustrar o a los problemas teóricos que se proponen plantear. Lo que resulta obvio es su poder sugestivo. Por consiguiente, aunque no esté clara la secuencia inferencial que sugieren, su papel impulsor de la revisión de hipótesis en la reflexión científica es innegable.

Así como Kekulé imaginó una estructura física, los experimentos mentales como la persecución del haz de luz nos llevan a imaginar una estructura de argumento; es decir, no siempre constituyen un argumento, pero sugieren uno. Hay una sensación de sorpresa agradable, de certeza inmediata, [...] un momento en el que obtienes el truco que resuelve un problema. [...] En sí mismo, no es un argumento sino una percepción que sugiere un argumento. (Peacock, 2018: 266)

Los usos exploratorios de los modelos, fundamentalmente como instrumentos para la comprensión en dominios en los que la teoría no se encuentra suficientemente establecida o desarrollada, ha sido recientemente examinada como una función que, si bien está íntimamente ligada a la explicación científica, es independiente o incluso previa a ésta. Del mismo modo que se ha

admitido que los experimentos reales pueden cumplir otras funciones además de las confirmatorias, se afirma que los modelos pueden contribuir en los procesos de integración de información y sondeo de hipótesis científicas. De manera análoga a los experimentos exploratorios, los modelos pueden iluminar aspectos de fenómenos complejos singularizando una gran masa de datos y sugiriendo “lugares donde buscar” regularidades (Gelfert, 2016: 80), proporcionando así un punto de partida para investigaciones futuras. Además, pueden ofrecer explicaciones potenciales ideando escenarios que, si fueran reales, darían lugar al fenómeno que constituye el *explanandum* en cuestión. Algunos experimentos mentales han sido usados con este propósito. El experimento que Newton emplea para introducir la noción de espacio absoluto puede considerarse como un ejemplo de este uso. En este caso, el fenómeno en cuestión (el movimiento absoluto) es evidencia de la hipótesis de la existencia del espacio absoluto.

Newton postula la existencia del espacio absoluto en el *Scholium* de los *Principia*, luego de introducir las definiciones. Allí presenta las nociones fundamentales de su sistema físico. Además de espacio y tiempo, fuerza y cantidad de materia (masa) son los conceptos centrales. De acuerdo con Newton, las personas comunes conciben estas cualidades bajo las nociones de relaciones que se dan entre los objetos sensibles. Ciertos prejuicios surgen de esta concepción que es necesario eliminar y distinguir entre absoluto y relativo, verdadero y aparente, matemático y común. El espacio relativo es alguna relación móvil o medida del espacio absoluto, que nuestros sentidos determinan por la posición de los cuerpos y que es tomada por el espacio absoluto. Dado que las partes del espacio no pueden ser vistas o distinguidas unas de otras por nuestros sentidos, en su lugar usamos mediciones sensibles para ellos. De la posición y la distancia de las cosas, desde cualquier cuerpo considerado como inmóvil, definimos todos los lugares y con respecto a tales lugares estimamos los movimientos considerados como la transferencia de los cuerpos de un lugar hacia otro. En vez de lugares y movimientos absolutos utilizamos los relativos, por lo

cual, es perfectamente posible que no haya ningún cuerpo en reposo respecto del cual se puedan determinar los movimientos y lugares absolutos de otros cuerpos; en síntesis, todos estos movimientos relativos pueden formar un sistema coordinado en movimiento. La postulación del espacio absoluto ha sido presentada como una necesidad lógica y ontológica para la física newtoniana (Jammer, 1952: 100). De su existencia depende la verdad de la primera ley de movimiento. El movimiento rectilíneo uniforme requiere un sistema de referencia diferente de cualquier posición arbitraria relativa al espacio. El estado de reposo, para Newton, presupone un espacio absoluto.

El experimento mental del balde en rotación presenta un escenario en el cual es posible apreciar el efecto de fuerzas y aceleraciones desligados de cualquier sistema material de referencia. En estas condiciones, deberíamos concluir que el movimiento no es relativo sino absoluto. Newton propone imaginar que en el universo sólo existe un balde lleno en parte por agua. El balde es suspendido de una cuerda (no importa de qué esté sostenida) que es retorcida y luego liberada. En este escenario:

[...] la superficie del agua será, al principio, plana, al menos hasta que el balde se empiece a mover; pero luego de eso, el balde le comunica gradualmente su movimiento al agua lo que lo hará comenzar girar sensiblemente y retroceder de a poco desde el centro y ascender por las paredes del balde formando una figura cóncava. (Newton, [1686] 1976: 10)

A medida que el movimiento relativo del agua respecto del balde decrece el circular del agua se incrementa. Este es movimiento respecto del espacio absoluto, no relativo a los elementos del sistema.

Este caso ha sido discutido en numerosas ocasiones en la literatura de los experimentos mentales. Los diferentes análisis contemporáneos del experimento coinciden en que la mejor explicación del fenómeno (la figura cóncava que se produce en el agua) es la acción de la enorme masa de las estrellas fijas. No obstante, de acuerdo a la narrativa de Newton, la curva que se forma en la

superficie del agua que contiene el balde no puede ser considerada como relativa respecto de otros objetos. Recordemos que la descripción del escenario indica que el balde se encuentra suspendido de una cuerda rotando en un universo en donde sólo existe el balde y la cuerda. Este es, desde nuestro punto de vista, un caso de uso exploratorio. El experimento presenta un fenómeno visualizable, irrealizable si se atiende estrictamente a la descripción de Newton y que presenta un fenómeno que no puede encontrarse en la naturaleza (por esto nos referimos a la figura cóncava que se produce en el agua en ausencia de cualquier masa). De este modo el experimento del balde tal como fuera formulado por Newton, presenta un *explanandum* para el cual la mejor explicación es según Newton, la existencia postulada del espacio absoluto.

7.4.2 Usos críticos de los experimentos mentales. Su rol en la elección de teorías

Las funciones refutatorias de los experimentos mentales han sido asociadas con los argumentos de reducción al absurdo. En general esta función resulta poco controversial. Los experimentos mentales son capaces de presentar escenarios imaginarios que si fueran reales, constituirían un contraejemplo para una hipótesis científica. Al proporcionar un contraejemplo muestran las limitaciones de un marco explicativo o la inaplicabilidad de un modelo. Aunque, basándose en consideraciones históricas, no puede afirmarse que los experimentos mentales hayan sido decisivos para la falsación de teorías, es posible identificar episodios en los cuales motivaron la revisión de principios teóricos, su modificación y en algunos casos, cumplieron un papel importante en los episodios de cambio conceptual.

Para la perspectiva apriorista de Brown, los experimentos mentales son capaces de justificar hipótesis presentando fenómenos que ejemplifican principios teóricos. Algunos casos excepcionales pueden refutar una hipótesis y

simultáneamente proporcionar evidencia confirmatoria para una hipótesis rival, es decir, desempeñar la función de experimentos cruciales. Este es el caso del experimento mental de los cuerpos en caída libre de Galileo, del experimento de la *Vis Viva* de Leibniz y del experimento EPR. Según la interpretación de Brown, el fenómeno presentado en estos casos presenta simultáneamente evidencia en contra de una hipótesis y a favor de otra. La oscuridad de este vínculo y el conjunto de compromisos metafísicos y epistémicos involucrados en la explicación del mecanismo de adquisición de nuevo conocimiento ha sido uno de las principales razones para rechazar la idea de que los experimentos mentales puedan cumplir funciones confirmatorias. Las principales dificultades consisten en mostrar que los experimentos mentales proporcionan evidencia a favor de una hipótesis, en el mismo sentido en que los experimentos reales lo hacen. No obstante, los casos presentados como experimentos mentales platónicos en su definición, desempeñaron papeles negativos y pueden ser considerados como ejemplos de los usos refutatorios de los experimentos mentales. Este es el caso del experimento de la *vis viva* de Leibniz

El experimento de Leibniz pertenece, según Brown, a la categoría de experimento mental *platónico* porque se propone refutar la tesis cartesiana acerca de la conservación de cantidad de movimiento y demostrar el principio de conservación de la magnitud física que Leibniz llama *vis viva*. El caso se enmarca en la controversia que Leibniz sostuvo con los cartesianos acerca de lo que él refirió como “conservación de la fuerza motriz” y se encuentra plasmado en su opúsculo⁴⁷ para el *Acta Eroditorum* de marzo de 1648. De acuerdo con Leibniz, es razonable suponer que la fuerza motriz se conserva dado que no puede ser disminuida sin ser transferida a otro cuerpo ni incrementada sin un nuevo impulso. No obstante, sostiene Leibniz, la relación establecida por los cartesianos entre “fuerza motriz” y “cantidad de movimiento” como medida para esta fuerza es

⁴⁷ El nombre del opúsculo es: “Una breve demostración de un notable error de Descartes y otros acerca de una ley natural”.

inadecuada. Con el fin de mostrar esta inadecuación, presenta un escenario que parte de un supuesto que un cartesiano estaría dispuesto a aceptar, a saber: “Un cuerpo que cae desde una cierta altura adquiere la misma cantidad de fuerza [motriz] que la que sería necesaria para elevarlo de nuevo a su altura original” (Leibniz, [1648] (1969): 296). El experimento continúa con la descripción de un escenario, acompañado de un gráfico, en el que se representan dos cuerpos uno de 1 kg⁴⁸ elevado a 4 m y uno de 4 kg elevado a 1 m. Se establece entre ambos la siguiente relación: la misma fuerza es necesaria para elevar 4 m un cuerpo de 1 kg que para elevar 1 m un cuerpo de 4 kg (Fig. 7.7). De esta segunda hipótesis, se sigue que un cuerpo de 1 kg al caer desde una altura de 4 m adquiere la misma cantidad de fuerza que un cuerpo de 4 kg al caer desde una altura de 1 m” (Leibniz, [1648] (1969): 297). De acuerdo con Leibniz, si se aplica la ley de caída de los cuerpos de Galileo, según la cual, la distancia recorrida por un cuerpo que cae es directamente proporcional al cuadrado del tiempo de caída, podemos ver que la “cantidad de movimiento” no es la misma. Los cartesianos suponían que la cantidad conservada es el producto de la masa y la velocidad ($m.v$). Sin embargo, valiéndose de la Ley de Galileo y la definición del módulo de la velocidad (la rapidez), puede calcularse cada caso del experimento propuesto. El cuerpo de 4 kg adquiere una rapidez de 1 m/s al caer 1 m y el cuerpo de 1 kg adquiere una rapidez de 2 m/s al caer 4 m. Con estos valores puede determinarse que la cantidad de movimiento del cuerpo de 4 kg que cae 1 m es de 4 unidades, dado que $4 \text{ kg} \times 1 \text{ m/s} = 4 \text{ kgm/s}$, y la del cuerpo de 1 kg que cae 4 m de 2 unidades, dado que $1 \text{ kg} \times 2 \text{ m/s} = 2 \text{ kgm/s}$. De estos valores se sigue que la “fuerza” (magnitud o cantidad) que se conserva no es la cantidad de movimiento. Está implícito en la Ley de Galileo que si la distancia de la caída es proporcional al cuadrado de su velocidad, entonces, lo que se conserva en un caso como el del experimento es la *vis viva* representada por la relación $m.v^2$. Así un cuerpo de 4 kg

⁴⁸ En el experimento de Leibniz la masa está expresada en libras.

que cae a una velocidad de 1 m/s ($4 \times 1^2 = 4$) conserva la misma *vis viva* que un cuerpo de 1 kg que cae con una velocidad de 2 m/s ($1 \times 2^2 = 4$).

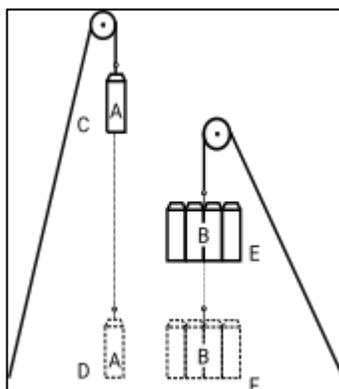


Figura 7.7

De acuerdo con Brown, “con un simple y elegante ejemplo Leibniz destruye la perspectiva cartesiana y establece la propia” (Brown, 2011: 42). En la primera parte del experimento Leibniz refuta la interpretación cartesiana de acuerdo con la cual lo que se conserva es la cantidad de movimiento. A continuación, establece que la fuerza conservada es la *vis viva*, que calcula valiéndose de la Ley de Galileo y la definición del módulo de la velocidad. Según Brown, debido a que no existen elementos adicionales que puedan identificarse (datos nuevos, mecanismos de inferencia) como razones para postular la nueva teoría, debemos asumir que la explicación más razonable de lo que ocurre en este caso excepcional, es la captación intuitiva del principio de conservación de la *vis viva*. Sin embargo, Brown no consigue proporcionar una explicación satisfactoria del caso ya que no logra especificar en qué momento del experimento tiene lugar la intuición platónica del principio en cuestión.

Una interpretación alternativa del experimento de Leibniz es que se trata de un cálculo basado en la teoría, que depende de la introducción de un principio teórico relevante. Según esta interpretación, el experimento mental tal como se encuentra formulado en el opúsculo de Leibniz responde al uso refutatorio. La

visualización del escenario solo cumple un papel heurístico en la obtención del resultado. La primera parte del experimento podría ser reconstruida considerando las siguientes premisas:

(1) La cantidad de fuerza que un cuerpo adquiere al caer desde cierta altura es igual a la cantidad de fuerza que sería necesaria para elevar ese cuerpo a la misma altura.

(2) La cantidad de fuerza conservada es el producto de la masa por la velocidad.

(3) La cantidad de fuerza que un cuerpo de 1 kg adquiere al caer desde una altura de 4 metros es igual a la cantidad de fuerza que un cuerpo de 4 kg adquiere al caer desde una altura de 1 metro.⁴⁹

(4) La distancia recorrida por un cuerpo que cae es directamente proporcional al cuadrado del tiempo que cae (Ley de Galileo).

La premisa (2) es el supuesto cartesiano de acuerdo con el cual la fuerza conservada es la cantidad de movimiento. Leibniz muestra que (2) es falso en virtud del cálculo que se realiza empleando (4) para determinar la cantidad de movimiento. Este cálculo aniquila la hipótesis cartesiana (2). La presentación del escenario en (3), en el que se suponen dos cuerpos suspendidos con las características mencionadas, tiene el fin de presentar el fenómeno que se encuentra subsumido al principio teórico, a saber, la conservación de la fuerza. Al introducir la ley de Galileo, Leibniz puede mostrar que la cantidad de fuerza conservada no es el producto $m.v$, sino el producto $m.v^2$. Leibniz postula la relación adecuada para calcular la fuerza conservada pero las razones para su postulación pertenecen a la heurística del descubrimiento y no se siguen de la presentación del escenario imaginario.

⁴⁹ Tomo las premisas 1, 2 y 4 de la formulación de McDonough (2014).

Las funciones confirmatorias de los experimentos mentales resultan mucho más polémicas. Dada la ausencia de intervención en el mundo natural resulta difícil sostener, sin comprometerse con una metafísica inflacionaria, que esta clase de experimentos pueda aportar evidencia a favor de una hipótesis. No obstante, en muchos casos han sido usados para apoyar teorías. Es posible afirmar que los experimentos mentales son capaces de desempeñar funciones en el contexto de justificación de las teorías físicas en los casos en los que proponen modelos fenoménicos que apoyan interpretaciones teóricas. Los modelos como mediadores satisfacen dos condiciones que los empiristas demandan de los experimentos que pudieran desempeñar un papel en la elección de teorías: a) guardan una relación con las observaciones y experimentaciones previas, de las que se sirven en la construcción de modelos fenoménicos, y b) su ejecución produce efectos (históricamente identificables) en la revisión de principios teóricos. Algunos de los casos que han cumplido este papel son los que señalaremos como casos de usos apoloéticos de los experimentos mentales.

Las principales objeciones a la idea de que los experimentos mentales puedan evidencia confirmatoria para las hipótesis científicas proceden de las interpretaciones empiristas de sus poderes epistémicos. Para estos puntos de vista, la única fuente de conocimiento incontrovertible sobre el mundo natural es la experiencia. Dado que lo que caracteriza a esta clase de experimentos es la ausencia de intervención en el mundo natural, su funcionamiento depende de alguna reorganización de los datos proporcionados por la experiencia. Este procedimiento coincide para la mayoría de los empiristas con la obtención de conclusiones a partir de premisas con contenido empírico a través de inferencias lógicas. Los experimentos mentales no son capaces de proporcionar nuevas evidencias confirmatorias porque no pueden proporcionar datos nuevos. Los ejemplos de experimentos mentales que amplían nuestro conocimiento sobre el mundo natural coinciden con instancias de razonamiento inductivo y su credibilidad depende de la preservación de la probabilidad. La creencia en el

resultado o conclusión del experimento mental está justificada solo en tanto el argumento reconstruido puede justificar la conclusión, es decir, proporcionarle apoyo inductivo (Norton, 2004: 50). Los razonamientos ampliativos involucran un salto inductivo, implícito en la presentación del experimento mental pero especificable como una premisa en su reconstrucción como argumento, en el que se afirma que el caso postulado por el escenario imaginario es típico y por lo tanto representativo de un conjunto de fenómenos sobre los cuales la conclusión realiza una afirmación (Norton, 1991: 131).

Para algunos defensores de la posición empirista, los experimentos mentales como metodología científica tienen algunos rasgos que permiten diferenciarlos de otras formas de conocimiento hipotético y que los acercan a la metodología científica. En esta línea, Irvine (1991) sostiene que los experimentos mentales que tienen lugar dentro de un contexto teórico bien definido se proponen identificar un número de variables independientes dentro del experimento para determinar las correlaciones entre variaciones de esas variables y un conjunto de variables dependientes usadas para caracterizar el resultado del experimento. Esta peculiaridad de los experimentos mentales es la razón principal de su utilidad en la metodología científica. Como instancias de la variación concomitante pueden ser considerados como un procedimiento regulado que, aunque desde el punto de vista empirista no posee el rasgo definitorio de la experimentación, es capaz de ampliar nuestro conocimiento. El empirismo machiano, que fue iniciador del análisis filosófico de los experimentos mentales, coincide en que este es un elemento definitorio de estas prácticas.

Ahora bien, retomando la discusión acerca de sus funciones, sostenemos aquí que los elementos señalados por Irvine como característicos de los experimentos mentales son indispensables para conciliar la exigencia empirista sobre el origen del conocimiento con la afirmación de que los experimentos mentales pueden proporcionar evidencia confirmatoria para las hipótesis científicas. Según Irvine, como los experimentos físicos, los experimentos

mentales deben estar en una relación privilegiada con las observaciones pasadas, es decir, con la evidencia disponible, y con una teoría de fondo razonablemente bien desarrollada. Este requisito es el que distingue a los experimentos mentales de otras formas de razonamiento menos experimentales y menos científicas. En los experimentos mentales, asimismo, debe haber variables independientemente aislables para que seamos capaces de determinar las correlaciones entre estas variables y el resultado del experimento (véase Irvine, 1991: 152). Esta condición es de especial importancia cuando se intenta evaluar si esta clase de experimentos puede prestar apoyo y en qué grado a hipótesis científicas. Estas condiciones resultan satisfechas si interpretamos a los experimentos mentales como modelos científicos.

La serie de evidencias experimentales que Galileo aduce en *El diálogo sobre los dos máximos sistemas del mundo* como evidencia para sostener que la Tierra se mueve, son ejemplos claros del uso apologético de los experimentos mentales. Salviati propone imaginar una serie de situaciones que podrían darse en la cabina cerrada de un barco. Propone primero observar cómo se comportan una serie de objetos mientras el barco está en reposo: una pecera con un pez que se desplaza libremente, algunos animales voladores como mariposas o moscas, un sistema compuesto por una botella que deja caer gotas de agua sobre un plato, una pelota que es lanzada hacia adelante. Luego propone observar nuevamente como se mueven los mismos objetos cuando el barco se desplaza a una velocidad arbitraria siempre y cuando el movimiento sea uniforme y no fluctuante. En este caso señala Salviati que:

No descubrirás el menor cambio en todos los efectos nombrados, ni podrás saber por ninguno de ellos si el barco se estaba moviendo o parado. Al saltar, pasarás por el suelo los mismos espacios que antes, ni harás saltos más grandes hacia la popa que hacia la proa aunque el barco se mueva con bastante rapidez, a pesar de que durante el tiempo que estés en el aire el piso debajo de ti irá en dirección opuesta a tu salto. Al arrojar algo a su compañero, no necesitará más fuerza para llevárselo, ya sea que esté en la dirección de la proa o de la popa, con usted situado enfrente. Las gotas caerán como antes en la embarcación que se encuentra

debajo sin caer hacia la popa, aunque mientras las gotas están en el aire, la nave recorre muchos tramos. Los peces en el agua nadarán hacia el frente de su cuenco sin más esfuerzo que hacia atrás, e irán con la misma facilidad al cebo colocado en cualquier lugar alrededor de los bordes del cuenco. Finalmente las mariposas y moscas continuarán su vuelo indiferente hacia todos lados, ni sucederá nunca que se concentren hacia la popa, como cansados de seguir el rumbo del barco, del que habrán estado separados durante mucho tiempo manteniéndose en el aire. Y si se haces humo quemando un poco de incienso, se verá subir en forma de nube, quedando quieto y no se mueve más hacia un lado que hacia el otro. (Galileo, [1632] 1967: 187)

La explicación de por qué el comportamiento de los objetos es idéntico en ambas situaciones es que “el movimiento de barco es común a todas las cosas contenidas en él y al aire también” (Galileo, [1632] 1967: 187). El experimento se emplea para proporcionar apoyo a la hipótesis de acuerdo con la cual un marco de referencia en reposo es indistinguible de uno en movimiento rectilíneo uniforme. Aunque las experiencias propuestas por Galileo son realizables, el valor apologético del experimento mental reside justamente en que se sirve de experiencias comunes que, recuperadas como insumo para la construcción del caso, proporcionan evidencia a favor de su hipótesis. El poder persuasivo del experimento descansa sobre la obviedad de las experiencias a las que apela. Así lo manifiesta Sagredo:

Aunque no se me ocurrió poner estas observaciones a prueba cuando estaba viajando, estoy seguro de que se llevarían a cabo de la manera que usted describe. En confirmación de esto, recuerdo haberme encontrado a menudo en mi camarote preguntándome si el barco se estaba moviendo o parado; y a veces, por capricho, he supuesto que iba en una dirección cuando su movimiento era el contrario. (Galileo, [1632] 1967: 188)

Otras de las características que acercan experimentos reales y mentales, desde la perspectiva de Irvine, son útiles para argumentar que los experimentos mentales pueden cumplir funciones apologéticas. La reproductibilidad de los experimentos en general depende de que el procedimiento de intervención en la naturaleza, realizado bajo circunstancias controladas, sea repetible bajo condiciones similares. La descripción del diseño experimental de un experimento

mental exitoso, esto es, capaz de ampliar nuestro conocimiento sobre el mundo natural, también debería habilitar para Irvine esta clase de repetición. Finalmente, el resultado del experimento mental debe tener repercusiones en la teoría de fondo original. El razonamiento concerniente a un estado particular de cosas o evento debe proporcionar evidencia a favor o en contra de alguna conclusión sobre el mundo (Irvine, 1991: 159). Aunque este elemento solo resulta evaluable desde una perspectiva temporal distante de la formulación del experimento mental, algunos experimentos mentales han tenido esta característica. Uno de los casos que por las particularidades de su narrativa y su importancia histórica satisface las condiciones que desde una perspectiva empirista debe cumplir un experimento capaz de aumentar nuestro conocimiento de la naturaleza, es el experimento pergeñado por Newton para motivar la explicación del movimiento orbital.

En *El sistema del mundo*, un opúsculo suprimido en la edición de la *Principia* de 1686 pero impreso un año después de la muerte de Newton (Hall, 1992: 345), se propone un experimento mental para mostrar que el movimiento de los planetas y los satélites es comparable al de los proyectiles terrestres. El experimento describe el lanzamiento de proyectiles desde una montaña muy alta. Supone, además, que el efecto de la resistencia del aire es despreciable. Si las balas de cañón pudieran ser lanzadas con una velocidad inicial suficiente, no terminarían su trayectoria al caer sobre la superficie de la Tierra sino que girarían continuamente a su alrededor (Fig. 7.8):

Que los planetas son retenidos en ciertas órbitas por medio de fuerzas centrípetas, podemos entenderlo fácilmente, si consideramos los movimientos de los proyectiles; porque una piedra proyectada es forzada por la presión de su propio peso fuera del camino rectilíneo, que la proyección debería haber seguido, y describe una línea curva en el aire; y por ese camino tortuoso es finalmente derribado a tierra; y cuanto mayor es la velocidad con la que se proyecta, más lejos va antes de caer a la tierra. Por lo tanto, podemos suponer que si la velocidad aumenta tanto que describiría un arco de 1, 2, 5, 10, 100,

1000 millas antes de llegar a la tierra, hasta que por fin, superando los límites de la tierra, debería pasar bastante sin tocarlo.

Sea AFB la superficie de la tierra, C su centro, VD , VE , VF , las líneas curvas que describiría un cuerpo si se proyectara en dirección horizontal desde la cima de una montaña alta sucesivamente con más y más velocidad y, como los movimientos celestes apenas son retardados por la escasa o nula resistencia de los espacios en los que se realizan, para mantener la paridad de casos, supongamos que no hay aire en la tierra, o al menos que está dotado de poco o ningún poder de resistencia; y por la misma razón que el cuerpo proyectado con menor velocidad describe el arco menor VD , y con mayor velocidad el arco mayor VE y, aumentando la velocidad, se aleja cada vez más hacia F y G , si la velocidad es aún mayor, llegaría por fin bastante más allá de la circunferencia de la tierra y volvería a la montaña desde la que se proyectaba.

Y dado que las áreas que por este movimiento describe por un radio dibujado en el centro de la tierra son (por la Prop. 1, Libro 1, *Princip. Math.*) proporcionales a los tiempos en que se describen, su velocidad, cuando regresa a la montaña, no será menor de lo que era al principio; y, manteniendo la misma velocidad, describirá la misma curva una y otra vez, por la misma ley. ([1686] 1934: 512-513)

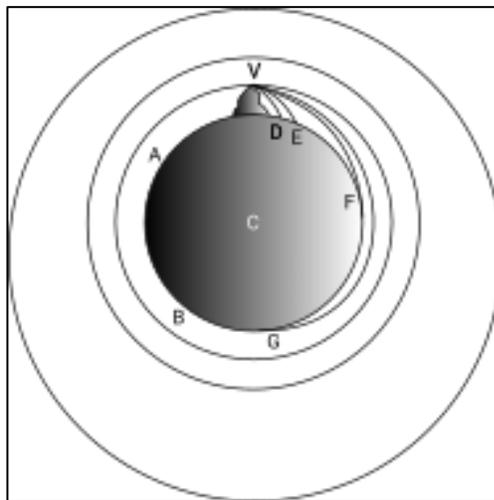


Figura 7.8

Simplificando la formulación de Newton, el experimento muestra que en ausencia de gravedad y de resistencia del aire, una bala de cañón se mueve en la línea en que fue disparada pero cuando la gravedad de la tierra actúa sobre ésta, sigue un camino que depende de la altitud y de la velocidad inicial con que fue

disparada. Si la velocidad es muy baja, simplemente caerá de nuevo a la tierra. Cuando la velocidad inicial alcanza un valor crítico, regresará a la montaña a través de una órbita elíptica.

Hooke (1674) había ofrecido los fundamentos conceptuales y algunas pruebas mecánicas que mostraban que los movimientos celestes y terrestres podían explicarse por las mismas fuerzas, mostrando en el laboratorio que una fuerza central producía un movimiento circular. Un cuerpo permanece en movimiento circular tratando de caer, debido a la acción atractiva de la Tierra hacia el centro, lo que impide su movimiento rectilíneo –esto es que salga disparado por la tangente- pero a su vez, y a causa de su inercia, el cuerpo no “cae” hacia el centro.

Newton presenta un modelo que integra varios aportes teóricos que explican un fenómeno (el movimiento orbital de los satélites). Por un lado, recupera los aportes de Galileo acerca de la inercia y el movimiento acelerado. En el escenario descrito los satélites son proyectiles con movimiento rápido. Por otro lado, el experimento de Newton extrae consecuencias de esta hipótesis de Hooke. Podría decirse que en el experimento mental Newton imagina las condiciones para poner en órbita un satélite artificial de la Tierra, vinculando el efecto de la velocidad tangencial y la gravedad. Finalmente, el modelo integró la primera y la segunda ley de Kepler con la acción de la fuerza centrípeta de la gravedad. La fuerza de atracción varía según la distancia, por lo que el satélite describe un círculo perfecto solo si la velocidad tangencial y la resultante fuerza centrífuga contrarrestan la atracción de la fuerza central. En la órbita elíptica la rapidez varía. Cuando el proyectil alcanza una velocidad crítica, se aleja del centro de gravedad perdiendo rapidez. Al caer de regreso a la Tierra gana rapidez (ya que se mueve en la misma dirección que la fuerza atractiva) hasta que vuelve a su trayectoria original.

7.5 Conclusiones

La clasificación de los experimentos mentales basada en la noción de *uso*, en contraposición a las clasificaciones basadas en la noción de función o en tipos de resultados, permite dar sentido a su importancia histórica y a los episodios en los cuales un mismo experimento mental fue empleado para defender hipótesis rivales. Esa interpretación apoya la idea de que los experimentos mentales gozan de cierta independencia teórica e ilustra de qué manera su función está definida por su objetivo cognitivo en un marco contextual. Asimismo, hace posible argüir que los experimentos mentales desempeñaron funciones en el contexto de justificación de las teorías. Los experimentos mentales destructivos son ejemplos canónicos de este tipo de usos. Los usos confirmatorios de los experimentos mentales son más controversiales. La perspectiva aquí defendida permite sostener que, en contra de la interpretación proporcionada por la mayoría de los filósofos de la experimentación, los experimentos mentales pueden dar apoyo a hipótesis en los casos en los que estos son capaces de proporcionar modelos fenoménicos apoyados en interpretaciones teóricas. Otros usos de los experimentos mentales menos controversiales son los restringidos al contexto de descubrimiento de las teorías. El uso heurístico de los experimentos mentales corresponde al objetivo de delinear o formular un problema teórico. El uso exploratorio de los experimentos mentales consiste fundamentalmente en ofrecer explicaciones potenciales ideando escenarios que, si fueran reales, darían lugar al fenómeno que constituye el *explanandum* en cuestión.

CAPÍTULO 8

Los experimentos mentales en la formulación y las soluciones de la paradoja de Olbers

*“Though my soul may set in darkness, it will rise in perfect light;
I have loved the stars too fondly to be fearful of the night.”*

(Williams, 1868: 69)

8.1 Introducción

Los filósofos de la ciencia han examinado con bastante detalle, y de manera reiterada, un conjunto de casos históricos de experimentos mentales, muchos de los cuales hemos mencionado en capítulos anteriores. Los casos suelen repetirse en la bibliografía sobre el tema: el experimento de la caída libre de Galileo, el del balde rotatorio de Newton, el del demonio de Maxwell, los de Einstein relativos a la luz y la gravitación (como el del ascensor acelerado), el del gato de Schrödinger y el de la doble rendija con electrones, entre algunos otros más. Todos forman parte de la historia de la física y algunos, como los de Galileo, el de Newton y el de la doble rendija, son realizables (y han sido realizados) como experimentos reales. Además, todos pertenecen a dominios de la física, como la mecánica (newtoniana, relativista o cuántica), donde la experimentación real no solo es posible, sino que ha prevalecido sobre la experimentación mental. Haciendo una generalización rápida, podría decirse que, aunque no hay duda de que los experimentos mentales desempeñaron un papel heurístico y refutatorio en la historia de la mecánica, las teorías físicas en este dominio han sido confirmadas casi exclusivamente mediante la evidencia proporcionada por los experimentos reales.

Hay otros dominios de las ciencias físicas, como la astrofísica y la cosmología, donde las posibilidades de experimentación real están severamente limitadas, no solo por razones técnicas, sino incluso por principio. En esas ciencias, los experimentos mentales, frecuentemente implementados mediante simulaciones computacionales, desempeñan un papel particularmente relevante en la justificación epistémica de modelos y teorías. El estudio de caso que se desarrolla en este capítulo se centra en la llamada “Paradoja de Olbers”. La paradoja se plantea a partir de la sencilla observación de la oscuridad del cielo nocturno. Este hecho, a pesar de su simplicidad, es difícilmente explicable en el contexto de las cosmologías newtoniana y relativista. Una respuesta a la pregunta por la oscuridad de la noche es que el Sol está brillando del otro lado de la Tierra y la luz estelar es más débil que la luz solar. Pero el enigma permanece si se supone que el universo contiene un número infinito de estrellas distribuidas uniformemente. En un modelo cosmológico de estas características el cielo nocturno debería ser completamente brillante. Incluso si se considera una estructura astronómica más compleja de estrellas agrupadas en galaxias o estructuras mayores, el resultado es un cielo completamente cubierto de estrellas y con ello, completamente brillante. La hipótesis de un universo en expansión e incluso la consideración de una geometría no euclidiana no resuelven este misterio.

La paradoja de Olbers y (muchas de) sus soluciones fueron planteados como experimentos mentales. La selección de un caso de estudio procedente de la cosmología tiene el propósito de poner a prueba algunas de las tesis elaboradas sobre el funcionamiento y alcance de los experimentos mentales. Para ello se describe el desarrollo histórico de la paradoja y se presentan algunos de los tratamientos de la misma. Se examina el experimento mental propuesto por Harrison (1987 y 2000), ya que este exhibe algunas de las características que en este trabajo se han señalado como típicas de esta clase de experimentos y presenta claras similitudes con las simulaciones computacionales.

8.2 El surgimiento de la paradoja

Una paradoja se produce cuando partir de ciertos supuestos razonables, y de una inferencia aparentemente válida se derivan consecuencias inconsistentes, o bien inesperadas o que contradicen el sentido común. Una paradoja física es un razonamiento con premisas que se aceptan como verdaderas (ya sea porque provienen de la observación o de teorías bien confirmadas) y que a través de una inferencia correcta conduce a una conclusión contradictoria. Hay básicamente tres maneras de abordar a las paradojas en física, las tres consisten en estrategias para eliminar o disolver la contradicción. Si se considera que las premisas que constituyen una paradoja son individualmente plausibles pero mutuamente inconsistentes, entonces, es preciso identificar la premisa falsa y abandonarla. Si no es posible abandonar ninguna premisa, entonces, o bien debe haber un error en la inferencia que necesita ser explicitado o bien la contradicción entre las premisas es solo aparente y como tal puede ser desestimada. Diversas perspectivas teóricas han clasificado a las paradojas en física empleando criterios similares (véanse Sorensen, 2003; Rescher, 2001; y Cuicic, 2009). La gravedad del elemento a corregir se ha vinculado a diferentes transformaciones que deben ser operadas en las teorías físicas: una interpretación errónea de los datos puede conducir a la formulación de una premisa errónea, un “hueco” en la teoría puede requerir de una explicitación ulterior para evitar la contradicción o para mostrar que la misma es solo aparente. Las modificaciones necesarias para evitar la paradoja pueden implicar cambios teóricos conservadores o revolucionarios.

Las paradojas cosmológicas tienen particularidades que han capturado la atención de filósofos y científicos. Dado que la experimentación es sumamente difícil o imposible en este ámbito y que la información empírica es escasa o difícil de integrar a la escala de los sistemas bajo escrutinio, las paradojas cosmológicas plantean enigmas cuya resolución demanda un alto grado de especulación y

sutileza. Estos enigmas en ocasiones comprometen principios teóricos fundamentales.

La paradoja de la noche oscura, junto a la paradoja del colapso gravitacional del universo, fueron enigmas acuciantes para el sistema newtoniano. La cosmología newtoniana asume que las estrellas se encuentran distribuidas por todo el universo a diferentes distancias. Newton había sostenido en sus primeros años una cosmología semejante a la estoica, donde un universo finito se encontraba rodeado por un espacio vacío que se extendía infinitamente. En ese universo, si se supone que la materia esparcida por el espacio se atrae mutuamente, el colapso gravitatorio sería inevitable. No obstante, dado que no se observa movimiento en las estrellas fijas, es razonable suponer que el colapso no se está produciendo.

La paradoja del colapso gravitacional fue planteada por el reverendo Richard Bentley en 1692 y motivó un cambio en algunos de los supuestos de Newton, quien a partir del reconocimiento de este conflicto, conjeturó un modelo cosmológico con infinitas estrellas distribuidas de manera más o menos uniforme en un espacio infinito. No obstante, este cambio en los supuestos no soluciona el problema del colapso gravitacional. En un escenario como este, la fuerza gravitatoria debería atraer las grandes masas unas hacia otras. Este proceso podría haber dado lugar a la formación de estrellas o conglomerados de estrellas. Pero, dado que la fuerza de gravedad es siempre atractiva y proporcional a la masa de los cuerpos, las estrellas y conglomerados de estrellas deberían colapsar formando una única masa esférica.

La solución que Bentley propuso para un modelo en el que el universo es infinito, es que todas las estrellas se encuentran en reposo porque cada una está rodeada por infinitas estrellas en todas direcciones, de modo que las fuerzas atractivas de todas ellas se cancelan mutuamente, de manera tan precisa como

para que la fuerza neta sobre cada estrella sea exactamente igual a cero. Newton ilustró la improbabilidad de un escenario como este con un experimento mental:

[...] suponer que todas las partículas en un espacio infinito estuviesen tan exactamente equilibradas entre sí como para permanecer inmóviles en perfecto equilibrio. [...] es tan difícil como hacer, no que una aguja, sino que un número infinito de ellas (tantas como partículas hay en un espacio infinito) permaneciesen perfectamente en equilibrio sobre sus puntas.... (Newton, January 17, 1693; en Janiak, 2014: 124).

El equilibrio del universo de Newton necesariamente debería romperse por la acción de la fuerza atractiva de la gravedad. Se ha probado que el equilibrio gravitatorio imaginado por Bentley es inestable y que el colapso gravitatorio también se produciría en un universo infinito (véase Guth, 1997: 295-297). Por lo que, la paradoja del colapso gravitacional resulta irresoluble en el marco de la cosmología newtoniana.

La paradoja de la oscuridad de la noche tampoco es explicable dentro del marco de la física newtoniana. Su punto de partida es la simple observación de que el cielo nocturno es oscuro. Este hecho resulta, sin embargo, difícilmente explicable si se supone que el universo es espacialmente infinito y está poblado por un número infinito de estrellas distribuidas de manera más o menos uniforme. La primera formulación de la paradoja de la noche oscura le fue sugerida a Newton por un aficionado a la astronomía, en base a la observación de la Vía Láctea:

Este problema que parece no haber sido advertido por Newton hasta muy tarde, fue señalado por el joven médico, y aficionado a la astronomía, William Stukeley (1687-1765) en una conversación informal con Newton. Stukeley conjeturó que la estructura tridimensional de la Vía Láctea era semejante a la de Saturno, es decir, un anillo aplanado que rodeaba un conglomerado esférico de estrellas. Cuando Newton afirmó que prefería un universo infinito donde las estrellas estuvieran simétricamente distribuidas, Stukeley replicó que en ese universo “el hemisferio entero del cielo habría tenido la apariencia de esa penumbra luminosa de la Vía Láctea” (citado por Hoskin 2003, trad. esp., p. 132). Stukeley ofreció, así, una versión de la paradoja de la noche oscura: si hay infinitas estrellas que están uniformemente distribuidas por todo el espacio ¿por qué observamos una Vía

Láctea brillante sobre un fondo oscuro? En el universo newtoniano, todo el cielo debía brillar como la Vía Láctea. Newton, que nunca resolvió el problema, al parecer jamás imaginó que la existencia de la Vía Láctea proporcionaba una evidencia que refutaba su modelo cosmológico de un universo simétrico, o, por lo menos, que constituía una seria anomalía. (Cassini, 2020: 5)

Se pueden advertir ciertas semejanzas entre la paradoja de la oscuridad de la noche y el colapso gravitatorio del universo. Ante todo, tanto la fuerza de gravedad como la luz tienen un alcance infinito, es decir, pueden propagarse a distancias arbitrariamente grandes. Además, tanto la intensidad de la fuerza gravitatoria como la intensidad de la luz disminuyen con el cuadrado de la distancia. Ambas son proporcionales a la cantidad de masa y a la cantidad de fuentes luminosas, respectivamente. La existencia de un número infinito de estrellas implica tanto la existencia de una cantidad infinita de masa como de fuentes luminosas en el universo. De esta manera, la disminución de la fuerza gravitatoria y de la intensidad de la luz de las estrellas con el cuadrado de la distancia está compensada por el aumento del número de estrellas, que a una distancia dada también es proporcional al cuadrado de la distancia (podemos pensarlas como si estuvieran distribuidas en capas esféricas). En consecuencia, si la densidad de la distribución de las estrellas se mantiene uniforme en todo el espacio, la intensidad de la fuerza gravitatoria se duplica al duplicarse la distancia a un punto dado y lo mismo ocurre con la intensidad de la luz emitida por las estrellas.

A diferencia del enigma gravitacional, para el que el modelo newtoniano puede ofrecer una solución (aunque improbable), en la que iguales fuerzas que atraen en todas las direcciones se cancelan, el enigma de la luz estelar no puede ser resuelto con una estrategia similar. Enormes cantidades de luz que proceden de diferentes direcciones no se cancelan mutuamente sino que se refuerzan dando como resultado un cielo completamente brillante, hecho que entra en conflicto evidente con la observación.

La oscuridad del cielo nocturno fue considerada como hecho cosmológicamente relevante en varias ocasiones previas a la formulación y discusión del modelo newtoniano. Kepler (1610), por ejemplo, escribe a Galileo que la observación de miles de estrellas, apiñadas y formando cúmulos, fortalece el argumento contra la infinitud del universo. Si las estrellas, en número infinito, se extendieran en un espacio infinito, la bóveda celeste debería brillar tanto como la superficie del Sol. Los espacios oscuros que pueden observarse entre las estrellas son, para Kepler, el límite cósmico de un universo finito. Digges (1576) por su parte, consideró que los espacios oscuros entre las estrellas estaban causados por la distancia que separaba a un observador situado en la Tierra de ellas. Digges podría señalarse como el primero en notar que la oscuridad de la noche requería de una explicación para quienes sostenían la infinitud del universo.

Aunque el enigma fue retomado y varios intentos de solución fueron propuestos por astrónomos reconocidos, entre ellos por Halley, la paradoja de la noche oscura es conocida como la paradoja de Olbers. El médico y astrónomo alemán publicó en 1823 un artículo denominado “Sobre la transparencia del espacio” donde recuperó el enigma y la fallida solución propuesta por Halley en 1721. Aunque su solución es muy similar a la que mucho antes había sido propuesta por Chéseaux en 1744, no hay referencia a esta formulación en su trabajo y para la posteridad quedó signada por el nombre del astrónomo alemán.⁵⁰

El planteo general de la paradoja propuesto por Harrison supone un observador en el centro de un bosque densamente poblado de árboles (Harrison, 2000: 4-5). Hacia cualquier dirección que el observador dirija su mirada interceptará el tronco de un árbol y a la distancia, un trasfondo de troncos de árboles. Análogamente, si se supone que el universo es infinito (esta hipótesis se funda en la implausibilidad de encontrar un límite que cerque abruptamente los confines del universo) y al mismo tiempo se asume que las estrellas se distribuyen

⁵⁰ La historia de la paradoja de Olbers se trata con todo detalle en el libro de Harrison (1987). Aquí solo hemos señalado algunos aspectos esenciales para introducir el tema.

de manera más o menos uniforme en este espacio infinito, un observador situado en cualquier punto de este espacio debería interceptar una estrella en la línea de la visión. Dado que las estrellas no son puntos geométricos sino cuerpos luminosos de tamaño finito pero de número infinito, distribuidas en el espacio, el observador encontraría en este escenario una estrella en cualquier línea recta en la que dirija su visión, no importa cómo estén distribuidas o qué tan separadas se encuentren unas de otras. Esto ocurriría incluso si se agruparan en galaxias o si se distribuyeran de manera que se solaparan unas a otras formando un fondo continuo. Si no es posible observar en alguna dirección definida de este “bosque de estrellas” sin encontrarse con una estrella, entonces, no es posible encontrar huecos entre ellas. Si este supuesto es válido para nuestro universo, el cielo estaría cubierto de estrellas tan brillantes como el Sol y con ello debería ser completamente brillante todo el tiempo (de hecho 180 000 veces más brillante que el disco del Sol). Pero el cielo nocturno es oscuro, por lo que se plantea una paradoja para el modelo cosmológico supuesto.

Adviértase que para que se produzca la paradoja de Olbers no es necesario que exista un número infinito de estrellas. La paradoja se produciría igualmente en un universo finito donde las estrellas (necesariamente finitas en número) estuvieran distribuidas de manera muy densa, es decir, a una distancia suficientemente corta unas de otras. Análogamente, un observador situado en medio de un bosque no puede ver el exterior si los árboles que lo rodean están agrupados de manera muy densa. No obstante, la paradoja podría resolverse en un universo finito suponiendo que no hay un número suficiente de estrellas como para interceptar toda línea de visión, o bien que su distribución en el espacio no es lo suficientemente densa porque están separadas por grandes distancias. La analogía con el bosque se aplica aquí nuevamente. En cambio, en un universo que contuviera un número infinito de estrellas, la paradoja debería producirse cualquiera fuera la densidad de la distribución de las estrellas, es decir, cualquiera fuera la distancia que las separara, siempre que tal distancia sea finita. No sería

posible resolver la paradoja suponiendo que la distancia entre las estrellas fuera infinita, porque, en ese caso, admitiendo que la luz se propaga con una velocidad finita, en el cielo nocturno no podría observarse ninguna estrella, lo cual resulta obviamente incompatible con la experiencia.

Las soluciones a esta paradoja pueden ser agrupadas según Harrison (1987: 8-10; 2000: 493) en dos grupos. Las del primer tipo son las llamadas *interpretaciones del cielo cubierto*. Según estas interpretaciones los espacios interestelares aparentemente vacíos, contrariamente a lo que se observa, están ocupados por estrellas que, por alguna causa desconocida, se hallan ocultas a nuestra visión. Este tipo de soluciones intenta dar respuesta a la paradoja explicando por qué la luz de dichas estrellas no llega a nuestro punto de vista terrestre apelando, por ejemplo, a la hipótesis de que la luz es bloqueada por alguna clase de materia oscura. El segundo tipo de soluciones son las llamadas *interpretaciones del cielo no cubierto*. Según estas, los espacios interestelares permanecen oscuros porque no contienen estrellas. Las soluciones de este tipo procuran dar respuesta a la paradoja explicando el enigma de las estrellas faltantes. Harrison ha realizado una crítica detallada de cada una de estas supuestas soluciones y ha argumentado que todas ellas fallan en algún aspecto y no consiguen, por tanto, evitar la paradoja. En términos generales, el problema, según Harrison, no es el de la presencia de luz visible, sino el de la acumulación de radiación térmica emitida por las estrellas. Cualquier tipo de materia que bloqueara la luz de las estrellas debería calentarse con el tiempo y, en algún momento, comenzaría a emitir tanta radiación como absorbe. En consecuencia, emitiría toda la radiación de las estrellas que oculta. El problema podría resolverse suponiendo que el universo es extremadamente joven, tanto como para que el calentamiento de la materia oscura resulte despreciable hasta el momento. Pero esa hipótesis tampoco resulta aceptable en la cosmología actual.

A las consideraciones de Harrison puede agregarse que los modelos cosmológicos que suponen un universo finito no pueden dar una solución correcta

a la paradoja. Una solución posible dentro de estas suposiciones es restablecer el límite cósmico que habían sostenido los cosmólogos antiguos. En este contexto se supone el universo no posee estrellas suficientes para cubrir el cielo y que los espacios oscuros entre ellas son el límite del universo. Pero la suposición de un universo espacialmente limitado que termina abruptamente en un límite cósmico es inaceptable para la cosmología contemporánea. Otras formas de limitar el universo, como la suposición estoica de un universo isla que flota en el espacio vacío finalmente fueron abandonadas a comienzos del siglo XX, cuando, hacia 1930, se admitió la existencia de múltiples galaxias.

Una alternativa a esta interpretación es considerar el universo como ilimitado pero de tamaño finito. Un universo finito de forma esférica podría dar sentido a la oscuridad de la noche. Sin embargo, no es razonable suponer que el cielo nocturno sería oscuro si el límite hasta el cual se extiende la visión de un observador es mayor que la distancia entre las antípodas de dicho universo. En la superficie del globo ilimitada, i.e. de área finita, los rayos de luz pueden viajar en cualquier dirección y eventualmente volver al punto inicial. En este escenario, el observador ve un trasfondo continuo de estrellas formado por las circunnavegaciones repetidas de rayos de luz estelar. Un universo relativista, deducido de hipótesis simples de la relatividad de Einstein permite imaginar universos con curvatura positiva y geometría esférica, representable por una esfera de tres dimensiones. En este tipo de universo un rayo de luz que viaje en cualquier dirección puede dar toda la vuelta a la esfera si se espera el tiempo suficiente. Por lo que un universo finito pero ilimitado tampoco soluciona la paradoja, a menos que se acepte que el universo es tan joven y tan grande como para que la luz no haya tenido tiempo todavía para recorrerlo en su totalidad.

La única solución cuantitativamente correcta a la paradoja relaciona el límite del universo visible con la radiación que se estima que emite la superficie del Sol. De acuerdo con esta interpretación, el universo visible está formado por un número finito de estrellas cuya luz ha tenido tiempo de alcanzar al observador.

Las estrellas más allá de este horizonte no pueden ser vistas. La radiación del universo visible es mucho menor que la necesaria para crear un cielo completamente brillante. De acuerdo con Harrison, la solución general a la paradoja de Olbers reconoce que la luz se propaga con velocidad finita y que el universo tiene una edad finita o que, al menos, las estrellas tienen una vida luminosa finita. El cielo nocturno es oscuro porque la separación promedio de las estrellas es tal que el límite de fondo hasta el que podemos ver es mayor que la edad del universo o la vida de las estrellas luminosas (Harrison, 2000: 511).

El propio Harrison ha elaborado un experimento mental en respuesta a la paradoja. Su solución muestra que no hay suficiente energía en el universo para crear un cielo completamente brillante. Una manera sencilla de formular la propuesta de Harrison es empleando la equivalencia entre masa y energía. Si toda la materia del universo, en sus diversas formas, fuese aniquilada y convertida en radiación térmica, la cantidad promedio de materia en el universo sería equivalente a un átomo de hidrógeno por metro cúbico. La cantidad de energía que contiene toda la materia del universo convertida en radiación térmica está muy lejos de la necesaria para crear un cielo completamente brillante (unos 20 grados Kelvin, mientras que la superficie del Sol está a 6000 grados Kelvin). Independientemente de si el universo se expande, se contrae o permanece estático, un cielo completamente brillante requiere 10 mil millones de veces más energía que la que obtendríamos de convertir toda la materia en radiación. En la sección 6 de este capítulo analizaremos con detalle el experimento mental propuesto por Harrison.

8.3 La formulación inicial y las primeras soluciones a la paradoja

Algunas soluciones a la paradoja interpretaron la oscuridad de la noche como un fenómeno que, en un universo poblado por un número infinito de estrellas, podía explicarse debido a la absorción de la luz por materia interestelar, o debido a la enorme distancia de las estrellas lejanas, que hace que no sean visibles, o bien debido a la existencia de un límite de fondo, detrás del cuál no era posible percibir más estrellas. Este tipo de soluciones son todas interpretaciones del cielo cubierto, según la clasificación de Harrison (1987: 8).

Entre las interpretaciones del cielo cubierto se encuentran las de Halley, Chéseaux y Olbers. Los tres ofrecieron planteos similares de la paradoja y soluciones que intentan dar una explicación de la luz estelar faltante. La formulación de la paradoja depende de la construcción de modelos cosmológicos que suponen un observador situado en su centro y una distribución homogénea de las estrellas a su alrededor. En todos los casos, el universo se considera estático. Las soluciones propuestas establecen que la oscuridad del cielo nocturno se debe a que existe algún tipo de obstáculo que intercepta la luz en su recorrido hasta el observador (Chéseaux y Olbers) o a que la luz de las estrellas distantes es demasiado débil para alcanzarlo. Si bien estas propuestas fallan por razones termodinámicas, dado que la acumulación de radiación térmica emitida por las estrellas lejanas terminaría por calentar cualquier forma de materia conocida (materia bariónica, diferente de la materia oscura), interesa examinar de qué manera el uso ilustrativo de estos experimentos mentales permitió mostrar una contradicción entre una observación simple y los principios teóricos de la cosmología newtoniana. Las formulaciones iniciales de la paradoja dependen de una serie de supuestos acerca de la naturaleza del universo (como su carácter homogéneo y estático) y de la velocidad y la trayectoria de la luz. En cada caso la ilustración de la contradicción se basa en la elaboración de un modelo de

distribución de las estrellas que involucra una representación visual cuyo propósito es describir el fenómeno y apoyar la plausibilidad de una solución al enigma.

El modelo cosmológico Newtoniano proporciona una base para calcular, empleando la geometría euclídeana, la cantidad de luz estelar que un observador puede recibir. Dentro de este modelo, las estrellas cercanas emiten una gran cantidad de luz, pero son pocas. Las estrellas distantes, en cambio, son numerosas pero cada una solo contribuye con una muy pequeña cantidad del total de luz recibida por el observador. El primero en ofrecer una solución basada en este tipo de cálculo fue Halley en un breve artículo de 1721 titulado "Of the Infinity of the Sphere of Fixed Stars".⁵¹ De acuerdo con esta interpretación, un número infinito de estrellas en un universo ilimitado debería cubrir todo el cielo. ¿Por qué entonces el cielo nocturno es oscuro? Halley supuso que la luminosidad aparente de una estrella es inversamente proporcional al cuadrado de su distancia. En este modelo, hay infinitas estrellas, pero las más lejanas son cada vez menos luminosas. La solución a la paradoja depende del efecto geométrico producido por la distribución de las estrellas en un universo infinito.

Para exhibir su solución, Halley propone un experimento mental. El mismo consiste en imaginar un observador en el centro de una serie de esferas concéntricas cuyo radio se incrementa progresivamente respecto del observador. Cada capa tiene el mismo espesor, es decir, la distancia entre la superficie de una esfera y la de la siguiente es igual. A medida que se incrementa el radio de la esfera, se incrementa, a razón del cuadrado del radio de cada esfera, el número de estrellas (el número de estrellas en cada esfera es proporcional a su superficie, esto es al duplicarse el tamaño de la esfera, el número de estrellas que cada esfera contiene se cuadruplica). Pero la cantidad de luz (la intensidad) que el observador situado en el centro recibe desde cada estrella, decrece a razón del cuadrado del

⁵¹ Reimpreso en Harrison (1987: 218-219).

radio de cada capa. De acuerdo con el experimento, todas las capas contribuyen con la misma cantidad de luz y la cantidad total de luz que un observador en el centro de este escenario recibiría es igual a la cantidad de luz de cada capa multiplicada por el número de capas. Halley dedujo, incorrectamente, que la intensidad de la luz de las estrellas más remotas decrece en una proporción mayor que el inverso del cuadrado de la distancia y que este hecho podría explicar la oscuridad de la noche (Fig. 8.1).

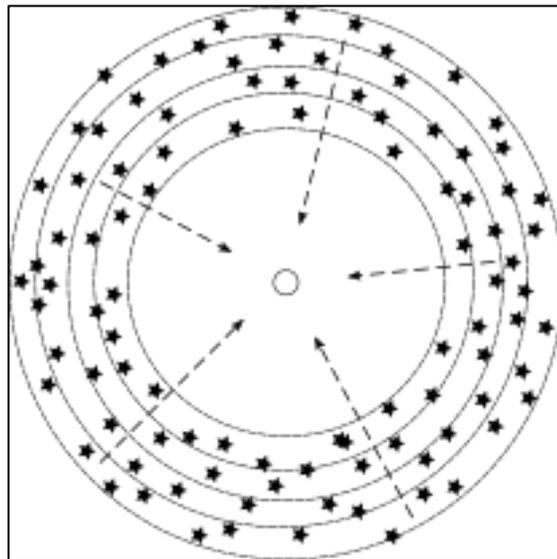


Figura 8.1

Halley intentó aclarar la confusión del argumento en un segundo artículo, publicado el mismo año 1721, titulado “On the Number, Order, and Light of the Fixed Stars”,⁵² en el que agregó que la oscuridad es producida por la debilidad de la luz de las estrellas más remotas, que no puede ser percibida. Esta segunda solución asume correctamente que los rayos de una estrella muy remota no pueden ser registrados a simple vista, pero falla en cuanto no considera que la combinación de los rayos de numerosas estrellas no sería igualmente débil. Si bien el disco de cualquier estrella a una distancia suficientemente grande es tan

⁵² Reimpreso en Harrison (1987: 219-220).

pequeño que nos sería imposible detectar su luz, un número virtualmente incalculable de estrellas acumularía luz suficiente como para volver al cielo completamente brillante. Un universo que se extiende ilimitadamente, contiene un número infinito de capas y, por lo tanto, el observador situado en el centro (que puede ser cualquier punto del espacio) recibe una cantidad infinita de luz, lo que haría al cielo infinitamente brillante en todos sus puntos.

Puede advertirse que en la descripción del escenario del experimento mental de Halley hay algunos supuestos erróneos. Sabemos que las estrellas se agrupan en galaxias por lo que si se supone una distribución uniforme de las mismas, el modelo de distribución de las estrellas propuesto implica que el observador debería recibir una cantidad infinita de luz, procedente de las regiones más remotas del universo. No obstante, es razonable suponer que las estrellas más cercanas interceptan parte de la luz procedente de las estrellas más lejanas, o en otras palabras, las estrellas que se encuentran en las esferas más próximas ocultan de nuestra vista a las estrellas más lejanas. En este escenario se alcanzaría un límite de fondo en el cual el cielo estaría completamente cubierto de estrellas y muy poca luz de las estrellas lejanas llegaría al centro.

Loys de Chéseaux advirtió el error en el razonamiento de Halley. Si suponemos que el universo se extiende infinitamente, debemos concluir que una infinita cantidad de luz se acumula proveniente de todas las capas. No obstante, cuando el cielo está completamente cubierto, las estrellas que componen las esferas forman un fondo continuo detrás del cual no podemos recibir la luz de más estrellas. Una vez que el cielo está completamente cubierto de estrellas, este fondo continuo bloquea la luz de las estrellas que se encuentran más lejos. La hipótesis de Chéseaux fue que incluso en un universo infinito, que contiene un número infinito de estrellas sólo podemos ver un número finito de ellas. En su libro

Traité de la comète, de 1744, Chéseaux proporciona el primer tratamiento cuantitativo de la paradoja.⁵³

Halley había señalado que aunque el cielo está completamente cubierto de estrellas, la mayoría no son visibles debido a que su luz es demasiado débil para alcanzar al observador. Chéseaux asumió que el área aparente del disco de una estrella es proporcional al inverso del cuadrado de la distancia, esto es, que al duplicar la distancia el área aparente de la estrella se reduce a un cuarto. Por lo que el incremento en el número de estrellas en cada capa sucesiva compensa el efecto de disminución en la luminosidad de cada estrella. En este escenario cada esfera contribuye con la misma cantidad de luz. Pero a diferencia de Halley, el argumento de Chéseaux no conduce a la conclusión de que una infinita cantidad de luz llega a un observador ubicado en el centro de las esferas. Para el astrónomo suizo, las sucesivas capas de estrellas forman a una determinada distancia crítica un trasfondo continuo detrás del cual el resto de las estrellas permanecen ocultas.

Las consideraciones de Chéseaux permiten calcular qué tan lejos es posible penetrar con nuestra visión entre las estrellas distribuidas en capas o esferas concéntricas antes de encontrarse con la superficie de una estrella. Si las estrellas se distribuyen uniformemente como los árboles en un bosque densamente poblado, formarán en algún punto un fondo continuo más allá del cual no es posible penetrar. De acuerdo con los cálculos de Chéseaux, este límite de fondo provocaría un cielo completamente brillante. Calculó que 760 billones ($7,6 \times 10^{14}$) de esferas con un espesor de 4 años luz cada una cubrirían completamente el cielo de estrellas. Las estrellas contenidas en estas esferas ocuparían una esfera de un radio de 3000 billones (3×10^{15}) de años luz. Más allá de este límite, un observador no podría recibir más luz de las estrellas en esferas más externas. Chéseaux asumió además que todos los lugares en el universo son

⁵³ El pasaje relevante, traducido al inglés, se encuentra en Harrison (1987: 221-222).

similares por lo que el resultado puede obtenerse para un observador en cualquier lugar del universo.⁵⁴

Empleando el método fotométrico ideado por James Gregory,⁵⁵ Chéseaux calculó que las estrellas más brillantes se encuentran a 240.000 unidades astronómicas de distancia (aproximadamente 4 años luz). Considerando, por razones de simplicidad, que todas las estrellas son tan brillantes como el Sol y que el área de la esfera celeste es 180 000 veces más grande que el área aparente del Sol, estimó que un hemisferio cubierto de estrellas sería 90 000 veces más brillante que el Sol. La diferencia entre esta conclusión y la experiencia demuestra que o bien el número de las estrellas fijas no es infinito o que la luminosidad decrece más rápidamente que el inverso del cuadrado de la distancia. El segundo disyunto es plausible si se supone que el espacio interestelar está impregnado de algún material capaz de interceptar la luz estelar. Chéseaux consideró que el medio interestelar velaba nuestra visión de este trasfondo continuo. Su solución a la paradoja depende de esta hipótesis: dado que el espacio entre las estrellas no es perfectamente transparente, los rayos de luz de las estrellas más lejanas son absorbidos por el medio mucho antes del límite de fondo, lo que hace que el cielo nocturno se mantenga oscuro.

⁵⁴ Harrison (1987: 90) calcula el límite de fondo calculado por Chéseaux empleando datos actuales. En base a la suposición de que a una distancia de 10 años luz del Sol se encuentran aproximadamente 10 estrellas, estima que el área ocupada por una estrella tiene un volumen de unos 100 años luz cúbicos. Considerando que el radio del Sol es de 700 000 km, su sección transversal es de 1,5 billones ($1,5 \times 10^{12}$) de km cuadrados. Suponiendo que todas las estrellas son similares al Sol si cada estrella ocupa un volumen promedio de 100 años luz cúbicos en las cercanías al Sol, la división de este volumen por la sección transversal de una estrella da como resultado un límite de fondo de 6000 billones (6×10^{15}) de años luz.

⁵⁵ El método consistía en estimar la distancia de las estrellas más brillantes cercanas asumiendo que estas son similares al sol, y comparando su brillo con el de otros planetas: Marte, Júpiter y Saturno. Conociendo el tamaño de los planetas y su distancia estimada respecto del Sol calculó la distancia aproximada de las estrellas más brillantes (Harrison, 1987: 84).

Olbers en un artículo de 1823 titulado “Über die Durchsichtigkeit des Weltraumes” (“Sobre la transparencia del espacio”)⁵⁶ fue el primero en plantear la paradoja para un universo en el cual las estrellas no se distribuyen de manera uniforme. De acuerdo a su formulación, incluso si se asume que las estrellas están agrupadas en galaxias como la Vía Láctea, y si su número es infinito, la línea de visión de un observador, cualquiera sea la dirección hacia la que se dirija, debería alcanzar la superficie de una estrella en algún punto. El planteo de la paradoja para una distribución no uniforme de las estrellas también puede ser comprendido a partir de la analogía del bosque: las estrellas agrupadas en galaxias son similares a agrupaciones de árboles, pequeños bosques, que componen un gran bosque. La superposición de muchos pequeños bosques, en un gran bosque infinito, genera un fondo continuo, de manera que una línea trazada desde nuestra visión situada dentro del bosque se encontrará necesariamente con un árbol. La solución a la paradoja es también similar a la de Chéseaux: dado que la bóveda celeste no es en todos sus puntos tan brillante como el Sol, el espacio interestelar no debe ser completamente transparente y en él debe ocurrir la absorción de la luz de las estrellas más lejanas, sumiendo al espacio infinito en la oscuridad, como un bosque sumergido en la niebla.

Olbers calculó, empleando el método fotométrico de Gregory, que las estrellas más cercanas se encuentran a 350 000 unidades astronómicas o 5 años luz (aproximadamente). Para resolver la paradoja propuso un escenario similar al de Halley: una serie de esferas concéntricas del mismo espesor en las que se encuentran distribuidas las estrellas agrupadas en galaxias. La superposición de estas capaz crearía un límite de fondo 90 000 veces más brillante que el Sol, por lo que postuló que alguna materia interestelar interceptaba la luz de las estrellas mucho antes del límite de fondo. Debido a que la física del calor no se hallaba suficientemente desarrollada en su momento, las soluciones de Chéseaux y Olbers ignoraron el hecho de que la absorción interestelar produciría el

⁵⁶ Los pasajes relevantes, en traducción inglesa, están reimprimos en Harrison (1987: 223-226).

calentamiento del medio que finalmente generaría un cielo brillante. Por lo tanto, sus soluciones fallan, por razones termodinámicas, en proporcionar una solución satisfactoria a la paradoja.

8.4 La paradoja de Olbers y la velocidad de la luz

Según Harrison, el tipo de soluciones *correctas* a la paradoja dentro de la cosmología newtoniana, fue presentado por primera vez en forma cualitativa por Edgar Allan Poe en 1848 y en forma cuantitativa por Kelvin en 1901. Las mismas se inscriben, dentro de la clasificación de Harrison, entre las interpretaciones del *cielo no cubierto* y aplican a la mayoría de modelos cosmológicos (Harrison, 1990). Cualitativamente, esta clase de interpretación de la paradoja correlaciona la distancia entre el observador y la vida luminosa de las estrellas, es decir, el tiempo finito durante el cual una estrella emite radiación. La solución cuantitativa se vincula con la hipótesis de Ole Røemer, formulada en 1676, según la cual la luz se propaga con una velocidad finita, lo que significa que cuando miramos en el espacio miramos atrás en el tiempo, es decir, que cuanto más lejano se encuentra un objeto, tanto más antigua es la imagen que de él recibimos. Si el universo es espacialmente infinito, cada observador puede ver solo un número finito de estrellas que se extiende hasta la distancia determinada por la vida luminosa de las estrellas o el límite visible del universo.

La primera solución clara y correcta del enigma, aunque solo cualitativamente expresada, provino de Edgar Allan Poe. La solución publicada en *Eureka*, un ensayo de cosmología especulativa, en 1848, establece que una manera de comprender los vacíos en un universo infinitamente poblado de estrellas es suponiendo que la distancia del fondo infinito es tan inmensa que no nos llega desde allí ningún rayo de luz.

No hay falacia astronómica más insostenible, y ninguna ha sido apoyada con más pertinacia, que la de la absoluta ilimitación del universo astral. Las razones que sustentan la limitación, como ya las he enunciado, me parecen *a priori* irrefutables; pero, para no hablar más de éstas, la observación nos asegura que hay en numerosas direcciones a nuestro alrededor, si no en todas, un límite positivo, o por lo menos no tenemos base alguna para pensar de otra manera. Si la sucesión de estrellas fuera infinita, el fondo del cielo nos presentaría una luminosidad uniforme, como la desplegada por la Galaxia, pues no podría haber en todo ese fondo ningún punto en el cual no existiera una estrella. En tal estado de cosas, la única manera de comprender los vacíos que nuestros telescopios encuentran en innumerables direcciones sería suponiendo tan inmensa la distancia entre el fondo invisible y nosotros, que ningún rayo de éste hubiera podido alcanzarnos todavía. ¿Quién se atreverá a negar que pueda ser así? Sostengo, simplemente, que no tenemos ni un adarme de razón para creer que sea así. (Poe, [1848] 1972: 41)

Poe relaciona por primera vez la velocidad finita de la luz y la vida finita de las estrellas con la oscuridad de la noche. Desde su punto de vista, nuestra visión se extiende a una distancia limitada, en un universo donde las estrellas han estado brillando por un tiempo limitado. Si el universo es infinito, o al menos muy vasto, solo es posible ver una parte del universo porque la luz de las estrellas no nos alcanza más allá de cierta distancia. Las estrellas situadas más allá del horizonte cósmico son invisibles. La parte del universo que podemos ver no contiene la cantidad suficiente de estrellas para cubrir el cielo. Para el poeta, la luz de las “paredes doradas” aún no nos ha alcanzado.

La idea de que si el universo tiene un pasado finito, entonces, solo podemos ver hasta una distancia finita, es la clave para las soluciones correctas al enigma. Más allá de cierta distancia no es posible observar luz porque ésta no nos ha alcanzado todavía. El universo visible está limitado por un horizonte que se aleja a la velocidad de la luz aumentando la parte accesible a la observación. El universo visible es finito si el universo tiene una edad finita.

Kelvin (1901) suministró un tratamiento cuantitativo del problema de la oscuridad de la noche y proporcionó una respuesta correcta en el marco de un

modelo cosmológico estático y uniforme. Kelvin supuso además que los espacios interestelares no contienen materia opaca que absorba la luz de las estrellas. El planteo de Kelvin muestra que las estrellas en la Vía Láctea son insuficientes para cubrir el cielo. La luz de las estrellas más distantes se encuentra más allá del límite de fondo y el tiempo que toma alcanzar a un observador dentro de este límite es mayor que la vida brillante de las estrellas. Cuando el cielo está completamente cubierto, el universo visible se extiende tan lejos como el límite de fondo y la luminosidad del cielo en todos sus puntos coincide con la luminosidad del disco del Sol. Cuando el tamaño del universo visible es igual al límite de fondo el cielo es en cada punto tan brillante como el Sol.

Para formular la paradoja, Kelvin asumió que la Tierra ocupa un lugar central en la galaxia compuesta de mil millones de estrellas tan brillantes como el Sol y que tiene un radio de mil parsecs. Por razones de simplicidad, supuso que la separación de las estrellas en el resto de la galaxia es similar a la que tienen en las proximidades de nuestro Sol. Con estos supuestos calculó que el límite de observación es de 3000 billones de años. Kelvin estimó además, a partir de consideraciones sobre la fuente de la energía radiante, previas al descubrimiento de la energía nuclear, que el Sol y otras estrellas similares tienen una vida brillante de entre 10 y 100 millones de años.

En líneas generales, el cálculo de Kelvin relaciona la fracción del cielo cubierto por estrellas y la luminosidad relativa del cielo. Calculó la fracción del cielo cubierto por estrellas dividiendo el tamaño del universo visible (que tomó como el radio de la galaxia) por el límite de fondo (el volumen del área promedio ocupada por una estrella dividida por el área de la sección transversal de la estrella). Argumentó que la fracción de cielo cubierto por discos estelares es equivalente a la luminosidad promedio del cielo estrellado dividida por la luminosidad del disco del Sol.

En el artículo de 1901 "On Ether and Gravitational Matter through Infinite Space",⁵⁷ mostró que la luz de las estrellas más distantes debe viajar 3000 trillones de años para alcanzar la Tierra y que la duración de este viaje es mucho mayor que la vida luminosa de las estrellas. La estimación de Kelvin, empleando datos contemporáneos, permite calcular que la fracción del cielo cubierto por estrellas es menor a un billón (considerando que la vida luminosa de una estrella es de 10^{10} años y que el límite de fondo es de 10^{23}). Si una billonésima parte del cielo está cubierta por estrellas, el cielo estrellado en promedio en cada punto, es una décima de billón tan brillante como el Sol. La solución de Kelvin establece que la cantidad de luz de las estrellas visibles no es suficiente para que el cielo nocturno sea completamente brillante.

Harrison (1990) señala que es llamativo que ninguno de los astrónomos antes de Kelvin, fuera capaz de proponer una solución cuantitativa del problema. Toda la información necesaria para resolver el enigma estaba disponible desde principios del siglo XIX, por lo que es sorprendente que los científicos no fueran capaces de calcular hasta entonces que el número de estrellas es insuficiente para cubrir el cielo (véase Harrison, 1990: 35). Para Harrison, la ausencia de soluciones correctas a la paradoja puede atribuirse a la inhibición en las mentes de los astrónomos que afectó el desarrollo de la cosmología desde el siglo XVII al XIX. Esta inhibición puede formularse como la resistencia a reconocer el hecho de que cuando miramos atrás en el espacio también miramos atrás en el tiempo. Este hecho ha recibido escaso reconocimiento en la historia de la ciencia. Ningún astrónomo luego del descubrimiento de la aberración de la luz estelar por Bradley en 1729 dudaría del resultado de Roemer de que la luz viaja a una velocidad finita. No obstante, las soluciones de Halley, Chéseaux y Olbers, entre otras, asumieron que vemos el cielo como es en el presente y no como era en el pasado, pasando por alto la importancia de la velocidad finita de la luz cuando observamos el cielo. No obstante, señala Harrison, solo con un esfuerzo consciente es posible notar

⁵⁷ Reimpreso en Harrison (1987: 227-228).

quede al observar el cielo se está mirando atrás en el pasado y viendo otros mundos como eran hace muchos años.

8.5 Reformulación del viejo problema. La paradoja como evidencia de la expansión del universo

Desde la década de 1930, la cosmología incorporó la hipótesis de que el universo se expande, por lo que la solución del enigma, si bien incluyó la consideración de la vida luminosa de las estrellas, no se centró en ella. Hermann Bondi (1952) propuso que el corrimiento al rojo de la luz estelar es lo que permite resolver la paradoja de la noche oscura. Las estrellas cubren el cielo nocturno pero la mayoría de ellas no son visibles debido al corrimiento hacia el espectro no visible de la luz que emiten. El enigma de la oscuridad de la noche se transformó de este modo en una evidencia de la expansión del universo. Debido a que el universo se expande, las estrellas distantes no pueden ser vistas. La vinculación con las observaciones pertinentes, en este caso del corrimiento de las líneas espectrales a longitudes de onda más largas, también resultó fundamental para la solución de la paradoja.

El tratamiento de la paradoja de Bondi es cualitativo, en general, y tiende a ser no matemático. Asume el problema tal como fuera formulado por Olbers y examina el argumento tomando como punto de partida las siguientes premisas (Harison, 1990: 38)

- (i) El universo es ilimitado en el espacio y uniformemente poblado de estrellas.
- (ii) El universo es ilimitado en edad y la luminosidad de las estrellas permanece inalterada en el tiempo.
- (iii) El universo es estático, no se contrae ni se expande.

Si estas tres premisas son aceptadas, entonces, la conclusión es que el cielo brilla con luz estelar. Si se supone tal como lo hizo Olbers que línea de la visión que parte de un observador intercepta inevitablemente la superficie de una estrella luminosa y que el cielo está cubierto de discos de estrellas puede seguirse de ello que contiene radiación en todas partes en equilibrio con las estrellas. Bondi asume que el principio cosmológico consiste en (i) y el principio cosmológico perfecto en (i) y (ii) juntos. De acuerdo con el primero nuestra ubicación en el universo no es privilegiada y este debe resultar, a gran escala, isótropo y homogéneo. De acuerdo con el segundo, el universo no solo es isótropo y homogéneo para todo posible observador ubicado en cualquier punto del espacio, sino también en todo tiempo, por lo que debe presentar el mismo aspecto que tiene en el presente en cualquier momento del pasado y del futuro. Este segundo principio es subyacente al modelo del universo estacionario que Bondi y Gold formularon en 1948. En la teoría del *Big Bang* caliente, en cambio, el principio cosmológico perfecto no resulta válido, ya que las propiedades del universo cambian con el tiempo, por ejemplo, se hace menos denso a medida que se expande. Su propuesta de solución asume (i) y (ii), pero sostiene que debido a la evidencia disponible en el contexto de la cosmología de mediados del siglo XX, (iii) ya no puede considerarse correcto.

La solución de Bondi, postula una respuesta al enigma que es correcta para un universo en estado estacionario que se expande. Su análisis atiende al fenómeno de extensión de la longitud de onda de la luz de una estrella más allá del espectro visible y al efecto Doppler-Fizeau que permite medir la velocidad radial de fuga de las estrellas por métodos espectrográficos y nos da noticia de la expansión del universo. De acuerdo con Bondi (1952), si las estrellas distantes se alejan rápidamente por efecto de la expansión del universo, la luz emitida por ellas se nos aparecerá corrida hacia el rojo del espectro y habrá perdido parte de su energía. El cielo estrellado en un universo que se expande es oscuro debido al estiramiento cósmico de la luz procedente desde galaxias distantes.

Desde la perspectiva de Harrison (2000: 504), la propuesta de Bondi no constituye una auténtica respuesta al enigma de la oscuridad de la noche. Según su clasificación, la solución que depende de la expansión del universo se encuentra dentro del tipo de soluciones del cielo cubierto. Estas asumen, como se ha mencionado, que el cielo está cubierto totalmente por estrellas pero que estas no son visibles debido a la expansión del universo que provoca el corrimiento hacia el rojo de la luz estelar. Pero, tal como mostrara Kelvin, en un universo newtoniano el cielo es oscuro porque no contiene energía suficiente para producir un cielo nocturno completamente brillante. De manera que la hipótesis de la expansión simplemente hace al cielo nocturno más oscuro de lo que sería en un universo estático y uniforme. Debido a que la vida luminosa de las estrellas es menor que el límite de fondo, la hipótesis de un cielo cubierto de estrellas es incorrecta.

8.6 La caja cósmica y la cosmología relativista

Harrison (1987) propuso una solución de la paradoja en el contexto de la física relativista. Su solución supone que, usando la equivalencia entre masa y energía, es posible mostrar que no existe en el universo suficiente energía para crear un cielo brillante. Como argumento a favor de esta hipótesis propuso un experimento mental. Formuló la solución a la paradoja por primera vez en 1964, en una carta al editor de la revista *Nature*. Luego publicó una breve historia de la paradoja con su propuesta de solución en un texto de 1981. En este mismo texto dedicó un capítulo a explicitar el funcionamiento de la caja cósmica. Luego, en 1987 publicó un trabajo enteramente dedicado al estudio de la paradoja, con un apéndice que contiene fragmentos de los textos originales donde aparecieron todas las soluciones que el autor considera históricamente relevantes.

El experimento mental de la caja cósmica consiste en imaginar una estrella dentro de una caja, rodeada por paredes perfectamente reflectantes (Fig. 8.2).

Esta es una idealización crucial para el funcionamiento del experimento. Las paredes de la caja de Harrison no absorben nada de la luz, aunque sabemos que no existen materiales perfectamente reflectantes y que en una habitación enteramente cubierta por espejos absorbe toda la luz de un flash en fracciones de segundo. Asimismo, el experimento supone que las estrellas se distribuyen de manera que, en promedio, cada estrella ocupe una unidad de volumen V (por ejemplo un parsec cúbico). La caja cósmica que encierra una estrella debe tener, por lo tanto, un volumen igual a V . Por último, se asume que la luz emitida por una estrella en lugar de esparcirse por el universo rebota entre las paredes dentro de la caja cósmica.

En un universo completamente brillante, como el que plantea la paradoja de Olbers, cada estrella emite tanta radiación como recibe. La radiación llena el espacio a un nivel que iguala el de la superficie de las estrellas. La temperatura en cada punto del espacio es la misma que en la superficie de las estrellas (aproximadamente, 6000 K° , considerando el Sol como estrella típica).

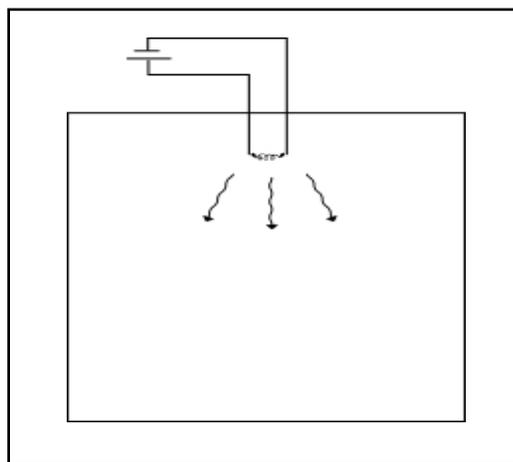


Figura 8.2

La idea central del experimento mental de Harrison es que cuando miramos hacia afuera en el espacio miramos hacia atrás en el tiempo y las cosas que vemos a grandes distancias son similares a las que han existido en esta parte del universo hace mucho tiempo. El escenario a miles de millones de años luz tal

como lo vemos es el mismo que aquí hace miles de millones de años. De manera que si fuera posible aislar una región del universo y encontrar alguna manera de “suplir el efecto” que otras regiones del universo tienen sobre el área que tomamos de muestra, podríamos construir un dispositivo para examinar algunos problemas cosmológicos. De acuerdo con Harrison, hay una manera de llevar el universo al laboratorio. Si imaginamos nuestra región del universo o cualquier otra, aislada de las demás por paredes reflectantes, la región seleccionada queda encerrada en una caja cósmica. Los rayos de luz emitidos dentro de la caja son reflejados hacia un lado y otro por las paredes reflectantes y no pueden escapar. Las cosas dentro de la caja son influenciadas por la luz emitida dentro de la caja y en promedio esta luz es la misma que normalmente recibirían de estrellas lejanas. Hemos tenido éxito entonces en aislar una región de muestra de manera que su estado actual es influido por sus condiciones pasadas y estas condiciones son idénticas a aquellas existentes en cualquier otra parte. Las paredes reflectantes de la caja cósmica deben ser perfectas en todo sentido: no deben transmitir ni absorber nada, y deben reflejar todo como la luz, ondas gravitacionales, partículas, neutrinos y cualquier otra cosa que se mueva de un lugar a otro en el universo (Harrison, 2000: 339). Vale la pena señalar nuevamente que esta es una idealización muy fuerte ya que no conocemos ningún material que se comporte de esta manera. No obstante de su aceptación depende en gran parte el funcionamiento del experimento.

Harrison afirma que lo que ocurre dentro de la caja cósmica es similar a lo que ocurre en cualquier lugar del universo. Los rayos de luz que normalmente proceden de estrellas lejanas recorren distancias similares dentro de la caja cósmica debido a que se reflejan en sus paredes. Un universo particionado funciona exactamente como un universo sin particiones, porque asumimos que las particiones no tienen masa y que su intersección no puede alterar el comportamiento dinámico del universo. Los contenidos de las celdas están en estados similares y en los mismos estados que cuando no tenían particiones. Por consiguiente, los observadores dentro y fuera de la caja cósmica perciben

esencialmente el mismo escenario (Fig. 8.3) Esta estrategia puede emplearse para investigar diferentes problemas; la única condición impuesta en cada caso es que cada celda debe ser tan grande como el objeto que se propone investigar. Si lo que se busca estudiar es el comportamiento de la radiación, la caja debe ser lo suficientemente grande como para contener las longitudes de onda más largas que sean de interés.

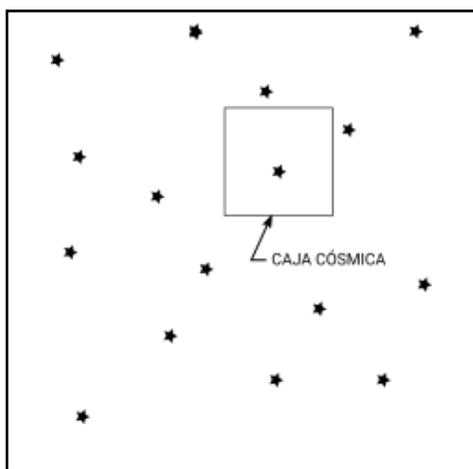


Figura 8.3

Una de las principales ventajas de usar esta estrategia para la investigación de problemas cosmológicos es que habilita a emplear la geometría euclídea. El espacio se puede curvar con geometría esférica o hiperbólica pero permanecer plano (muy aproximadamente euclídeo) en cualquier región pequeña. Lo que es adecuadamente explicable en los confines de una caja cósmica es esencialmente independiente de la geometría a gran escala del espacio. En palabras de Harrison: "Solo necesitamos estudiar lo que ocurre en una región del espacio relativamente pequeña rodeada de paredes perfectamente reflectantes para saber lo que ocurre en todas partes" (2000: 341).

La caja cósmica permite plantear y resolver la paradoja de la noche oscura para modelos donde el universo es estático, se expande y se contrae. Para un universo estático, en el que se supone que las estrellas están uniformemente

distribuidas, cada estrella está encerrada en una caja de igual volumen. Cada estrella dentro de su caja retiene su radiación y llena su propia región del espacio. Supone que la distancia que recorren los rayos de las estrellas cuando no están confinadas a una caja es, en promedio, el mismo que recorren los rayos de luz dentro de la caja, rebotando en las paredes reflectantes antes de ser absorbidos. Así el tiempo de llenado de la caja cósmica puede ser fácilmente calculado y puesto en relación con la vida brillante de cada estrella. En este escenario, las estrellas no brillan lo suficiente como para llenar su caja cósmica. El tiempo promedio que viaja un rayo de luz es el límite temporal (*lookback limit*) y este es el tiempo de llenado de la caja. El tiempo de llenado de la caja cósmica es igual al tiempo que le toma llenar al universo de radiación, que Harrison estima en 100 000 millones de billones (10^{23}) de años. Una estrella típica brilla por 10 000 millones (10^{10}) de años, de manera que el universo no puede ser llenado de radiación.

Los múltiples reflejos de las estrellas en las paredes espejadas imitan el universo con muchas estrellas. El tiempo de llenado de la caja es el tiempo que le toma a la luz alcanzar un límite de fondo. Este argumento es en esencia, el mismo que había empleado Kelvin para relacionar la radiación en la superficie de una estrella y la cantidad de radiación necesaria para cubrir el cielo con luz.

Este experimento mental muestra por qué los espacios oscuros deben existir entre las estrellas en un universo que no contiene suficiente energía para crear un cielo brillante. A causa del tiempo finito de vida de las estrellas luminosas y de la velocidad finita de la luz, el número de las estrellas visibles es demasiado escaso para cubrir todo el cielo. La inmensa mayoría de las estrellas están tan alejadas que la luz que emiten todavía no ha llegado a un observador terrestre; asimismo, el tiempo de viaje de la luz emitida por estas estrellas distantes es mayor que el de su vida luminosa. El tiempo de vida de cada estrella es mucho menor que el tiempo requerido para llenar el espacio con radiación hasta que éste alcance una temperatura semejante a la que tiene en la superficie de las estrellas. Finalmente, las estrellas no contienen suficiente energía para llenar el espacio con

radiación a esta elevada temperatura. Más precisamente si pudiéramos aniquilar toda la materia en el universo y convertirla en radiación térmica, obtendríamos una temperatura de 20 grados Kelvin. Para crear un cielo completamente brillante se requieren 10 000 millones (10^{10}) de estrellas por cada estrella que brilla actualmente dado que, durante su vida luminosa las estrellas convierten una milésima parte de su masa en energía luminosa. En conclusión, el cielo nocturno es oscuro porque la luz de las estrellas es demasiado débil como para llenar el universo.

El mismo experimento puede ser formulado para una caja cósmica que se expande. A un escenario como el descrito anteriormente se agrega que la caja que contiene cada estrella se expande con el universo. En estas condiciones cada vez que un rayo de luz es reflejado por una pared es levemente corrido a rojo (se estira la longitud de onda de los rayos) por el efecto Doppler. En un universo en expansión, al igual que en uno estático, también ocurre que la noche es oscura porque la luz es demasiado débil para llenar el cielo. Pero, de acuerdo con Harrison, la luz estelar dentro de una caja que se expande no es mucho más débil que la luz en una caja estática. Según sus cálculos, la radiación en el interior de una caja que se expande hasta un volumen V , cualquiera sea su volumen inicial, nunca es menor que la mitad de la radiación contenida en un caja estática de volumen V (Harrison, 1987: 200-202 y Harrison, 2000: 505-506). La expansión del universo ciertamente contribuye al oscurecimiento del cielo, pero en una medida mucho menor que la requerida para evitar la paradoja de Olbers. La expansión no puede, por tanto, ser la respuesta al enigma excepto en un universo estacionario que viola la conservación de la energía.

Un cielo completamente brillante podría crearse ajustando las variables del experimento. Si se ajusta la separación entre las estrellas y éstas se ubican más cerca una de otras, se acorta también el tiempo de llenado de la caja, o en otras palabras, se acorta también el límite de fondo.

El experimento mental de Harrison puede evaluarse de acuerdo con los criterios de adecuación formulados en el capítulo 6 de este trabajo. En primer lugar, es importante señalar que el objetivo cognitivo del experimento es evaluar la hipótesis según la cual la densidad de materia del universo actual es demasiado baja como para poder elevar su temperatura a más de 20 K°. En relación con esta hipótesis, puede decirse que el experimento es exitoso en proporcionar una respuesta al problema a partir del cual se plantea. Además, puede admitirse que Harrison hace un uso apologético del experimento mental. En segundo lugar, puede señalarse que el experimento es de carácter contrafáctico. Sus presupuestos son irrealizables, incluso como experimento real a pequeña escala. No obstante, el uso de estos supuestos puede considerarse legítimo en tanto su descripción específica los particulares involucrados, su dinámica y los procesos se desarrollan bajo una teoría aplicable al escenario.

Por otra parte, las conclusiones del experimento son fácticas. El experimento muestra que no existe suficiente energía en el universo para crear un cielo nocturno completamente brillante. Asimismo la efectividad del experimento depende de varias premisas fácticas. Primero, de la observación de que el cielo nocturno es oscuro. Segundo, de que la temperatura promedio del universo es mucho más baja que la temperatura de la superficie de una estrella. Tercero, de que la temperatura promedio de una estrella es de entre 2000 y 30 000 grados Kelvin. Cuarto, de que la vida promedio de una estrella es de 10 000 millones de años. Otras premisas empíricas son: la velocidad finita a la que viaja la luz y la edad estimada del universo.

Además el experimento depende de una serie de idealizaciones. En primer lugar se supone que las variaciones en la luminosidad y la temperatura de las estrellas a lo largo de su vida luminosa pueden ser ignoradas. En segundo lugar, estima que la separación entre las estrellas es homogénea en la región del espacio considerada y que esto puede suponerse respecto del universo en su totalidad. En tercer lugar, el aparato experimental supone la construcción de una caja con

paredes perfectamente reflectantes, lo cual implica una idealización respecto de lo que es posible lograr empíricamente en las mejores condiciones ya que los rayos de luz son absorbidos rápidamente en espejos comunes. No obstante, en la narrativa del experimento queda claro de qué manera estas distorsiones afectan al desarrollo del escenario dinámico descrito. Asimismo, todas las variables y parámetros relevantes para su ejecución están adecuadamente identificadas y su correlación apropiadamente establecida.

Finalmente, la caja cósmica puede interpretarse como un caso dirigido a mostrar que algunas de las hipótesis involucradas en la formulación de la paradoja de Olbers son falsas. Específicamente, la interpretación de Harrison intenta mostrar que el supuesto de acuerdo con el cual en un universo homogéneo un observador debería interceptar una estrella en cualquier línea que se proyectara desde su visión es falso. Existe un límite del universo visible formado por un número finito de estrellas cuya luz ha tenido tiempo de alcanzar al observador. Las estrellas más allá de ese horizonte no pueden ser vistas. Esta interpretación se apoya al mismo tiempo en otras consideraciones que fueron desatendidas por las primeras formulaciones e intentos de solución de la paradoja. Las perspectivas referidas previas al siglo XVIII consideraban que la luz se propagaba de manera instantánea y suponían un universo eterno que no evoluciona temporalmente. Por otra parte, el experimento de Harrison muestra que la expansión del universo no es en sí misma una respuesta al enigma de la oscuridad de la noche. El cielo nocturno sería oscuro incluso si el universo no se expandiera, debido a que el número de estrellas y su vida luminosa no son suficientes para llenarlo de luz.

8.7 Conclusiones

Los experimentos mentales empleados en la formulación y soluciones de la paradoja de Olbers exhiben algunas de las características que se ha intentado mostrar son propias de esta clase de experimentos. Los casos revisados parten

de un conjunto de supuestos teóricos, emplean idealizaciones, abstracciones y distorsiones para producir un modelo del universo a gran escala en el cual, se ilustra y se explica el fenómeno en cuestión. Algunos de estos casos involucran además la representación visual del escenario en cuestión. Los experimentos mentales utilizados para mostrar la contradicción entre algunos principios teóricos y la observación de la noche oscura se caracterizan, además, por la instantaneidad con la cual se derivan resultados de su ejecución. Asimismo algunos de estos experimentos pueden ser reinterpretados a partir de otros principios teóricos y empleados con fines diferentes de los que fueron formulados (piénsese por ejemplo en el caso de Halley reinterpretado a partir de las consideraciones introducidas por Kelvin). Por otra parte, los criterios de validación vinculados a la manipulación de variables y a las características que debe poseer el escenario descrito permiten explicar por qué algunos de los casos aquí referidos fallan (esto se aplica a todas las interpretaciones del cielo cubierto).

El caso ideado por Harrison exhibe claramente las características que hemos atribuido a los experimentos mentales. En primer lugar, depende de la construcción de un modelo que permite representar la evolución de un sistema. En este caso, la caja cósmica permite razonar analógicamente sobre la cantidad de luz estelar que puede ser recibida por un observador en el universo. En segundo lugar, emplea una serie de idealizaciones que se introducen en vistas a la tratabilidad del problema. Como se señaló, estas incluyen simplificaciones acerca de la distribución de las estrellas y las variaciones en su luminosidad, entre otras. De manera semejante a lo que ocurre en algunas simulaciones, el objetivo (*target*) del experimento no es un fenómeno real, sino uno que representa ciertas características del fenómeno sobre el cual se pretenden extraer conclusiones. El caso funciona, como se ha señalado respecto de las simulaciones computacionales y otros modelos, a partir de un criterio de representación selectiva. Aunque el experimento es contrafáctico, y por lo tanto irrealizable en principio, los supuestos respecto de lo que ocurriría con el sistema en las

condiciones descritas son nomológicamente posibles dentro del marco teórico establecido por Harrison.

Su vinculación con los experimentos mentales formulados previamente para dar solución a la paradoja de Olbers permite apreciar también algunas de las características del caso que ejemplifican nuestro análisis de los experimentos mentales. Por un lado, el experimento identifica claramente el conjunto de supuestos teóricos que son asumidos para la construcción del escenario. Esta no es una virtud compartida por las primeras formulaciones de la paradoja. Por otro lado, toma información empírica como parte de sus premisas, entre ellas: la temperatura promedio del universo, la velocidad de la luz y la vida promedio de una estrella. Estos elementos pueden asociarse al aspecto “experimental” de este caso. El estado actual de desarrollo de la cosmología observacional proporciona una información que no estaba disponible para los astrónomos que propusieron los primeros intentos de solución. La descripción de las condiciones iniciales a partir de estos datos y la clarificación del marco nomológico que se asume, contribuyen a la perspicuidad con la que es posible anticipar el desarrollo del escenario. Por otra parte, esto le permite a Harrison vincular el caso con otras hipótesis rivales. El experimento muestra por qué la expansión del universo no es una hipótesis relevante para explicar la oscuridad del cielo nocturno y, por tanto, resolver la paradoja. Un defensor del modelo del estado estacionario, en principio, podría aceptar la descripción del escenario de la caja cósmica y las conclusiones que Harrison deriva del caso. El caso de la caja cósmica, según nuestra interpretación, constituye un uso apologético de un experimento mental. No obstante, no está claro si esta es la intención de Harrison. Podría considerarse también como un experimento empleado con fines refutatorios. Para zanjar esta cuestión es necesario reconstruir el contexto de su formulación. El experimento no se formuló específicamente en el marco de una controversia científica, sino que más bien partió de la intención de Harrison de reavivar el interés por el enigma cosmológico de la oscuridad de la noche. Sin embargo, si se revisa el largo

derrotero de contribuciones sobre este tema, el cosmólogo manifiesta su preocupación por mostrar que la expansión del universo y la solución propuesta por Bondi son incorrectas. Pero la motivación más importante de su formulación, según afirma el propio Harrison (1990: 44), es la relevancia del enigma para la historia de la astronomía y para la cosmología. La oscuridad de la noche no es el velo de infinitas estrellas ocultas sino el rastro de aquello que existió antes de que se formaran las estrellas. De cualquier modo, el experimento de la caja cósmica y los demás experimentos formulados para dar respuesta a la paradoja de Olbers, constituyen sin duda poderosas herramientas de razonamiento analógico que permiten extender el pensamiento sobre lo que resulta empíricamente inabarcable, nada menos que el universo.

CONCLUSIONES GENERALES

Los experimentos mentales son sin duda una práctica que comparte algunos rasgos importantes con la experimentación. Si bien su uso en el contexto de justificación de las ciencias físicas es menos frecuente y más controversial que el de los experimentos reales, algunos casos han sido utilizados con fines apologéticos y han desempeñado un papel importante en diversos episodios de cambio teórico y conceptual. La vivacidad de los escenarios representados en los experimentos mentales probablemente haya contribuido a la legitimación de esta práctica, a pesar de que la experimentación real se consideró desde la Revolución Científica del siglo XVII como la herramienta por excelencia para escrutar la naturaleza. Dado que su utilidad en muchos casos residió en posibilitar la representación de sistemas muy complejos que no era posible explorar empíricamente, o de sistemas para los cuales no se hallaban teorías disponibles, su contribución como instrumento para la comprensión del mundo natural difícilmente puede ser desdeñada.

Aunque su importancia fue apreciada por algunas perspectivas en la historia de la ciencia, la filosofía analítica de la ciencia del siglo XX subestimó el valor epistémico de los experimentos mentales. Como consecuencia de ello, pocas contribuciones en este campo los tomaron como objeto de estudio hasta principios de la década de 1990. A partir de los tratamientos originales de Brown, Sorensen (1992) y Norton (1991) proliferaron algunas líneas de análisis sobre esta temática, sin embargo una característica de estos desarrollos fue la falta de sistematicidad.

Un primer aporte que el presente trabajo se propuso realizar fue reunir y clasificar los estudios sobre la temática. La segunda contribución fue delimitar apropiadamente el objeto de estudio. Si bien en la actualidad existe un volumen considerable de bibliografía que aborda el problema epistemológico de los

experimentos mentales, la mayoría de las perspectivas elaboran algún aspecto particular de esta cuestión, como su relación con los experimentos ordinarios, o examinan algunos casos históricos sin formular una definición explícita. Aunque en muchos casos la estrategia de recurrir a una caracterización implícita no es un impedimento para el análisis de un problema epistemológico particular, esto resulta en una limitación para extender las conclusiones alcanzadas a otros casos. Dado que la intención de esta investigación fue proporcionar una perspectiva sinóptica sobre el funcionamiento de este tipo de experimentos en el campo de las ciencias físicas, resultó imprescindible comenzar por una definición preliminar de la noción de experimento mental, cuestión que se trató en el capítulo 1. La misma buscó que hiciera foco en los aspectos de esta problemática que consideramos los núcleos de la indagación filosófica. De acuerdo con esta definición, un experimento mental en física consiste en la representación de un escenario posible y en la manipulación imaginaria de ciertas variables. Este ejercicio tiene lugar dentro de un marco nomológico y procede a partir de la postulación de una situación hipotética o contrafáctica y de la descripción de un conjunto de condiciones iniciales que se comunica a través de una narrativa. En ocasiones esta descripción también incluye un relato acerca del funcionamiento de un aparato o instrumento y es acompañada de una representación visual o un gráfico del dispositivo experimental. Su ejecución, sin embargo, no requiere de instrumentos y se lleva a cabo enteramente en la imaginación. Su realización hace posible, en principio, extraer consecuencias acerca del comportamiento de los fenómenos físicos pero, en cada caso, la realización del experimento tiene un objetivo cognitivo determinado: ilustrar una teoría, refutar una hipótesis, proporcionar evidencia a favor de una afirmación, derivar consecuencias inconsistentes o absurdas de un principio teórico o mostrar su plausibilidad. Esta caracterización inicial resulta lo suficientemente amplia como para incluir una variedad de casos históricamente relevantes y lo suficientemente rigurosa como para organizar el trabajo.

Atendiendo a los elementos presentados en la definición, se examinó en el Capítulo 2 en qué consiste la representación de un escenario posible y la manipulación imaginaria de ciertas variables. Para ello, se analizaron las principales analogías positivas y negativas entre experimentación real y experimentación mental. Si bien en la narrativa de los experimentos mentales se describen escenarios y aparatos experimentales y se dan instrucciones que el experimentador debe representarse realizando en primera persona, las actividades de visualización que tienen lugar en dichos experimentos se diferencian del proceso de observación científica. Los experimentos mentales se basan en la visualización y en modos de aprehender el mundo que se asemejan al sentido común. Además no existen criterios claros de corrección que puedan emplearse para evaluar la legitimidad de las visualizaciones que tienen lugar en los experimentos mentales. Por otra parte, las intervenciones que se proponen en la narrativa de los experimentos mentales no son completamente análogas a las que tienen lugar en la experimentación real debido a que el resultado de las mismas depende de las habilidades y la capacidad representativa de los experimentadores, pero no de la “respuesta” de la naturaleza. Asimismo la noción de fenómeno idealizado empleada en algunas perspectivas teóricas se separa de la que se emplea en la experimentación real. Si bien puede aceptarse que los fenómenos son abstracciones, estas parten de la interacción causal con el mundo y como tales no pueden conocerse prescindiendo de ésta. En síntesis las analogías negativas entre experimentos mentales y reales son más relevantes que las positivas. No obstante, esto no implica que sean incapaces de proporcionar conocimiento nuevo, como se mostró a partir del examen del experimento de la doble rendija con electrones, según la formulación de Feynman (1965). El experimento presenta con relativa simplicidad un fenómeno extraño para la experiencia ordinaria, que siempre se refiere a objetos macroscópicos. De este modo, pretende introducir cierta familiaridad con el comportamiento del mundo cuántico. El experimento fue realizado como experimento real mucho tiempo después de su formulación y los resultados confirmaron las predicciones

realizadas por el experimento mental. De acuerdo con Feynman, como no podemos aprender directamente sobre esta clase de fenómenos tenemos que aprender sobre ellos de una manera abstracta o imaginativa y no por conexión con nuestra experiencia directa. Aunque este caso no prueba que los experimentos mentales permitan conocer el mundo solamente a través de un ejercicio especulativo, proporciona evidencia de que se trata de una práctica que no es completamente dependiente de la teoría. La autonomía relativa de los experimentos mentales es una característica que no puede ser explicada de manera satisfactoria a partir de la vinculación con la experimentación real de lo cual se sigue que las relaciones de parentesco con esta clase de experimentos son más lejanas que las postuladas por algunas perspectivas recientes y que un marco más apropiado para comprender cabalmente su funcionamiento es el que proporciona la noción de modelo científico.

La caracterización propuesta de experimento mental afirma que la representación de un escenario imaginario permite realizar inferencias acerca de fenómenos reales, por lo que indagamos la vinculación entre argumentos y experimentos mentales. En el Capítulo 3, se examinó el marco interpretativo que asimila el funcionamiento de los experimentos mentales a argumentos teóricos. Dentro esta perspectiva, se distinguieron dos versiones. Una de carácter *eliminativista*, que se identifica con la posición argumental. De acuerdo con esta versión, los experimentos mentales son argumentos disfrazados, de manera que, si se los reformula adecuadamente, puede eliminarse toda su apariencia experimental sin pérdida de su valor epistémico. La otra versión, que denominada *inferencialista*, sostiene que para evaluar el éxito de un experimento mental es preciso reconstruirlo como un argumento, aunque su poder epistémico no se reduzca al de un argumento. Aunque la reconstrucción argumental de los supuestos teóricos y empíricos que operan en la interpretación del resultado de un experimento mental hace posible una mejor evaluación del mismo, la

perspectiva argumental e inferencialista no consiguen proporcionar una buena explicación de su funcionamiento.

Respecto de la perspectiva argumental, se estableció que si bien proporciona un criterio de confiabilidad aceptable y una explicación económica del funcionamiento de los experimentos mentales, no es una interpretación satisfactoria si atendemos al hecho de que la narrativa de estos experimentos puede dar lugar a diferentes argumentos. La posición argumental no explica cuál de las reconstrucciones posibles del experimento es la más adecuada. Por otra parte, la ejecución correcta de un experimento mental depende de un conjunto de habilidades y conocimiento de fondo por parte del experimentador que se asemejan a aquellas puestas en acción en la ejecución de un experimento real. Su realización depende también de su inserción en un contexto de discusión que incluye supuestos teóricos o metafísicos. En las reconstrucciones argumentales estos elementos están generalmente implícitos y requieren de una contextualización adecuada especificarlos. La identificación de (todas) las premisas relevantes para la inferencia requiere de la reconstrucción el contexto de supuestos en el cual el experimento es presentado como evidencia a favor o en contra de una tesis. La posición de Norton (2004a y 2004b) parece asumir que si los poderes justificatorios de los experimentos mentales residen fundamentalmente en sus argumentos, los experimentos mentales exitosos lo son independientemente de los contextos en los que aparecen, ya que su eficacia depende del contenido empírico de las premisas y de la inferencia en cuestión. Aunque estas no son razones concluyentes para descartar la interpretación argumental, indican que quizás el funcionamiento de los experimentos mentales puede comprenderse mejor si se los asimila a otra clase de prácticas. Por otra parte, la posición argumental tampoco da cuenta de los rasgos que distinguen a los experimentos mentales de otras formas de razonamiento hipotético, ya que no todo argumento contrafáctico constituye un experimento mental. Al mismo tiempo, la posición argumental asume implícitamente la tesis de que solo los

argumentos pueden proporcionar un esquema adecuado para explicar el nuevo conocimiento en los experimentos mentales, en tanto afirma que dado que los experimentos mentales no son experimentos, solo pueden ser argumentos. Finalmente, al asimilar a los experimentos mentales a argumentos disfrazados, esta posición tampoco hace justicia a la importancia histórica de algunos experimentos mentales.

Por otra parte, las perspectivas inferencialistas respaldan la intuición de que la realización de un experimento mental consiste en seguir un procedimiento para arribar a un resultado. Aunque esta es una explicación insuficiente, acierta en señalar que la representación en la imaginación del escenario y de las operaciones descritas en la narrativa del experimento justifica (y no solo motiva) la interpretación que los casos se proponen apoyar. Las perspectivas de Sorensen (1992) y Haggqvist (1996) proporcionan aquí un elemento importante para sostener esta interpretación: la representación de un escenario imaginario genera creencias de manera similar a como los experimentos ordinarios lo hacen. Los eventos físicos que tienen lugar en el laboratorio causan que los observadores profieran ciertos enunciados observacionales. Los eventos psicológicos que tienen lugar dentro de la mente del experimentador que realiza un experimento mental, motivan la aceptación de ciertos enunciados. Si bien estas interpretaciones no permiten reconstruir algunos casos, porque suponen que los experimentos mentales funcionan exclusivamente como refutadores aléticos, develan una serie de supuestos relevantes para examinar la función que los experimentos mentales pueden desempeñar en la elección de teorías. En suma, si bien los experimentos mentales no son meros argumentos, como en el caso de los experimentos reales, están acompañados de argumentos que especifican una interpretación y por ello su reconstrucción detallada puede resultar de utilidad para la comprensión de su funcionamiento.

En el Capítulo 4 se elucidó el origen del conocimiento en los experimentos mentales. El carácter “mental” de estos implica que por definición no emplean

instrumental, no son ejecutados físicamente, ni proporcionan nuevos datos empíricos. Una explicación satisfactoria acerca del fundamento del conocimiento en los experimentos mentales debería, en primer lugar, esclarecer el enigma planteado por Kuhn (1964), es decir, proporcionar una interpretación explícita que muestre cómo una actividad puramente especulativa permite obtener nueva información sobre el mundo físico. Para evaluar las posiciones existentes sobre el origen del nuevo conocimiento en los experimentos mentales, se adoptó un criterio de acuerdo con el cual una buena explicación de esta cuestión, en primer lugar, debe ser *clara* en tanto debe evitar la ambigüedad en las caracterizaciones e identificar de manera perspicua el mecanismo de producción de nuevo conocimiento en cuestión, *no debe asumir compromisos metafísicos innecesarios o superfluos* y debe poder *incorporar* el mecanismo de generación de nuevo conocimiento a mecanismos más generales de producción o ampliación de conocimiento.

Contra la posición argumental, se estableció que su adhesión a las tesis empiristas como fuente del conocimiento no está claramente desarrollada y que la conjunción de la perspectiva argumental con la epistemología empirista conduce a algunas inconsistencias. Podría afirmarse que, desde la perspectiva argumental, la información empírica necesaria para la obtención de nuevo conocimiento debe encontrarse disponible o accesible de alguna manera en el sistema de creencias del experimentador, pues, de otro modo, no podría realizarse una inferencia. Por tanto, si la realización de un experimento mental coincide con una inferencia, la información empírica necesaria debería estar simultáneamente disponible (proposicionalmente formulada) y a la vez implícita y, de cierto modo, inaccesible al experimentador antes de la realización del experimento.

La perspectiva empirista de Mach (1893 y 1897), recuperada por Sorensen, tampoco aporta claridad a la explicación del origen del conocimiento en los experimentos mentales. De acuerdo con este punto de vista, los experimentos mentales funcionan sobre la base de la habilidad evolutivamente adquirida de

especular sobre el comportamiento de los fenómenos imaginados y de derivar a partir de este ejercicio consecuencias sobre el comportamiento de los fenómenos reales. No obstante su atractivo, la tesis evolucionista resulta superflua para la explicación de la capacidad de obtener nuevo conocimiento en los experimentos mentales. Las habilidades relevantes para especular sobre el mundo físico dependen de la inserción en la comunidad científica o se apoyan en una descripción que hace posible recuperar los elementos de la experiencia relevantes. Por otra parte, apelar a la noción de intuición empírica, como en la interpretación empirista de Brendel (2005), no esclarece la manera en que la experiencia previa es articulada en la ejecución de esta clase de experimentos. Esta es una noción oscura, sobre la cual no hay acuerdo entre los filósofos contemporáneos y que suscita diversos problemas filosóficos no resueltos.

El *apriorismo*, por su parte, asume una serie de compromisos metafísicos y epistémicos que es incapaz de justificar. La perspectiva platonista de Brown (2011) se compromete con el realismo nomológico y con una forma de conocimiento *a priori* en las ciencias fácticas. En general, la propuesta de Brown se basa en un estudio de caso que propone como paradigmático del conocimiento en ciencias fácticas. Este argumento es insuficiente y en varios aspectos es inconsistente con la noción de fenómeno físico. En relación con el conocimiento *a priori*, Brown sostiene que, si el platonismo es aceptable en el contexto del conocimiento matemático, entonces, también lo es en el ámbito del conocimiento físico. Este argumento es sumamente problemático. Por una parte, el platonismo matemático no es la única explicación posible del conocimiento en este ámbito (por lo que no es una epistemología que debería aceptarse por *default*) y, por otra parte, Brown no consigue desacreditar la teoría causal del conocimiento para justificar la posibilidad de conocimiento *a priori* de las leyes físicas. Otras razones para considerar que la perspectiva de Brown presenta dificultades importantes son su compromiso con una metafísica inflacionaria y la falta de integración con otras formas de adquisición de conocimiento en la práctica científica.

La posición *trascendentalista*, una variante del apriorismo propuesta por Buzzoni, es criticada también por resultar superflua. De acuerdo con este punto de vista, los experimentos mentales funcionan sobre la base de una facultad de especular sobre el mundo natural que el autor identifica con una forma de conocimiento *a priori* funcional. No resulta necesario que la habilidad de anticipar el comportamiento de la naturaleza dependa de estructuras trascendentales que configuran toda la experiencia posible. Es decir, la misma interpretación sobre cómo ganamos conocimiento en los experimentos mentales podría sostenerse bajo un supuesto empirista, esto es, podría explicarse nuestra capacidad de anticiparnos a la experiencia futura apelando a la experiencia pasada o acumulada.

Las perspectivas *constructivistas*, por su parte, sostienen que los experimentos mentales pueden ayudar a detectar anomalías, mostrando que algunas posibilidades empíricas han sido pasadas por alto o que algunos fenómenos no se pueden explicar empleando las leyes aceptadas dentro de un paradigma. Si bien algunos experimentos mentales son capaces de introducir una reforma conceptual, porque presentan un contraejemplo de alguna ley que forma parte del núcleo duro de una teoría, no todos los experimentos mentales funcionan de esta manera. Por consiguiente, identificar los experimentos mentales con la presentación exitosa de anomalías no permite explicar el funcionamiento de aquellos experimentos mentales que se proponen explorar fenómenos en contextos en los que no existen teorías suficientemente desarrolladas y aceptadas por una comunidad científica.

A partir de estas consideraciones se estableció que si bien la ejecución de (algunos) experimentos mentales amplía el conocimiento sobre la naturaleza empleando conocimiento previamente disponible, este proceso tiene lugar con la concurrencia de diversas facultades, además de la percepción. La imaginación es una facultad que permite explicar el funcionamiento de las inferencias ampliativas y que tiene un papel importante en el razonamiento científico. Aunque no es

posible definirla de manera unívoca, podría ser caracterizada brevemente, como lo hizo Williamson (2016), como la capacidad de formular o suponer posibilidades. En su aplicación al método científico, la imaginación desempeña un papel en los casos en los que el método de ensayo y error no puede llevarse a cabo. El ejercicio de esta facultad no es completamente libre, se encuentra limitado y orientado por sus objetivos y por el conjunto de supuestos s teóricos y empíricos que condicionan su empleo. Dadas ciertas condiciones iniciales empírica y teóricamente orientadas, el desarrollo posible del escenario imaginado está condicionado por estas de modo que sería posible distinguir cuándo las consecuencias que se derivan de esta situación inicial son ilegítimas. En este sentido, aunque se trata de una facultad falible, tanto como la percepción y la memoria, el procedimiento imaginativo (bajo las condiciones previamente estipuladas) constituye un método confiable para formar creencias aproximadamente verdaderas acerca de lo que sucedería en circunstancias hipotéticas. En el caso de los experimentos mentales, lo que compele a la aceptación de ciertas conclusiones y al rechazo de otras es que la decisión final acerca de “qué sucedería” no depende de aquello que el experimentador imagine libre y voluntariamente, sino de lo que, dadas las condiciones antecedentes y el marco teórico asumido, sucedería en tales escenarios. De manera similar a como podemos ganar conocimiento acerca de las relaciones espaciales sin intervenir directamente sobre los objetos, la imaginación permite obtener conocimiento sobre los fenómenos físicos, en condiciones especiales, cuando está orientada por supuestos teóricos bien definidos y suficientemente informada por la experiencia previa.

En el Capítulo 5 se argumentó que la interpretación de los experimentos mentales como *modelos científicos* presenta una serie de ventajas explicativas respecto de las que los equiparan a los experimentos reales y a los argumentos teóricos. Fundamentalmente, proporciona una explicación plausible de cómo los experimentos mentales producen nuevo conocimiento y logra integrar esta

exégesis a la teorización filosófica sobre otras formas de razonamiento científico. Permite elaborar, además, una lectura consistente de su eficacia heurística, un elemento que ha caracterizado a esta clase de experimentos y que ha motivado interpretaciones esotéricas de su funcionamiento. La falta de intervención en el mundo natural es un rasgo que tanto la explicación “experimental” como la “argumental” reconocen como rasgo propio de los experimentos mentales. La manipulación de variables en la imaginación difiere de la intervención material en que la ocurrencia del resultado de dicha manipulación está determinada enteramente en el experimento mental por la descripción del aparato experimental, por las teorías e hipótesis a las que se apela en la narrativa y por el conocimiento de fondo de los experimentadores. Esta es una diferencia importante con los experimentos reales. Los resultados de los experimentos mentales están limitados por los aspectos representados en un sistema idealizado. El tipo de razonamiento que propician hace posible inferir ciertas conclusiones acerca del comportamiento de los sistemas reales y bloquea otras. La descripción de un escenario imaginario, de un aparato experimental o de un mecanismo establece las condiciones iniciales e identifica las variables relevantes. La pregunta planteada en el experimento mental “¿qué pasaría si...?” implica que el experimentador se represente la situación descrita y responda cómo se comportaría el sistema representado en tales condiciones dado cierto marco teórico y nomológico. Este ejercicio permite derivar una “moraleja” o, en términos de Bailer-Jones (2009), un mensaje, que es el resultado de interpretar la situación en cuestión empleando los elementos provistos para ello. Esta es típicamente una forma de razonamiento sustitutivo, diferente de las inferencias lógicas que pueden realizarse a partir de un conjunto de proposiciones lingüísticas. La interpretación de los experimentos mentales científicos como un tipo de modelos científicos permite especificar en qué sentido estas prácticas se vinculan con los fenómenos y esclarecer el tipo de relación que establecen con los sistemas reales de representación selectiva. La noción de representación empleada no involucra que los experimentos mentales reflejan o representan

exhaustivamente los sistemas físicos sino que, a partir de la selección de algunos aspectos, permiten realizar inferencias cuyas conclusiones son aproximadamente verdaderas en el mundo natural, tal como ocurre con los modelos científicos en general.

Dado este marco interpretativo, en el Capítulo 6 se establecieron las condiciones de adecuación de los experimentos mentales. Si bien no es posible establecer un conjunto de condiciones necesarias y suficientes para la evaluación de todos los experimentos mentales, un repertorio de aspectos característicos de estos experimentos debe ser tenido en cuenta con el fin de determinar si un caso tiene éxito en concretar el objetivo cognitivo para el cual se formula. Estas condiciones se vinculan con los distintos aspectos que constituyen el funcionamiento de los experimentos mentales, a saber: la narrativa que describe un escenario imaginario y, en ocasiones, el funcionamiento de un aparato experimental; el objetivo cognitivo del experimento; las representaciones que la descripción evoca y su relación con el universo conceptual y las habilidades relevantes de los experimentadores; los supuestos contrafácticos y las estrategias de idealización empleadas; la interpretación teórica del experimento y la comparación de los resultados del experimento mental con los resultados de experimentos reales análogos. Para establecer criterios de evaluación en cada uno de estos aspectos recurrimos a las teorizaciones filosóficas referidas a experimentos reales y simulaciones computacionales.

Una primera condición de adecuación vinculada con el escenario descrito en la narrativa de los experimentos mentales establece que, para que el experimento proporcione evidencia relevante para una hipótesis, todas las variables y parámetros relevantes para su ejecución deben ser adecuadamente identificadas y la correlación entre éstas apropiadamente establecida. Además, en caso de que el experimento mental se proponga como un contraejemplo a una hipótesis, la posición rival debe ser descrita de manera aceptable para quienes

abogan por dicha teoría rival, de manera que el experimento mental evite apoyarse en simplificaciones o distorsiones excesivas.

Respecto del objetivo cognitivo de los experimentos mentales una de las primeras cuestiones a considerar para evaluar si son exitosos es si el experimento en cuestión proporciona una respuesta genuina a la pregunta que motiva su formulación. Se pueden identificar tipos de razones por los cuales los experimentos fallan en este sentido. En primer lugar, los experimentos mentales pueden fallar por razones vinculadas a la competencia del experimentador. Un experimento mental puede no arrojar un resultado (en absoluto) si el experimentador no es capaz de llevar a cabo el experimento propuesto por la narrativa. En segundo lugar, el experimento mental puede fallar en dar una respuesta debido a que está descrito de manera incompleta o imprecisa, lo que imposibilita que arroje un resultado unívoco. Finalmente, un experimento mental puede fallar porque produce resultados que no son respuestas relevantes para las preguntas que inicialmente motivaron el experimento mental. En estos casos, la hipótesis que el experimento se proponía poner a prueba o explorar permanece indeterminada, aunque el escenario imaginario promueva la exploración de alguna hipótesis alternativa. Por otra parte, para determinar si el caso hace un uso legítimo de los contrafácticos es preciso detallar el experimento en términos de su objetivo cognitivo, es decir, establecer con claridad si el caso se formula para apoyar una hipótesis, proporcionar un contraejemplo, ilustrar una perspectiva, etc.

En lo que respecta a las representaciones evocadas por los experimentos mentales, la cuestión ha sido abordada en la literatura reciente en términos de intuiciones. Aunque este concepto resulta problemático, ya que introduce un conjunto de problemas epistémicos y metafísicos que es necesario aclarar, asociamos las intuiciones a las representaciones que constriñen la estructura a partir de la cual los sujetos conocen el mundo. En el contexto de los experimentos mentales en física, las intuiciones restringen en algún sentido también aquello que pueden considerar como físicamente posible. La interpretación que a partir de

estas observaciones parece razonable asumir es que las intuiciones fondo determinan que un experimento mental solo tenga un valor evidencial relativo a contextos específicos. Las intuiciones solo pueden ser desafiadas o modificadas en casos en los cuales se comparta conocimiento de fondo y se presenten alternativas inesperadas dado cierto marco teórico compartido. Por otra parte, si la descripción del experimento mental introduce supuestos teóricos que hacen plausible una interpretación y socavan otra, entonces, podría asimilarse estos casos a las falacias de petición de principio. Estos casos, identificados como extractores de intuiciones, resultan espurios.

Las analogías positivas que pueden encontrarse entre simulaciones computacionales y experimentos mentales, permiten trasladar algunos de los criterios empleados para evaluar la confiabilidad de las simulaciones a los experimentos mentales. Ambas prácticas dependen de la representación de la evolución de un sistema físico y emplean para ello estrategias de modelización que, aunque varían cualitativamente, desempeñan el mismo papel: hacer tratable un problema en el que la consideración de todas las variables relevantes resulta inasequible. Una de las estrategias de validación de las simulaciones que es trasladable a los experimentos mentales consiste en la comparación de sus resultados con datos previamente disponibles provenientes de la observación y medición de los fenómenos. En dominios en los cuales la experimentación real es imposible o extremadamente difícil, los resultados de los experimentos mentales pueden ser comparados con los resultados de otros experimentos mentales sobre el mismo fenómeno, e incluso vinculados con los resultados de simulaciones computacionales concernientes a estos fenómenos.

En lo que respecta a la interpretación teórica de los resultados de los experimentos mentales, se estableció que es necesario especificar los supuestos relevantes en cada caso y describir el experimento en un lenguaje teóricamente neutral, es decir, en términos que los partidarios de dos teorías rivales acepten. Pero en lo que refiere al trasfondo teórico de un experimento mental, no es

evidente en cuál de las capas del conocimiento de fondo es necesario detenerse para hacer conspicua una interpretación. En cuanto a la descripción del aparato experimental o de la situación hipotética, no es claro que pueda realizarse de manera que permita decidir qué interpretación teórica constituye una mejor explicación de los fenómenos descriptos. En consecuencia, aparte de la sugerencia de explicitar los supuestos relevantes, no es posible establecer un criterio claro acerca de cómo proceder en el caso de interpretaciones rivales de un mismo experimento. La reconstrucción de las inferencias que dan lugar a una interpretación teórica puede favorecer la evaluación del experimento. Por ejemplo, si el experimento mental presenta un contraejemplo a un principio teórico, podría reconstruirse la interpretación del experimento como un *modus tollens*; si el caso presentado se propone como un caso típico y del se infieren conclusiones para un dominio de fenómenos, la interpretación podría reconstruirse como una inferencia ampliativa

La comparación entre experimentos mentales y sus contrapartes reales hace posible reexaminar y esclarecer algunas cuestiones vinculadas a los mecanismos empleados en uno y otro caso para la obtención de resultados. La coincidencia entre predicciones realizadas por experimentos mentales y resultados empíricos ofrece respaldo epistémico a los experimentos mentales. Pero la falta de coincidencia perfecta entre los resultados de los experimentos mentales y reales no es una razón suficiente para dudar de la legitimidad de los primeros, dado que los mecanismos que emplean para obtener conocimiento del mundo natural son diferentes. Si algún criterio general puede considerarse como condición de adecuación de los experimentos mentales a los reales, es el de realizabilidad en principio del experimento mental. Según este criterio, si la secuencia de eventos que el escenario imaginario describe pudiera ocurrir, entonces, se producirían las consecuencias o el fenómeno que el experimento describe. En otras palabras, el criterio señala que un experimento mental aceptable debe capturar adecuadamente las relaciones causales, de manera que

si las situaciones invocadas por la narrativa pudieran tener lugar, sería posible, dado nuestro conocimiento del mundo, decir qué ocurría.

En el Capítulo 7 se elaboró una tipología de los experimentos mentales basada en la noción de *uso*, en contraposición a las clasificaciones existentes que catalogan a los experimentos mentales de acuerdo con sus funciones. Se sostiene que, dado que los experimentos mentales se caracterizan por presentar una narrativa similar a la descripción de un experimento real y están acompañados por un argumento que proporciona la interpretación teórica, el primer elemento puede ser separado del segundo dando lugar a diferentes empleos del experimento relacionados con diferentes objetivos cognitivos. Esta interpretación es sugerida por los casos en los cuales el mismo experimento ha desempeñado funciones opuestas respecto de una misma hipótesis. La tipificación en función de la noción de uso permite explicar, además, los casos en los que los experimentos mentales son presentados como evidencia a favor de una hipótesis científica, cosa que la mayoría de las clasificaciones existentes no pueden hacer sin comprometerse con una metafísica inflacionaria. Además, en los diferentes intentos de catalogar los experimentos mentales, sus funciones se subordinan a la teoría. Todas las perspectivas asumen, con excepción de la de Brown, que los resultados que estos experimentos pueden arrojar son meramente cualitativos y, en general, afirman que dependen de una teoría bien desarrollada para realizar alguna contribución significativa. Lo que, según estos puntos de vista, podemos aprender de los experimentos mentales se limita al descubrimiento de la inadecuación o inaplicabilidad de un concepto teórico, o bien a la posibilidad de derivar consecuencias absurdas o contradictorias con algún principio teórico importante al intentar aplicarlo a escenarios imaginarios.

La interpretación de los experimentos mentales como modelos científicos que se defiende en este trabajo sostiene que estos gozan de autonomía respecto de las teorías al menos en dos aspectos: por un lado no son subsidiarios de la teoría y tienen la capacidad de desempeñar diversas funciones y por otro lado

pueden ser replicados en contextos teóricos diferentes. Según esta lectura, los experimentos mentales son capaces de proporcionar conocimiento nuevo a partir de la presentación de casos no experimentados (en ocasiones imposibles de experimentar por razones físicas o técnicas). El valor evidencial que estos puedan tener se define dentro de un contexto teórico particular en el cual el experimento activa ciertas intuiciones. Explicitar lo más exhaustivamente posible este contexto de supuestos permite actualizar su valor heurístico y evaluar, retrospectivamente, su valor epistémico.

La tipificación de los experimentos mentales basada en la noción de *uso*, en contraposición a las clasificaciones basadas en la noción de *función* o en tipos de resultados, permite dar sentido a su importancia histórica y a los episodios en los cuales un mismo experimento mental fue empleado para defender hipótesis rivales. Esa interpretación apoya la idea de que los experimentos mentales gozan de cierta independencia teórica e ilustra de qué manera su función está definida por su objetivo cognitivo en un marco contextual. Asimismo, hace posible argüir que los experimentos mentales desempeñaron funciones en el contexto de justificación de las teorías. Los experimentos mentales destructivos son ejemplos canónicos de este tipo de usos. Los usos confirmatorios de los experimentos mentales son más controversiales. La perspectiva aquí defendida permite sostener que, en contra de la interpretación proporcionada por la mayoría de los filósofos de la experimentación, los experimentos mentales pueden dar apoyo a hipótesis en los casos en los que estos son capaces de proporcionar modelos fenoménicos apoyados en interpretaciones teóricas. Otros usos de los experimentos mentales menos controversiales son los restringidos al contexto de descubrimiento de las teorías. El uso heurístico de los experimentos mentales corresponde al objetivo de delinear o formular un problema teórico. El uso exploratorio de los experimentos mentales consiste fundamentalmente en ofrecer explicaciones potenciales ideando escenarios que, si fueran reales, darían lugar al fenómeno que constituye el *explanandum* en cuestión.

En el Capítulo 8, se aplican las conclusiones sobre el funcionamiento de los experimentos mentales al caso de la Paradoja de Olbers. Los experimentos mentales empleados en la formulación y soluciones de la paradoja de Olbers exhiben algunas de las características que son propias de esta clase de experimentos. Los casos revisados parten de un conjunto de supuestos teóricos, emplean idealizaciones, abstracciones y distorsiones para producir un modelo del universo a gran escala en el cual se ilustra y se explica el fenómeno en cuestión. Algunos de estos casos involucran además la representación visual del escenario en cuestión. Los experimentos mentales utilizados para mostrar la contradicción entre algunos principios teóricos y la observación de la noche oscura se caracterizan, además, por la instantaneidad con la cual se derivan resultados de su ejecución. Asimismo, algunos de estos experimentos pueden ser reinterpretados a partir de otros principios teóricos y empleados con fines diferentes de los que fueron formulados. Por otra parte, los criterios de validación vinculados a la manipulación de variables y a las características que debe poseer el escenario descrito permiten explicar por qué fracasan algunas de las soluciones propuestas.

En particular, el ejemplo de la caja cósmica ideado por Harrison exhibe claramente las características que les hemos atribuido a los experimentos mentales. En primer lugar, depende de la construcción de un modelo que permite representar la evolución de un sistema y razonar analógicamente sobre la cantidad de luz estelar que puede ser recibida por un observador en el universo. En segundo lugar, emplea una serie de idealizaciones que se introducen en vistas a la tratabilidad del problema. Estas incluyen simplificaciones acerca de la distribución de las estrellas y las variaciones en su luminosidad, entre otras. De manera semejante a lo que ocurre en algunas simulaciones, el objetivo (*target*) del experimento no es un fenómeno real, sino uno que representa ciertas características del fenómeno sobre el cual se pretenden extraer conclusiones. Por otra parte, el caso funciona, como se ha señalado respecto de las simulaciones

computacionales y otros modelos, a partir de un criterio de representación selectiva. Aunque el experimento es contrafáctico, y por lo tanto irrealizable en principio, puede decirse, de acuerdo con los criterios de adecuación ya indicados, que los supuestos respecto de lo que ocurriría con el sistema en las condiciones descritas son legítimos, en tanto el escenario y el mecanismo propuesto son nomológicamente posibles dentro del marco teórico establecido por Harrison.

Los experimentos mentales científicos constituyen una fecunda temática de investigación para la filosofía de la ciencia. El propósito más importante de este trabajo fue explicitar y estructurar un conjunto de problemas que vertebran el análisis de esta cuestión. En relación con este objetivo, se proporcionan explicaciones plausibles y suficientes de algunas de estas problemáticas: el origen del nuevo conocimiento que los experimentos mentales pueden proporcionar, su funcionamiento, sus criterios de adecuación y una tipificación de sus principales usos. El examen de los casos vinculados a la paradoja de Olbers permitió aplicar la perspectiva de análisis elaborada a ejemplares de experimentos mentales no examinados hasta ahora por las teorizaciones existentes. No obstante, debido a la fertilidad de esta temática, muchas cuestiones abiertas quedan planteadas para estudios futuros. Quizás uno de los más importantes es el examen detallado de la vinculación entre experimentos mentales y modelos científicos en general. Si bien explicitamos el mecanismo de conocimiento involucrado en los experimentos mentales y aportamos argumentos para considerarlos como un tipo de modelos científicos, el papel de los experimentos mentales en la construcción de modelos teóricos y computacionales es una temática cuya exploración resulta prometedora. Asimismo, el funcionamiento epistémico de los experimentos mentales puede tomarse como hilo conductor para asistir al tratamiento de un conjunto de problemas centrales en la discusión de la filosofía de la ciencia contemporánea: a) el papel del razonamiento especulativo en las prácticas científicas, b) la construcción de un enfoque integral de la metodología científica, c) la naturaleza

de las leyes científicas, d) el conocimiento de dichas leyes, e) los modelos científicos y su vinculación con las leyes científicas, f) las conceptualizaciones de la noción de fenómeno, g) el papel de la imaginación en el conocimiento científico, h) los alcances del concepto de evidencia científica y i) la noción y clasificación de las idealizaciones científicas. Todas estas constituyen posibles líneas de investigación para las cuales la elucidación de los experimentos mentales presentada en este trabajo resulta relevante.

BIBLIOGRAFÍA

- Achinstein, P. (1968). *Concepts of Science: A Philosophical Analysis*. Baltimore: Johns Hopkins Press.
- Armstrong, D. (1978). *Universals and Scientific Realism. Volume 1*. Cambridge: Cambridge University Press.
- Armstrong, D. (1983). *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Arthur, R. (1999). "On Thought Experiments as A Priori Science". *International Studies in the Philosophy of Science*, 13: 215-229.
- Arthur, R. (2012). "Can Thought Experiments Be Resolved by Experiment? The Case of Aristotle's Wheel". En Brown, J., Frappier, M, y Meynell, L. (eds.), *Thought Experiments in Philosophy, Science and the Arts*, London: Routledge, pp. 107-123.
- Arthur, R. (2019). "Thought Experiments in Newton and Leibniz". En Stuart, M., Fehige, Y. y Brown, J. (eds.), *The Routledge Companion to Thought Experiments*. London: Routledge, pp. 148-164.
- Atkinson, D. (2001). "Experiments and Thought Experiments in Natural Science. The Dynamics of Thought Experiments – A Comment to Atkinson". En Galavotti M. (ed.) *Observation and Experiment in the Natural and Social Sciences*. New York: Kluwer, pp.209-226.
- Bach, R., Pope D., Hwang, L. y Batelaan, H. (2013). "Controlled Double-Slit Electron Diffraction". *New Journal of Physics*, 15. 033018. pp.1-8.
- Bailer-Jones, D. (2009). *Scientific Models in Philosophy of Science*. Pittsburgh: University of Pittsburgh Press.
- Beisbart, C. (2019). "What is a Computer Simulation and What does this Mean for Simulation Validation?". En Beisbart, C. y Saam, N. (eds.), *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*. Cham: Springer, pp. 901-923.
- Bertotti, B. y Balbinot, P. Bergia, S. y Messina, A. (eds.) (1990). *Modern Cosmology in Retrospect*. Cambridge: Cambridge University Press.

- Bishop, M. (1998). "An Epistemological Role for Thought Experiments". En Shanks, N. (ed.), *Idealization IX: Idealization in Contemporary Physics*. Amsterdam: Rodopi, pp. 19-34.
- Bishop, M. (1999). "Why Thought Experiments Are Not Arguments". *Philosophy of Science*, 66: 534-541.
- Bogen, J. y Woodward, J. (1988) "Saving the Phenomena". *The Philosophical Review*, 97: 303-335.
- Bokulich, A. (2001) "Rethinking Thought Experiments". *Perspectives on Science*, 9: 285-307
- Bondi, H. [1952] (1960). *Cosmology*. Second Edition. Cambridge: Cambridge University Press.
- Borge, B. y Mettini, G. (2018). "El estatus epistémico de los experimentos mentales en las ciencias fácticas". *Kriterion*, 140: 341-364.
- Brendel, E. (2004). "Intuition Pumps and the Proper Use of Thought Experiments", *Dialectica* 58: 89-108.
- Brown, H. (1987). *Observation and Objectivity*. New York: Oxford University Press.
- Brown, J. (1991). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. First Edition. London: Routledge.
- Brown, J. (1993). "Why Empiricism Won't Work". *Proceedings of the Philosophy of Science Association*. 2: 271–279.
- Brown, J. (1994). *Smoke and Mirrors: How Science Reflects Reality*. London. Routledge.
- Brown, J. (2004a) "Why Thought Experiments Transcend Empiricism". En Hitchcock, C. (ed.), *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell, pp. 23-43
- Brown, J. (2004b). "Peeking into Plato's Heaven". *Philosophy of Science*, 71: 1126-1138.
- Brown, J. (2008). *Philosophy of Mathematics: A Contemporary Introduction to the World of Proofs and Pictures*, London and New York: Routledge.

- Brown, J. (2011). *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. Second Edition. London: Routledge.
- Brown, J. (2012). "What Do We See in a Thought Experiment?". En Brown, J., Frappier, M. y Meynell, L. (eds.), *Thought Experiments in Philosophy, Science and the Arts*, London: Routledge, pp. 53-68.
- Brown, J., Frappier, M. y Meynell, L. (eds.) (2013). *Thought Experiments in Philosophy, Science and the Arts*. London: Routledge.
- Buridan, J. [1509] (2009). "Impetus and its Applications: Jean Buridan". En McGrew, T., Alspector-Kelly, M. y Allhoff, F. (eds.), *Philosophy of Science: An Historical Anthology*. Oxford: Wiley-Blackwell, pp. 86-90.
- Buzzoni, M. (2016). "Thought Experiments and Computer Simulations". En Magnani L. y Casadio C. (eds.), *Model-Based Reasoning in Science and Technology*. Cham: Springer, pp. 57-78.
- Buzzoni, M. (2009). "Empirical Thought Experiments: A Transcendental-Operational View". *Epistemologia*, 33: 5-26
- Carnap, R. (1936/1937). "Testability and Meaning". *Philosophy of Science*, 3: 419-471, 4: 1-40.
- Cartwright, N. (2010). "Models: Parables vs Fables". En Frigg, R. y Hunter, M. C. eds.), *Beyond Mimesis and Convention: Representation in Art and Science*. New York: Springer, pp. 19-32.
- Cassini, A. (2013). "Modelos, idealizaciones y ficciones: Una crítica del ficcionalismo". *Principia*, 17:345-364.
- Cassini, A. (2015). "Una reivindicación de los experimentos cruciales" *Revista de Filosofía*, 40: 105-137.
- Cassini, A. (2020). *Historia de la ciencia*. Clase N° 15: "El triunfo de la física newtoniana" [Notas de clase].
- Cassini, A. y Redmond, J. (eds.) (2021). *Models and Idealizations in Science*. Cham: Springer.
- Chalmers, A. (2003). "The Theory-Dependence of the Use of Instruments in Science". *Philosophy of Science*, 70: 493–509.
- Chesseaux, J. [1744] (1987). "On the Force of Light and its Propagation in the Ether and the Distances to the Fixed Stars". Reimpreso en Harrison, E.,

Darkness at Night: A Riddle of the Universe. Cambridge: Harvard University Press, pp. 221-222.

Clatterbuck, H. (2013). "The Epistemology of Thought Experiments: A Non-Eliminativist, Non-Platonist Account". *European Journal of Philosophy of Science*, 3: 309–329.

Clausius, R. (1879). *The Mechanical Theory of Heat*. Translated by W. Rowne. London: Macmillan.

Cohen, M. (2005). *Wittgenstein's Beetle and Other Classic Thought Experiments*. Oxford: Blackwell.

Collins, H. (1992). *Changing Order: Replication and Induction in Scientific Practice*. Second Edition, Chicago: The University of Chicago Press.

Collins, H. y Pinch, T. (1993). *The Golem: What You should Know about Science*. Cambridge: Cambridge University Press.

Contessa, D. (2007). "Scientific Representation, Interpretation, and Surrogate Reasoning". *Philosophy of Science*, 74: 48-68.

Cooper, R. (2005). "Thought Experiments". *Metaphilosophy* 36: 328-347.

De Mey, T. (2003). "The Dual Nature View of Thought Experiments". *Philosophica* 72: 61-78.

Descartes, R. [1641] (1967). *Meditaciones metafísicas*. Traducción de E. de Olaso. Buenos Aires: Editorial Sudamericana.

Di Paolo, E., Noble, J. y Bullock, S. (2000). "Simulation Models as Opaque Thought Experiments". En Bedau, M., McCaskill, J. S., Packard, N. H y Rasmussen, S. (eds.), *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life*. Cambridge: The MIT Press, pp. 497-506.

Downes, S. (2021). *Models and Modeling in the Sciences: A Philosophical Introduction*. London and New York: Routledge.

Duhem, P. [1914] (1982). *The Aim and Structure of Physical Theory*. Translated from the Second Edition, by Wiener, P. Princeton: Princeton University Press.

Einstein, A. [1919] (1999). "The Foundation of the General Theory of Relativity". En Kox, M. y Schulman, R. (eds.), *The Collected Papers of Albert Einstein. Volume 6*. Princeton: Princeton University Press, pp. 147-200.

- Einstein, A. e Infeld, L. [1938] (2014) *The Evolution of Physics*. Cambridge: Cambridge University Press.
- Einstein, A., Podolsky, B. y Rosen, N. (1935). "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?". *Physical Review*, 47: 777-790.
- El Skaf, R. (2017). "What Notion of Possibility Should We Use in Assessing Scientific Thought Experiments?". *Lato Sensu: Revue De La Société De Philosophie Des Sciences*, 4: 18-30.
- Elliot-Graves, A. y Weisberg, M. (2014). "Idealization". *Philosophy Compass*, 3: 176-185.
- Fehige, Y. y Stuart, M. (2014) "On the Origins of the Philosophy of Thought Experiments: The Forerun". *Perspectives on Science* 22 (2):179-220.
- Feynman, R., Leighton, R. y Sands, M. (1965). *The Feynman Lectures on Physics Volume III*. Reading, MA: Addison-Wesley. https://www.feynmanlectures.caltech.edu/III_toc.html. Consultado el 13 de Julio de 2018.
- Franklin, A. (1984) "The Epistemology of Experiment". *The British Journal for the Philosophy of Science*, 35: 381-401.
- Franklin, A. (1989). *The Neglect of Experiment*. Cambridge: Cambridge University Press.
- Franklin, A. (2002). "Física y experimentación". *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*: 17: 221-242.
- Franklin, L. (2005). "Exploratory Experiments". *Philosophy of Science*, 72: 888-899.
- French, S. y Murphy, A. (2021). "The Value of Surprise in Science". *Erkenntnis*. <https://doi.org/10.1007/s10670-021-00410-z>.
- Frigg, R. (2006). "Scientific Representation and the Semantic View of Theories". *Theoria*, 55: 37-53.
- Frigg, R. y Nguyen, J. (2020). *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Springer. Cham: Springer.
- Galilei, G. [1632] (1967). *Dialogue Concerning the Two Chief World Systems*. Translated by S. Drake. Second Edition. Berkeley: University of California Press.

- Galilei, G. [1638] (1947) *Two New Sciences*. Translated by S. Drake. Madison: University of Wisconsin Press.
- Galison, P. (1987) *How Experiments End*. Chicago: The University of Chicago Press.
- Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: The University of Chicago Press.
- Gelfert, A. (2016). *How to Do Science with Models: A Philosophical Primer*. Cham: Springer.
- Gendler, T. (2000). *Thought Experiment: On the Power and Limits of Imaginary Cases*, Londres: Routledge.
- Gendler, T. (2004). "Thought Experiments Rethought and Reperceived". *Philosophy of Science*, 71: 1152-1164.
- Gendler, T. (2013). *Intuition, Imagination, and Philosophical Methodology*. Oxford: Oxford University Press.
- Giere, R. (2009). "Is Computer Simulation Changing the Face of Experimentation?". *Philosophical Studies*, 143: 59-62.
- Goffi, J. y Roux, S. (2011). "On the Very Idea of a Thought Experiment". En leorodiakonu, K. y Roux, S. (eds.), *Thought Experiments in Methodological and Historical Contexts*. Leiden: Brill, pp. 165-191.
- Gooding, D. (1990). *Experiment and the Making of Meaning: Human Agency in Scientific Observation and Experiment*. Dordrecht: Kluwer.
- Gooding, D. [1984] (1992). "The Procedural Turn; or, Why Do Thought Experiments Work?". En Giere, R. y Feigl, H. (eds.), *Cognitive Models of Science*. Minnesota: University of Minnesota Press, pp. 45-76.
- Gooding, D. (1993). "What is Experimental about Thought Experiments?". *PSA 1992*: 280-290.
- Guth, A. (1997). *The Inflationary Universe*. New York: Basic Books.
- Hacking, I. (1983). *Representing and Intervening*. New York: Cambridge University Press.
- Hacking, I. (1992). "Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding". *PSA 1992*: 302-308.

- Häggqvist, S. (1996). *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International.
- Häggqvist, S. (2009). "A Model for Thought Experiments". *Canadian Journal of Philosophy*, 39: 55-76
- Halley, E. [1721] [1987]. "Of the Infinity of the Sphere of Fixed Stars". Reimpreso en Harrison, E., *Darkness at Night: A Riddle of the Universe*. Cambridge: Harvard University Press, pp. 219-220.
- Halley, E. [1721] (1987). "On the Number, Order, and Light of the Fixed Stars" Reimpreso en Harrison, E., *Darkness at Night: A Riddle of the Universe*. Cambridge: Harvard University Press, pp. 218-219.
- Harrison, E. (1964). "Olbers' Paradox". *Nature*. 204: 271.
- Harrison, E. (1981). *Cosmology: The Science of the Universe*. Cambridge: Cambridge University Press.
- Harrison, E. (1986). "Kelvin on an Old, Celebrated Hypothesis". *Nature*, 32: 417-418.
- Harrison, E. (1990a). "The Dark Night-Sky Riddle, "Olbers's Paradox". Een Bowyer, S. y Leinert, C. (eds.), *The Galactic and Extragalactic Background Radiation*. The Netherlands: IAU, pp. 3-17.
- Harrison, E. (1990b). "Olbers' Paradox in Recent Times". En Bertotti, B. y Balbinot, P. Bergia, S. y Messina, A. (eds.), *Modern Cosmology in Retrospect*. Cambridge: Cambridge University Press
- Harrison, E. (2000). *Cosmology: The Science of the Universe*. Second Edition. Cambridge: Cambridge University Press.
- Hartmann, S. (1996). "The World as a Process: Simulation in the Natural and Social Sciences". En Hegselmann, R., Muller, U. y Troitzsch, K. (eds.), *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Dordrecht: Kluwer, pp. 77-100.
- Hesse, M. (2001). "Models and Analogies". En Newton-Smith, W. (ed.), *A Companion to the Philosophy of Science*, Oxford: Blackwell, pp. 299-307.
- Hitchcock, C. (ed.) (2004). *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell.

- Hon, G. (2003). "An Attempt at a Philosophy of Experiment". En Galavotti, M. (ed.), *Observation and Experiment in the Natural and Social Sciences*. Dordrecht: Kluwer, pp. 259–284.
- Horowitz, T. y Massey, G. (eds.) (1991). *Thought Experiments in Science and Philosophy*. Savage: Rowman and Littlefield.
- Humphreys, P. (1993). "Seven Theses on Thought Experiments". En Earman, J., Janis, A. et. al. (eds.), *Philosophical Problems of the Internal and External Worlds*. Pittsburgh: Pittsburgh University Press, pp. 205-227.
- Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press.
- leorodiakonu, K. y Roux, S. (eds.) (2011). *Thought Experiments in Methodological and Historical Contexts*. Leiden: Brill.
- Irvine, A. (1991). "On the Nature of Thought Experiments in Scientific Reasoning". En Horowitz, T. y Massey, G. (eds.), *Thought Experiments in Science and Philosophy*. Savage: Rowman and Littlefield, pp. 149-166.
- Jammer, M. (1993). *Concepts of Space: The History of Theories of Space in Physics*. Third Edition. New York: Dover,
- Janiak, A. (2014). *Isaac Newton: Philosophical Writings*. Cambridge: Cambridge University Press.
- Janis, A. (1991). "Can Thought Experiments Fail?". En Horowitz, T. y Massey, G. (eds.) *Thought Experiments in Science and Philosophy*. Savage: Rowman and Littlefield, pp.113-118
- Johnson-Laird, P. (1980). "Mental Models in Cognitive Science". *Cognitive Science*, 4: 71-115.
- Johnson-Laird, P. (1982). "The Mental Representation of the Meaning of Words". *Cognition*, 25: 189-211.
- Kirk, G., Raven, J. y Schofield, M. (1987). *Los filósofos presocráticos*. Madrid: Gredos.
- Kosso, P. (1989). *Observability and Observation in Physical Science*. Dordrecht: Kluwer.

- Koyré, A. (1968). "Galileo's Treatise 'De Motu Graviorum': The Use and Abuse of Imaginary Experiment". En *Metaphysics and Measurement*. London: Chapman and Hall, pp. 44-88.
- Koyré, A. (1973). *Études d'histoire de la pensée scientifique*, Paris, Gallimard. Traducción española: *Estudios de historia del pensamiento científico*. México: Siglo XXI, 1991.
- Kuhn, T. [1962] (2004). *La estructura de las revoluciones científicas*. Buenos Aires: Fondo de Cultura Económica.
- Kuhn, T. [1964] (1977). "A Function for Thought Experiments", en: *The Essential Tension*. Chicago: The University of Chicago Press, pp. 240-265.
- Kuhn, T. [1964] (1996). "La función de los experimentos mentales", en: *La tensión esencial*, México, Fondo de Cultura Económica, pp. 17-57.
- Kühne, U. (2009). *A Short History of the Method of Thought Experiments in Science*. <http://philsci-archive.pitt.edu/id/eprint/4700>.
- Ladyman, R. (2008). "Idealizations". En Psillos, S. y Curd, M. (eds), *Routledge Encyclopedia of Philosophy of Science*. New York: Routledge, pp. 359-367.
- Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge: Cambridge University Press.
- Lakatos, I. (1978). *Mathematics, Science and Epistemology: Philosophical Papers 2*. Cambridge: Cambridge University Press.
- Laymon, R. (1991). "Thought Experiments by Stevin, Mach and Gouy: Thought Experiments as Ideals Limits and as Semantics Domains". En Horowitz, T. y Massey, G. (eds.), *Thought Experiments in Science and Philosophy*. Savage: Rowman and Littlefield, pp: 167-191.
- Lenhard, J. (2018). "Thought Experiments and Simulation Experiments: Exploring Hypothetical Worlds". En Stuart, M. y Fehige, Y. y Brown, J. (eds.), *The Routledge Companion to Thought Experiments*. London: Routledge, pp. 554-567.
- Lenhard, J. (2019). *Calculated Surprises. A Philosophy of Computer Simulation*. New York: Oxford University Press.
- Locke, J. [1690] (2005). *Ensayo sobre el entendimiento humano*, Traducción de O'Gorman. E.México: Fondo de Cultura Económica.

- Lucretius [50 A. C.] (1971). *De rerum natura*. Cambridge, MA: Harvard University Press, Loeb Classical Library.
- Mach, E. [1893] (1960). *The Science of Mechanics*. Traducción al inglés de J. McCormack. La Salle: Open Court.
- Mach, E. [1897] (1976). "On Thought Experiments". En *Knowledge and Error*. Dordrecht: Reidel, pp. 134-147.
- Mach, E. [1897] (1973) "On Thought Experiments" Translated by Price, O. and Krilmsky, En *Philosophical Forum* 4/3, pp. 446–457. (trad. del original: "Über Gedankenexperimente", *Zeitschrift für den physikalischen und chemischen Unterricht*, Vol X (1896-1987) pp. 1-5).
- Mach, E. [1905] (1948). *Conocimiento y error*. Buenos Aires: Espasa Calpe.
- Magnani, L., Nersessian, N. y Thagard, P. (eds) (1999). *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer Academic/Plenum Publishers.
- Maxwell, J. (1865). "A Dynamical Theory of the Electromagnetic Field" *Philosophical Transactions of the Royal Society of London*, 155: 459-512.
- Maxwell, J. [1871] (1873). *Theory of Heat*. Third Edition. London: Longman.
- McAllister, J. (1996). "The Evidential Significance of Thought Experiments in Science". *Studies in History and Philosophy of Science*, 27: 233-250.
- McAllister, J. (2004). "Thought Experiments and the Belief in Phenomena". *Philosophy of Science*, 71: 1164-1175.
- McAllister, J. (2012). "Thought Experiments and the Exercise of Imagination". En Brown, J., Frappier, M. y Meynel, L. (eds.), *Thought Experiments in Philosophy, Science and the Arts*. London: Routledge, pp. 11-29.
- McDonough, J. (2021). "Leibniz's Philosophy of Physics", en Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/archives/fall2021/entries/leibniz-physics/>.
- McMullin, E. (1985). "Galilean Idealization". *Studies in History and Philosophy of Science*, 16: 247-273.
- Miščević, N. (1992). "Mental Models and Thought Experiments". *International Studies in the Philosophy of Science*, 6: 215-226.

- Morgan, M. (2003). "Experiment without Material Intervention: Model Experiments, Virtual Experiments, and Virtually Experiments". En Radder, H. (ed.). *The Philosophy of Scientific Experimentation*. Pittsburgh: Pittsburgh University Press, pp. 216-235.
- Morgan, M. y Morrison, M. (1999). *Models as Mediators*. Cambridge: Cambridge University Press.
- Morrison, M. (2015). *Reconstructing Reality: Models, Mathematics, and Simulations*. New York: Oxford University Press.
- Morrison, M. (2009). "Models, Measurement and Computer Simulation". *Philosophical Studies*, 143: 33-57.
- Moue, A., Masavetas, K. y Karayianni, H. (2006). "Tracing the Development of Thought Experiments in the Philosophy of Natural Science". *Journal for General Philosophy of Science*, 37: 61-75.
- Nersessian, N. (1991). "Why Do Thought Experiments Work?". *Proceedings of the Cognitive Science Society*. Hillsdale: Lawrence Erlbaum, 13: 430-438.
- Nersessian, N. (1992). "In the Theoretician Laboratory: Thought Experimenting as Mental Modeling". *PSA 1992*: 291–301.
- Nersessian, N. (1999). "Model-Based Reasoning in Conceptual Change": En Magnani, L., Nersessian, N. y Thagard, P. (eds.), *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer Academic/Plenum Publishers, pp. 5-22.
- Newton, I. [1687] (1934). *Mathematical Principles of Natural Philosophy and his Systems of The World*. Tranlated by A. Motte. Berkeley: University of California Press.
- Norton, J. (1991). "Thought Experiments in Einstein's Work". En Horowitz, T. y Massey, G. (eds.) *Thought Experiments in Science and Philosophy*. Savage: Rowman and Littlefield, pp. 129-148.
- Norton, J. (1996). "Are Thought Experiments Just What You Thought?". *Canadian Journal of Philosophy*, 26: 333-366.
- Norton, J. (2003). "A Material Theory of Induction". *Philosophy of Science*, 70: 647-670.

- Norton, J. (2004a). "Why Thought Experiments do not Transcend Empiricism". En Hitchcock, C. (ed.), *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell, pp. 44-66.
- Norton, J. (2004b). "On Thought Experiments: Is There More to the Argument?" *Philosophy of Science*, 71: 1139-1151.
- Norton, J. (2013). "Chasing the Light: Einstein's Most Famous Thought Experiment". En Brown, J., Frappier, M. y Meynell, L. (eds.), *Thought Experiments in Philosophy, Science and the Arts*. New York: Routledge, pp. 123-140.
- Norton, J. (2015) "Replicability of Experiment". *Theoria*, 30: 229-248.
- Oberkampff, W. y Roy, C. (2010). *Verification and Validation in Scientific Computing*. Cambridge: Cambridge University Press.
- Olbers, H. [1823] (1987). "On the Transparency of Space". Reimpreso en Harrison, E., *Darkness at Night: A Riddle of the Universe*. Cambridge: Harvard University Press, pp. 223-226.
- Olivé, L. y Perez Ransanz, R. (eds.) (1989). *Filosofía de la ciencia: Teoría y observación*. México: Siglo XXI.
- Ørsted, H. [1811] (1998). "First Introduction to General Physics". En *Selected Scientific Works*. Princeton: Princeton University Press, pp. 282-309.
- Palmieri, P. (2018). "Galileo's Thought Experiments: Projective Participation and the Integration of Paradoxes". En Stuart, M., Fehige, Y. y Brown, J. (eds.) *The Routledge Companion to Thought Experiments*. London: Routledge, pp. 92–110.
- Peacock, K. (2018). "Happiest Thoughts: Great Thought Experiments in Modern Physics". En Stuart, M., Fehige, Y. y Brown, J. (eds.) *The Routledge Companion to Thought Experiments*. London: Routledge, pp: 254-288.
- Pickering, A. (ed.) (1992). *Science as Practice and Culture*. Chicago and London: The University of Chicago Press.
- Platón (2007). *Menón*. En *Diálogos II* (traducción española J. Calonge Ruiz, E. Acosta Méndez, F. J. Oliveri y J. L. Calvo. Barcelona: Gredos.
- Poe, E. [1848] (1976). *Eureka: A Prose Poem*. Reproducido en Beaver, H. (ed.) *The Science Fiction of Edgar Allan Poe*. Harmondsworth: Penguin, pp. 205–309.

- Popper, K. [1959] (2002). "On the Use and Misuse of Imaginary Experiments, Especially in Quantum Theory". En *The Logic of Scientific Discovery*. New York: Routledge, pp. 464-480.
- Pozzi, G. et al (2012). "The Young-Feynman Two-Slits Experiment with Single Electrons: Build-up of the Interference Pattern and Arrival-Time Distribution Using a Fast-Readout Pixel Detector". *Ultramicroscopy*, 116: 73-76.
- Putnam, H. [1975] (1984). "The Meaning of 'Meaning'". *Minnesota Studies in the Philosophy of Science* VII: 131-193. Traducción española de Flematti Alcalde, J. *Cuadernos de Crítica*. México: UNAM.
- Quine, W.V. (1990). *Pursuit of Truth*. Cambridge: Harvard University Press.
- Radder, H. (1996). *In and About the World: Philosophical Studies of Science and Technology*. Albany: State University of New York Press.
- Radder, H. (ed.) (2003). *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press.
- Radder, H. (2006). *The World Observed/The World Conceived*. Pittsburgh: University of Pittsburgh Press.
- Rédei, M. (2001). "Thinking about Thought Experiments in Physics: Comment on 'Experiments and Thought Experiments in Natural Science'". En Galavotti M. (ed.), *Observation and Experiment in the Natural and Social Sciences*. New York: Kluwer, pp.237-241.
- Reiss, J. (2002). "Causal Inferences in the Abstract or Seven Myths about Thought Experiments". Technical Report 03/03, Center for Philosophy of Natural and Social Sciences, London School of Economics http://www.lse.ac.uk/CPNSS/pdf/DP_withCover_Causality/CTR03-02-3.pdf. Accessed on 20August, 2011.
- Reiss, J. (2018). "Thought Experiments and Idealization". En Stuart, M.. & Fehige, Y. y Brown, J. (eds.), *The Routledge Companion to Thought Experiments*. London: Routledge, pp: 538-653.
- Rescher, N. (1991). "Thought Experiments in Presocratic Philosophy". En Horowitz, T. y Massey, G. (eds.). *Thought Experiments in Science and Philosophy*. Savage: Rowman and Littlefield, pp: 31-41.
- Rickles, D. (2016). *The Philosophy of Physics*, Cambridge: Polity Press.

- Rowbottom, D. (2014). "Intuitions in Science: Thought Experiments as Argument Pumps". En Booth, A. y Rowbottom, D. (eds.), *Intuitions*. Oxford: Oxford University Press, pp. 119-134.
- Ruiz Sosa, M (2012). *La observación en la palabra: La función de los experimentos imaginarios en el desarrollo de la Física Cuántica. 1927-1936*. [Tesis doctoral] Universidad Autónoma de Barcelona.
- Schrödinger, E (1935). "The Present Situation in Quantum Mechanics". Translated and reprinted in Wheeler, J and Zurek, W. (eds.) *Quantum Theory and Measurement*, Princeton: Princeton University Press, 1983, pp. 152-167.
- Searle, J. (1980). "Minds, Brains and Programs". En Hofstadter, D. y Dennett, D. (eds.), *The Mind's I*. New York: Basic Books, 1981, pp. 353-373.
- Shapere, D. [1982] (1989). "El Concepto de observación en ciencia y en filosofía". En Olivé, L. y Pérez Ransanz, A. (eds.) *Filosofía de la ciencia: Teoría y observación*, México: Siglo XXI Editores - UNAM, 1989, pp. 47-69.
- Shinod, N. K. (2017). "Why Thought Experiments do have a Life of Their Own: Defending the Autonomy of Thought Experimentation Method". *Journal of Indian Council of Philosophical Research*, 34: 75-98.
- Shumelda, M. y Brown, J. (2009). "Thought Experiments and Inertial Motion: A Golden Thread in the Development of Mechanics". *Rivista di Estetica* 42: 71-96.
- Sorensen, R. (1992). *Thought Experiments*. New York: Oxford University Press.
- Sorensen, R. (2013). "Thought Experiments and the Epistemology of Laws". *Canadian Journal of Philosophy*, 22: 15-44.
- Steinle, F. (1997). "Entering New Fields: Exploratory Uses of Experimentation". *Philosophy of Science*, 64: S65-S74.
- Steinle, F. (2002). "Experiments in History and Philosophy of Science". *Perspectives on Science*, 10: 408-432.
- Stöltzner, M. (2003). "The Dynamics of Thought Experiments: A Comment to Atkinson". En Galavotti M. (ed.), *Observation and Experiment in the Natural and Social Sciences*. New York: Kluwer, pp. 243-258.
- Stuart, M., Fehige, Y. y Brown, J. (eds.) (2018). *The Routledge Companion to Thought Experiments*. London: Routledge.

- Suárez, M. (2003). "Scientific Representation: Against Similarity and Isomorphism". *International Studies in the Philosophy of Science*, 17: 225-244.
- Suárez, M. (2004). "An Inferential Conception of Scientific Representation". *Philosophy of Science*, 71: 767-769.
- Suarez, M. (2015). "Deflationary Representation, Inference and Practice". *Studies in History and Philosophy of Science*, 49: 36-47.
- Swoyer, C. (1991). "Structural Representation and Surrogative Reasoning". *Synthese*, 87: 449-508.
- Tallant, J. (2012). "Intuitions in Physics". *Synthese*, 190: 2959-2980.
- Thagard, P. (2014). "Thought Experiments Considered Harmful". *Perspectives on Science*, 22: 288-305.
- Tonomura, A. et al. (1989). "Demonstration of Single Electron Build up of an Interference Pattern". *American Journal of Physics*, 57:117-120.
- Tooley, M. (1977). "The Nature of Laws". *Canadian Journal of Philosophy*, 7: 667-698.
- Tooley, M. (1988). *Causation: A Realist Approach*. Oxford: Oxford University Press.
- Virvidakis, S. (2012). "On Kant's Critique of Thought Experiments in Early Modern Philosophy". En Ierodiakonou, K. y Roux, S. (eds.), *Thought Experiments in Methodological and Historical Contexts*. Leiden: Brill, pp: 27-142.
- Waters, K. (2007). "The Nature and Context of Exploratory Experimentation: An Introduction to Three Case Studies of Exploratory Research". *History and Philosophy of the Life Sciences*, 29: 275-284.
- Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.
- Wilkes, K. (1988). *Real People: Personal Identity without Thought Experiments*. Oxford: Oxford University Press.
- Williams, S. (1868). *Twilight .Hours: A Legacy of Verse*. London: Satrahan & Co. Publisher.

- Williamson, T. (2016). "Knowing by Imagining. En Kind, A. y Kung, P. (eds.), *Knowledge through Imagination*. Oxford: Oxford University Press, pp. 113-123.
- Winsberg, E. (2003). "Simulated Experiments: Methodology for a Virtual World". *Philosophy of Science*, 70:105-125.
- Winsberg, E. (2010). *Science in the Age of Computer Simulation*. Chicago: The University of Chicago Press.
- Wordward, J. (2000). "Data, Phenomena and Reliability". *Philosophy of Science*, 67: S163-S179.
- Young, T. [1804] (1959). "Experimental Demonstration of the General Law of the Interference of Light". En Shamos, M. (ed.), *Great Experiments in Physics*. New York: Dover, pp. 96-101.
- Zuppone, R. (2013). "La vida propia del experimento: Un análisis crítico de la autonomía de la experimentación y su relación con la constitución y la justificación de la evidencia empírica". *Cuadernos de Filosofía*, 61: 109-111.