

**EXPERIMENTOS MENTALES
Y**

FILOSOFÍAS DE SILLÓN

DESAFÍOS · LÍMITES · CRÍTICAS

2017

Rodrigo Alfonso González Fernández

**EXPERIMENTOS MENTALES Y
FILOSOFÍAS DE SILLÓN
DESAFÍOS, LÍMITES, CRÍTICAS**

© Bravo & Allende editores

© Rodrigo Alfonso González Fernández

Registro de Propiedad Intelectual N° A-283150

ISBN: 978-956-307-081-1

Se terminó de imprimir en octubre de 2017

Diseño y Diagramación

Cristián Puentes Riffo

Impreso en Chile / Printed in Chile

A mi hijo, León Alfonso

Presentación

La filosofía, en clave contemporánea, puede verse como una disciplina en la que diversos argumentos y teorías compiten entre sí. Tanto argumentos como teorías buscan establecer conocimiento, e intentan hacerlo en distintos ámbitos de discusión. En efecto, hay argumentos mediante ejemplificación, mediante apelación a la experticia, acerca de causas, deductivos, etc. Todos los esfuerzos por argumentar son de alguna manera válidos en la filosofía, y también en las ciencias, al punto de que muchos de ellos han inspirado al experto y fascinado al lego. Con relación a esto último vale la pena destacar que incluso *Hollywood* se ha nutrido de argumentos y teorías científicas y filosóficas, al punto de exacerbar nuestra imaginación hasta grados impensables. Ciertamente, metáforas y analogías, es decir, comparaciones ingeniosas han servido para explorar problemas tanto filosóficos como científicos, desde que Platón planteó *La Caverna*, a propósito de cómo se alcanza la luz del conocimiento por medio de la razón, instrumento que permite trascender la confusión y ambigüedad de lo sensible.

En relación con el último punto, es importante destacar que literatura y cinematografía convergen no solo para entretener, sino además para el planteamiento de escenarios relevantes para el debate filosófico y científico. Por ejemplo, hay películas como *Flash Gordon*, *20.000 Leagues Under the Sea*, *Journey to the Centre of the Earth*, *The Time Machine*, *Battle of the Worlds*, *Fahrenheit 451*, *2001: A Space Odyssey*, *Doppelgänger*, *Brave New World*, *Blade Runner*, *AI*

Artificial Intelligence, Matrix, I Robot, por nombrar algunas, que son brillantes ejercicios de qué sucedería si ciertas circunstancias se dieran, y más aún, de cómo dichas circunstancias son relevantes para establecer conocimiento. Estimo que la imaginación es crucial para explorar tópicos filosóficos como el estudio del conocimiento, la mente, la persona, la libertad, la responsabilidad, etc. Todos estos tópicos tienen relevancia en la filosofía, y también en la ciencia. Pero este esfuerzo, que no es banal ni trivial, tiene un importante antecedente en la epistemología, es decir, una rama de la filosofía que se aboca al estudio del conocimiento mismo. Dicha rama se concentra en el estudio de los experimentos mentales, un caso típico de *filosofía de sillón*, y ciertamente afin a los escenarios *hollywoodenses*.

Justamente, el presente libro es un esfuerzo colectivo para poner al alcance de legos y expertos los experimentos mentales, una instancia típica de tal filosofía. Pese a su supuesta falta de confiabilidad, creo que los experimentos mentales son poderosas herramientas para el desarrollo de teorías y argumentos capaces de establecer conocimiento mediante la imaginación. Considero, por otra parte, que la filosofía de sillón es un estilo típico de la exploración teórica filosófica, preferentemente *a priori*, esto es, independientemente de la experiencia. En particular, este libro tiene como finalidad dotar al lector de elementos básicos para determinar si la filosofía de sillón atenta contra los desarrollos de la filosofía misma, y de las disciplinas científicas. En consecuencia, el objetivo central de esta compilación de artículos es que el lector *pondere* si los experimentos mentales y las filosofías de sillón son herramientas valiosas en la caja de herramientas del filósofo, y también del científico.

Esta compilación de artículos, que pretende divulgar el debate de los experimentos mentales en el medio nacional, se divide en dos partes, precedidas de sus respectivas introducciones. La primera parte del libro tiene como fin remediar una patente escasez de trabajos en español sobre la epistemología de los experimentos mentales. La filosofía de las ciencias ha tenido singular protagonismo en la

evaluación de las cuestiones epistemológicas y metodológicas ligadas a la experimentación mental. De hecho, toda la primera parte se concentra en seis traducciones, las cuales introducen al lector en qué es y a qué dedica sus esfuerzos la experimentación mental. De esta forma, se le brinda una panorámica general, lo que es, a mi juicio, una instancia necesaria para proceder a ponderar el valor y la relevancia, en ciencia y filosofía, de los experimentos mentales.

La segunda parte, en cambio, aborda un tópico específico en función de lo examinado en la primera: el problema de la confiabilidad de los experimentos mentales como herramientas para examinar hipótesis y teorías. Dicha confiabilidad es particularmente polémica en una disciplina donde la experimentación mental ha sido prolífica: la filosofía de la mente. Diversas discusiones destacan aquí. El problema mente-cuerpo, o la supuesta diferencia metafísica entre la mente y la materia; el problema de las otras mentes, o las dificultades para establecer si podemos conocer las mentes ajenas como la propia; la subjetividad, o la característica propia de los estados mentales de tener 1ª persona; o la plausibilidad de la Inteligencia Artificial, entre otras, han sido enriquecidas por controvertidos experimentos mentales. Pese a sus detractores, dichos experimentos han contribuido a la propuesta de diversos argumentos y teorías, por ejemplo, tanto a favor del fisicismo reduccionista como en contra, esto es, de la teoría según la cual la mente *no es más que* una cosa física.

Es importante destacar que el auge de los experimentos mentales en filosofía de la mente ha significado una revitalización de la *filosofía de sillón*. Se ha reafirmado un nicho de investigación propio de científicos y filósofos, de distintas especializaciones. Pero, obviamente, donde hay filosofía hay debate y disenso, por lo que el rol de los experimentos mentales y la filosofía de sillón han recibido importantes cuestionamientos y críticas, desde el comienzo. Por este motivo, la segunda parte del libro muestra argumentos a favor y en contra de la experimentación mental, tanto en relación con las principales críticas a este método de argumentación, como con algunos

controvertidos problemas abordados por la filosofía de la mente contemporánea. Entre estos destaca, por ejemplo, la plausibilidad del fisicismo reduccionista mencionado más arriba.

Hay que aclarar, ciertamente, que tanto la primera como la segunda parte se dirigen tanto al lego como al experto. Aquél será introducido en una reflexión filosófica relevante en nuestro tiempo, que muestra las ventajas y los límites de una herramienta de análisis y argumentación típica de la filosofía. Se mostrarán, entonces, las ventajas y desventajas de dicha herramienta. El experto, en cambio, podrá revisar argumentos a favor y en contra de la experimentación mental, lo cual contribuirá a fundamentar de manera más acabada el juicio que se hace de esta como filosofía de sillón. Tal juicio, espero, podrá nutrirse de mejor forma a partir de lo discutido en este libro sobre los experimentos mentales y dicho tipo de filosofía.

Cabe destacar que esta obra es un esfuerzo académico colectivo, y que muchos han contribuido desinteresadamente a concretarla. Se ha obtenido permiso para publicar los textos por parte de sus autores. Con relación a tales permisos, estoy particularmente agradecido del Centro de Filosofía de la Ciencia de la Universidad de Pittsburgh, el cual permitió traducir artículos de *Thought Experiments in Science and Philosophy*, un clásico sobre experimentación mental editado por Gerald Massey y Tamara Horowitz. James Brown, John Norton e Ian Hacking, a su vez, también otorgaron permisos para traducir sus trabajos. Por último, filósofos destacados como Timothy Williamson, Daniel Dennett, Hilary Kornblith y Tim Crane generosamente aportaron trabajos inéditos, los cuales figuran en la segunda parte del libro.

Sin el riguroso y arduo trabajo de traducción de Felipe Morales, Guido Vallejos y José Araya este libro no hubiera visto la luz. Aclaro que en las traducciones se ha optado por privilegiar el sentido que el autor quiere dar, respetando, en la medida de lo posible, cómo se expresan tales ideas filosóficas en español. Siempre que se encontró

difícil la traducción de un término, se optó por ponerlo en corchetes con el objeto de que el lector tome en cuenta la expresión original.

Espero que el esfuerzo colectivo de este libro contribuya no solo a introducir al lector en un debate filosófico de nuestro tiempo, sino además a formar un juicio propio y ponderado acerca del valor de la experimentación mental y la filosofía de sillón. En efecto, si esta obra logra mover a la reflexión crítica sobre los tópicos abordados, cruciales para fomentar el análisis y la argumentación, considero que habrá realizado una importante contribución, no solo a la filosofía chilena, sino también a la educación en nuestro medio nacional.

James Brown

University of Toronto

Tim Crane

Cambridge University

Daniel Dennett

Tufts University

Rodrigo González

Universidad de Chile

Ian Hacking

York University

Andrew Irvine

University of Pittsburgh

Allen Janis

University of Pittsburgh

Hilary Kornblith

University of Massachusetts, Amherst

Gerald Massey

University of Pittsburgh

John Norton

University of Pittsburgh

Timothy Williamson

Oxford University

Índice

PARTE I

Cuestiones metodológicas sobre los experimentos mentales: sus alcances y límites.....	17
--	-----------

Experimentos mentales

James Brown.....	27
1. Ejemplos de experimentos mentales.....	27
2. Objeciones a los experimentos mentales.....	31
3. Tipos de experimentos mentales.....	32
4. Algunas posturas recientes sobre los experimentos mentales.....	36

Sobre los Experimentos Mentales:

John D. Norton.....	47
1. Introducción.....	47
2. Un experimento mental erróneo.....	49
3. ¿Por qué argumentos?.....	52
3.1 Dos partes de la tesis justificadas.....	52
3.2 La tesis de la confiabilidad.....	53
4. Las epistemologías de los experimentos mentales comparadas.....	56
4.1 La postura argumental.....	56
4.2 El platonismo de Brown.....	58
4.3 El constructivismo de Szabó Gendler.....	60
4.4 McAllister y la Inercia Evidencial de los Experimentos Mentales en Ciencia.....	61
Apéndice: ¿Por qué una Transformación de Lorentz rota cuerpos oblicuos?.....	64

¿Tienen los experimentos mentales vida propia?

Comentarios sobre James Brown, Nancy Nersessian y David Gooding

Ian Hacking.....	69
------------------	----

Experimentos mentales en el razonamiento científico

Andrew D. Irvine 83

 1. Introducción 83

 2. Los experimentos mentales en la física 87

 3. Los experimentos mentales en la Antigüedad 90

 4. Los experimentos mentales en la astronomía 96

 5. Acerca del valor de los experimentos mentales para la
 ciencia..... 99

Acerca de algunos experimentos mentales viciados

Richard M. Gale 109

¿Pueden fallar los experimentos mentales?

Allen I. Janis 119

Analiticidad por la puerta trasera

Gerald J. Massey 129

 1. La pareja dispareja 129

 2. Dos Dogmas del Empirismo 131

 3. Concebibilidad 133

PARTE II

**¿Son los experimentos mentales y las filosofías de sillón herramientas
confiables?**

Críticos y defensores 149

Conocimiento de Sillón y Naturaleza de la Filosofía

Timothy Williamson 165

¿Hay lugar en epistemología para la filosofía de sillón?

Hilary Kornblith 187

Bombas de Intuiciones

Daniel C. Dennett 215

“Bomba de intuición”

La evolución de una herramienta epistemológica

Rodrigo A. González Fernández 225

1. Introducción 225
2. La Habitación China y la definición original del concepto “bomba de intuición” 227
3. La evolución de “bomba de intuición”: una herramienta exploratoria en el trabajo filosófico 235

Conclusión 245

Conocimiento Subjetivo

Tim Crane 249

1. Introducción 249
2. El argumento del conocimiento 251
3. El argumento del conocimiento y los qualia 253
4. Desafiando la validez del argumento: “la hipótesis de la habilidad” 255
5. Desafiando las premisas 258
6. Hechos físicos y hechos subjetivos 262
7. El fisicismo revisitado y re-descrito 269

PARTE I

Cuestiones metodológicas sobre los experimentos mentales: sus alcances y límites

Introducción

Como cuestión preliminar es importante aclarar qué significa “experimento mental”. El término proviene de la palabra alemana *Gedankenexperiment*, popularizada por Ernst Mach, aunque introducida antes por Georg Lichtenberg (Schildknecht 1990, citado en el ensayo de Brown, de esta compilación). Tal término siempre ha sido traducido al inglés como *thought experiment*. En español prefiero la traducción “experimento mental” a “experimento de pensamiento”, porque aquél enfatiza mejor el rol de la mente como un laboratorio en donde se evalúan hipótesis y teorías, incluso sobre la propia naturaleza de lo mental.

Una definición minimalista de “experimento mental” debería incluir una distinción entre instancias constructivas o positivas y negativas o destructivas. Tanto en ciencia como en filosofía se emplean experimentos mentales cuando se prueba una hipótesis o teoría, y son particularmente útiles cuando se utilizan para aclarar o refutar ideas y argumentos.

Si el experimento sirve para explicitar una teoría o hipótesis, entonces es positivo o constructivo. Esto ocurre cuando el experimento sirve como elemento de apoyo para comprender mejor ciertas ideas, cuestión necesaria en ámbitos especializados y técnicos, como los de las disciplinas científicas. En la filosofía se emplean experimentos mentales por este mismo motivo, y también para captar mejor las complicaciones y consecuencias de una hipótesis, teoría o argumento, muchas veces difíciles de comprender. En consecuencia, tal como

con los experimentos reales, los mentales positivos o constructivos producen mejor comprensión en ciencia y filosofía, al “aterrizar” problemas y ponerlos al alcance de expertos y legos.

Si, por el contrario, un experimento mental se emplea para generar una contradicción flagrante una vez asumida una teoría, entonces cuenta como un experimento mental negativo o destructivo. Este siempre socava una teoría y permite desecharla al exponer una tensión entre ella y una situación imaginada relevante. Tal como Kuhn (1964) sostiene, los experimentos mentales negativos son en general más frecuentes que los positivos, porque apelan a conocimiento compartido para mostrar las tensiones teóricas.

No obstante, del hecho de que los experimentos mentales permitan examinar la naturaleza de las cosas, e incluso la naturaleza de la mente, no se sigue que tengan un estatus privilegiado y que permanezcan indemnes a la crítica del escéptico. Esto es particularmente frecuente en los experimentos mentales negativos: los defensores de la teoría socavada elaboran objeciones contra el experimento mental, e incluso proponen contra experimentos mentales.

De hecho, los experimentos mentales son tan o más polémicos que los experimentos reales, cuestión que lleva a examinar su confiabilidad. En la medida que los experimentos mentales someten a prueba hipótesis y teorías de manera *a priori*, esto es, llevando al experimentador a imaginar situaciones sofisticadas, los científicos *siempre* prefieren los experimentos reales a los mentales. Pero, ¿a qué se debe que los experimentos mentales ostenten una jerarquía más modesta que los reales?

La mejor manera de enfrentar esta pregunta es identificando las circunstancias en que se apela a los experimentos mentales como instancias para exponer una teoría, o para decidir su destino refutándola. Los científicos generalmente recurren a la experimentación mental cuando existen impedimentos técnicos y financieros que impiden llevar a cabo un experimento real. En circunstancias adversas los

científicos usufructúan de su imaginación, ciertamente barata y eficaz. No resulta extraña, entonces, la afición de los filósofos analíticos por los experimentos mentales, toda vez que la filosofía recurre a métodos conceptuales de sillón [*armchair methods*], los cuales dan origen a las *filosofías de sillón*. Así, los experimentos mentales son un método *a priori*, que a diferencia de los experimentos reales, ponen la teoría en contacto con situaciones imaginarias decisivas para decidir su destino.

El artículo de Brown justamente destaca este aspecto. Específicamente, considera que los experimentos mentales son instancias cognitivas en que es posible generar conocimiento de forma *a priori*, sin el concurso de la experiencia. A pesar de que irían más allá de lo que Brown denomina conocimiento de base [*background knowledge*], y de que parecen existir con independencia de los hombres de carne y hueso, cuestión que explica por qué se pueden reproducir a través del tiempo, lo crucial de los experimentos mentales es que permiten *aprehender* cómo son las cosas, y así parecen incrementar el conocimiento. Esto, por lo demás, sucede incluso si los experimentos mentales no aportan ningún nuevo dato empírico.

El enfoque browniano, de corte platónico y falibilista [*fallibilist*], ha sido criticado por los filósofos empiristas, quienes afirman que no es posible adquirir *nuevo* conocimiento mediante experimentos mentales. Que los experimentos mentales transcurran “en la cabeza” y no en laboratorios explica por qué su grado de confiabilidad es menor al de los experimentos reales. En efecto, mientras los laboratorios ayudan a explicar y predecir las conexiones causales de fenómenos mediante ambientes altamente idealizados, el laboratorio de la mente solo puede recurrir a imaginaciones y, por lo tanto, tienen una característica cuasi cartesiana. Tal como Levine recalca (1993), las certezas a las que llevan los experimentos mentales son más bien lógicas y conceptuales.

Norton representa la antípoda al pensamiento de Brown, pues defiende la idea de que los experimentos mentales son formas veladas de argumentar: detrás de todas las situaciones imaginadas no existe

más que un argumento escondido. Los experimentos mentales serían, puesto en palabras simples, tan solo “argumentación pintoresca”. El nuevo conocimiento al que alude Brown, critica Norton, no es más que contenido proposicional introducido de contrabando en las premisas de un argumento, el que se ha puesto de manera ingeniosa. Tal rasgo explicaría por qué, pese a que hay que imaginarse situaciones bastante peculiares mediante los procedimientos descritos por los experimentos mentales, lo que prima siempre es el argumento propuesto. La posibilidad de que todo experimento mental pueda formalizarse como argumento, sea científico o filosófico, muestra, según Norton, por qué no son más que derivaciones lógicas de conclusiones.

¿No resultan las posiciones de Brown y Norton un tanto extremas? Existen otras posiciones intermedias. El artículo de Hacking presentado en esta compilación recoge una serie de posiciones contrarias a Norton, que describen las aproximaciones del propio Brown, de Nerssesian y de Gooding. Sus quejas en contra de considerar los experimentos mentales como mera argumentación ingeniosa parecen al menos dignas de consideración, al menos respecto del siguiente punto: todos los experimentos mentales versan sobre operaciones que deben *realizarse* por parte de agentes epistémicos. Tal como sucede con los experimentos reales, la operación permite obtener un resultado y derivar una conclusión. En vista de cómo opera un experimento mental, de cómo este lleva a *creer que p*, es difícil que alguien capte su sentido *solo* al examinar el argumento formal que lo esquematiza.

Hacking, por lo mismo, propone en su trabajo que los experimentos mentales aluden a procedimientos y a formas de imaginar situaciones; por lo tanto, refutan teorías a través de imaginiería, sea esta lo que sea. Las imaginaciones difícilmente pueden reconstruirse a través de los conceptos de la ciencia cognitiva tales como “narrativa” o “modelación mental” [*mental modelling*], y aunque esto fuera posible, no se avanza mucho. Con relación a esta precisión, es patente la influencia de Wittgenstein en Hacking: los experimentos mentales apelan a imágenes mentales; cómo reconstruirlas resulta poco

interesante desde el punto de vista de cómo y para qué se usan los experimentos mentales.

Estos, sean entendidos como argumentos o como procedimientos, son constantemente usados en ciencia, y hay disciplinas como la física o la astronomía donde la experimentación mental es parte de la metodología de la disciplina. El artículo de Irvine se concentra en el papel específico que desempeñan los experimentos mentales en el razonamiento científico: colaboran con la empresa científica mediante argumentos hipotéticos o contrafácticos, fundamentados en conocimiento previamente adquirido, lo que usualmente se denomina como “marco teórico”.

Las presuposiciones teóricas son lo que distingue a los experimentos mentales de otras formas de razonamiento, como los métodos deductivos, abductivos e inductivos. En este punto coinciden los experimentos físicos y los mentales, ya que ambos deben ser capaces de responder a preguntas relevantes en ámbitos teóricos y prácticos, que surgen en el seno de un contexto teórico bien definido y acotado. No obstante, los experimentos reales también inducen a errores, o a resultados incorrectos, ¿por qué no podría suceder lo mismo con los experimentos mentales, especialmente si se toma en consideración su mayor grado de falibilidad en relación con los experimentos reales?

Los artículos que tratan con este problema son precisamente los de Gale y Janis. El primero explora en qué forma los experimentos mentales pueden contribuir a la generación de contraejemplos, y cómo algunos de aquellos, pese a la creencia de quienes los utilizan entusiastamente, resultan ineficaces, viciados e incluso perjudiciales. Esto, según Gale, ocurre cuando los experimentos mentales se encuentran en franca contradicción con las prácticas habituales de nuestros juegos de lenguaje. Es decir, quien participe en la ejecución de un experimento mental debe hacerlo de acuerdo con las reglas de un juego de lenguaje, pues estas, además de permitir acceder a lo imaginado y a lo contrafáctico, son decisivas para evaluar la aplicación de un concepto *C*.

Se debe distinguir entre las presuposiciones empíricas para participar en un juego de lenguaje y las reglas del juego propiamente tales. Ignorar esta crucial diferencia lleva a situaciones ambiguas y a experimentos mentales viciados, con conclusiones erradas que contravendrían las presuposiciones empíricas de nuestros juegos de lenguaje tradicionales. Por ejemplo, es típico que algunos experimentos mentales entren en conflicto con la presuposición empírica de la continuidad espacio-temporal de las personas. En consecuencia, el ensayo de Gale destaca en qué medida hay constreñimientos que se deben tener en cuenta cuando se plantea un experimento mental. No respetarlos lleva a experimentos extravagantes, viciados y reñidos con las prácticas comunes de los juegos de lenguaje en los que participamos.

El artículo de Janis, por otra parte, también analiza de manera rigurosa en qué sentido pueden fallar los experimentos mentales. Alineándose con aquellos que insisten en que los experimentos mentales son descripciones de procedimientos experimentales que llevan a un resultado, Janis propone que hay al menos tres sentidos en los que un experimento mental puede fallar. En primer lugar, y tal como ocurre con los experimentos reales, un experimento mental falla cuando no es posible ejecutarlo hasta el final. A su vez, falla cuando el experimentador es llevado a una conclusión incorrecta, problema que también se da en los experimentos reales. Finalmente, y tal como un experimento real, fracasa cuando se obtiene un resultado irrelevante para responder la pregunta que lo motivó. Así, un experimento mental falla cuando su resultado lleva a una conclusión que no satisface el objetivo de su ejecución. De los tres fallos el último es el más interesante, pues se siembran dudas sobre la pertinencia del experimento mental para resolver disputas teóricas.

El último artículo de esta primera parte examina la relación entre lo analítico y lo concebible. Massey destaca el contexto y desarrollo histórico de la filosofía analítica en el siglo XX, diagnosticando cómo la filosofía se encontraba en un estado anoréxico y desanimado luego del

dictum quineano en los 50's sobre la analiticidad, un concepto filosófico que oscila entre lo circular y lo obscuro. Según Quine, la distinción analítico/sintético es artificial e infundada, pues no es posible precisar de manera clara y no circular qué es lo analítico. Pero dicha acusación, citada con gran prosapia por algunos filósofos, no ha desincentivado que se recibiese la analiticidad por la puerta trasera, disfrazada como concebibilidad [*conceivability*]. En este punto, el artículo de Massey no solo resulta esclarecedor respecto de la relación entre analiticidad y concebibilidad, sino que también representa un incentivo para quienes defienden la confiabilidad de los experimentos mentales, por ejemplo, en vista de las aproximaciones clásicas de filósofos como Descartes y Hume. En este sentido, si los científicos no escatiman emplear experimentos mentales, ¿por qué el filósofo puede disponer de este método *a priori* para dar lugar a filosofías de sillón?

Brindar al lector elementos para responder esta pregunta de manera positiva o negativa es, justamente, el objetivo de la primera parte de este libro. Los artículos que la componen son el fundamento para que el lector entienda el significado de la pregunta misma.

Referencias

- Kuhn, T. (1964): "A function for thought experiments". Reimpreso en: *The Essential Tension*. Chicago: University of Chicago Press.
- Levine, J. (1993): "On Leaving out What It's Like". En: M. Davies and G. Humphreys (eds.) *Consciousness*. Oxford: Blackwell. Reimpreso en: N. Block, O. Flanagan y G. Güzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
- Schildknecht, C. (1990): *Philosophische Masken: Literarische Formen der Philosophie bei Platon, Descartes, Wolff und Lichtenberg*. Stuttgart: Metzler.

Experimentos mentales¹

James Brown

Los experimentos mentales son dispositivos de la imaginación que se usan para investigar la naturaleza de las cosas. Basta hacer una lista de unos pocos de los experimentos mentales más conocidos para recordar su enorme influencia e importancia en las ciencias: el balde de Newton, el demonio de Maxwell, el ascensor de Einstein, el microscopio de rayos gamma de Heisenberg, el gato de Schrödinger. Se puede decir lo mismo de su importancia en filosofía. Gran parte de la ética, de la filosofía del lenguaje y de la filosofía de la mente está firmemente basada en los resultados de experimentos mentales. Nuevamente, una breve lista hace esto evidente: el violinista de Thompson, la habitación china de Searle, la tierra gemela de Putnam, las personas imaginadas por Parfit, quienes se dividen como amebas. El siglo XVII vio a algunos de sus partidarios [*practitioners*] más destacados en Galileo, Descartes, Newton y Leibniz. En nuestro tiempo, la creación de la mecánica cuántica y la relatividad son casi impensables sin el rol crucial de los experimentos mentales. La filosofía contemporánea, incluso más que las ciencias, se vería severamente empobrecida sin ellos.

1. Ejemplos de experimentos mentales

Entre los científicos, podría decirse que Galileo y Einstein fueron los experimentadores mentales más admirables, pero de ningún modo

¹“Thought experiments”. En: Edward N. Zalta (2008, ed.) *The Stanford Encyclopedia of Philosophy*. Disponible en línea en: <http://plato.stanford.edu/archives/win2008/entries/thought-experiment>. Traducido y reimpresso con permiso de su autor.

los primeros. Hubo experimentos mentales durante la Edad Media, y también pueden encontrarse en la Antigüedad. Uno de los más hermosos ejemplos de entre los antiguos (en Lucretius, *De Rerum Natura*) trata de mostrar que el espacio es infinito: si hay un límite del universo, podríamos arrojarle una lanza. Si la lanza vuela a través de él, no es un límite después de todo, y si la lanza rebota, debe haber algo más allá de este supuesto borde del espacio, un muro cósmico que detuvo la lanza, una pared que está ella misma en el espacio. De cualquier modo, no hay un borde del universo; el espacio es infinito.

Este ejemplo ilustra muy bien muchos de los rasgos comunes de los experimentos mentales: visualizamos una situación; llevamos a cabo un experimento; vemos qué sucede. También ilustra su falibilidad. En este caso hemos aprendido cómo conceptualizar el espacio de modo que sea tanto finito como ilimitado. Consideremos un círculo, que es un espacio unidimensional: cuando nos movemos alrededor de él, no hay borde, pero es no obstante finito. El universo podría ser una versión multidimensional.



Figura 1: Bienvenidos al borde del Universo

A menudo, un experimento real análogo a un experimento mental es imposible por razones físicas, tecnológicas o financieras; pero esto no tiene que ser una condición que defina los experimentos mentales. El punto más importante es que parece que podemos aprehender la naturaleza simplemente pensando, y en eso reside el enorme interés de los experimentos mentales para la filosofía. ¿Cómo es posible, aparentemente, aprender cosas nuevas sobre la naturaleza sin datos empíricos nuevos?

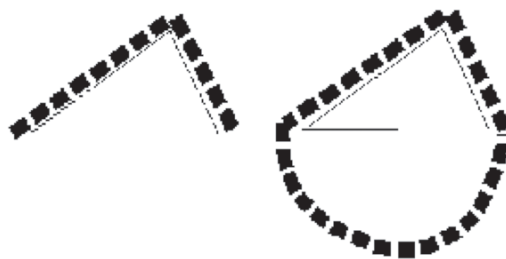


Figura 2a y b: ¿Cómo se moverá?

Ernst Mach hizo mucho para popularizar la idea del Gedankenexperiment. También popularizó el término, pero no fue el primero en usarlo. Ese honor parece corresponder a Georg Lichtenberg, quien escribió sobre ellos casi un siglo antes (Schildknecht 1990). Mach desarrolló una interesante postura empirista en su libro clásico, *La ciencia de la mecánica*. Poseemos, dice, una amplia provisión de “conocimiento instintivo” recogido de la experiencia. Parte de esta proviene de la experiencia efectiva, y otra parte la hemos heredado por medio del proceso evolutivo, gracias a la experiencia de nuestros ancestros. Este conocimiento no necesita ser articulado, pero se hace patente cuando enfrentamos ciertas situaciones. Uno de sus ejemplos favoritos se debe a Simon Stevin. Si se coloca una cadena sobre un plano doble sin fricción, como en la figura 2a, ¿cómo se moverá? Si se añaden algunos eslabones, como en la figura 2b, la respuesta se vuelve obvia. La configuración original debería estar en equilibrio estático. De otro modo, tendríamos una máquina de movimiento

perpetuo; y, según Mach, esto nos parece imposible de acuerdo a nuestro “conocimiento instintivo” basado en la experiencia.

Judith Thompson propuso uno de los experimentos mentales más sorprendentes y efectivos en el ámbito moral. Su ejemplo apunta a un popular argumento anti-aborto que es más o menos el siguiente: el feto es una persona inocente con derecho a la vida. El aborto resulta en la muerte de un feto. Por lo tanto, el aborto es moralmente errado. En su experimento mental se nos pide que imaginemos que un violinista famoso cae en coma. La sociedad de amantes de la música determina, a partir de registros médicos, que tú y sólo tú puedes salvar al violinista, conectándolo a ti por nueve meses. Los amantes de la música entran a tu casa mientras estás dormido y te conectan al violinista, quien permanece inconsciente (además, no sabe qué sucede, y por lo tanto es inocente). Podrías querer desconectarlo, pero te enfrentas con este argumento propuesto por los amantes de la música: el violinista es una persona inocente con derecho a la vida. Desconectarlo le causará la muerte. Por lo tanto, hacerlo es moralmente incorrecto.

Sin embargo, el argumento no parece convincente en este caso. Serías muy generoso si permaneces atado y en cama por nueve meses, pero no estás moralmente obligado a hacerlo. El paralelo con el caso del aborto es evidente. El experimento mental es efectivo al distinguir entre dos conceptos que habían estado previamente fusionados: el “derecho a la vida” y el “derecho a lo que se necesita para sostener la vida”. El feto y el violinista pueden tener ambos el primero, pero no es evidente que ninguno de los dos tenga el segundo. El resultado es que incluso si el feto tiene derecho a la vida (y esto es algo que Thompson no cree, pero que permite en aras de la discusión) podría ser permisible abortar. Teorizar sobre los experimentos mentales usualmente depende de los detalles o patrones de casos específicos. Muchas discusiones sobre experimentos mentales incluyen varias ilustraciones (e.g., Brown 1991, Horowitz y Massey 1991, y Sorensen 1992a). Hay dos libros dedicados principalmente a la presentación de ejemplos breves, no técnicos, Cohen 2005 y Tittle 2005.

2. Objeciones a los experimentos mentales

Por supuesto, se han refutado experimentos mentales particulares, pero, por lo general, la experimentación mental en las ciencias ha sido aceptada con gusto. El gran historiador de la física, Pierre Duhem, es casi el único que los condena. Él afirmó que un experimento mental no era sustituto para un experimento real, y que estos deberían ser prohibidos en ciencia. Sin embargo, en vista del importante rol de los experimentos mentales reales en la historia de la física —desde los cuerpos que caen de Galileo al balde de Newton, hasta el ascensor de Einstein— es improbable que mucha gente vaya a sentir demasiada simpatía por la crítica de Duhem, o que debiera sentirla.

Los filósofos son más críticos. Se preocupan, con alguna justicia, de cuán confiables son realmente nuestras intuiciones. ¿Podemos confiar en ellas en situaciones exóticas? Kathleen Wilkes, por ejemplo, sentía mucha desconfianza de las personas imaginadas por Parfit, quienes se dividen como amebas. Ella propuso que simplemente no sabemos qué decir al pensar una cosa como esa. Declaró que un experimento mental no debería violar lo que consideramos las leyes de la naturaleza. Esto haría desechar los ejemplos de Parfit. Una propuesta como esa me parece, sin embargo, demasiado fuerte. Aprendemos mucho sobre el mundo y nuestras teorías cuando nos preguntamos, por ejemplo, qué habría pasado después del *Big Bang*, si la ley de gravedad hubiese sido una ley de cubo inverso en vez de una de cuadrado inverso. ¿Habrían dejado de formarse estrellas? Razonar sobre tal posibilidad es perfectamente coherente y muy instructivo, incluso si viola una ley de la naturaleza.

Hay otras objeciones. Jonathan Dancy piensa que los experimentos mentales en ética son circulares. Daniel Dennett piensa que usamos conceptos populares (“folk”) que son inevitablemente conservadores. Esas objeciones pueden probablemente responderse, pero muestran un debate vigente.

3. Tipos de experimentos mentales

Hay muchas formas de clasificar los experimentos mentales: ciencia vs. filosofía, normativo (moral o epistémico) vs. fáctico, etc. Voy a esbozar una taxonomía basada en cómo cuentan [*function*] como evidencia. La división principal es *constructivo vs. destructivo*, esto es, entre casos en que un experimento mental puede usarse positivamente para fundamentar una teoría y casos en que podría usarse negativamente para socavar una teoría. Cada una de estas permite divisiones ulteriores.

Los experimentos mentales se usan negativamente de muchas maneras. La más simple es la de exponer una contradicción en una teoría, refutándola. Una segunda manera es la de mostrar que la teoría en cuestión está en conflicto con otras creencias que tenemos. El gato de Schrödinger, por ejemplo, no muestra que la teoría cuántica (como la interpretaba Bohr) es internamente inconsistente, sino que muestra un conflicto con algunas creencias de sentido común muy poderosas que tenemos acerca de objetos de tamaño macro como los gatos. La rareza de las superposiciones en el mundo atómico es bastante preocupante, nos dice Schrödinger, pero cuando implica la misma rareza a nivel cotidiano, es intolerable.

Hay un tercer tipo que, en efecto, mina una suposición central o premisa de un experimento mental. Thompson mostró que el “derecho a vivir” y el “derecho a lo que se necesita para sostener la vida” habían sido confundidos. Cuando se los distingue, el argumento contra el aborto colapsa. Un cuarto tipo de experimento mental negativo es algo más complejo. Voy a llamar a esos casos *contra-experimentos mentales*. Mach elaboró uno contra Newton, y Dennett uno contra Jackson. Newton presentó un par de experimentos mentales como evidencia para el espacio absoluto. Uno era el balde con agua que sube por su costado, y el otro era un par de esferas unidas por un cordel que mantenían su tensión en un espacio que, salvo por estos objetos, estaba vacío. La explicación de esos fenómenos, según Newton, es el espacio

absoluto: el balde y las esferas unidas rotan con respecto al espacio mismo. Como respuesta, Mach dijo, en contra de Newton, que las dos esferas se moverían una hacia la otra gracias a la tensión de la cuerda, y que si rotásemos un anillo muy grueso y pesado en torno a un balde estacionario, veríamos que el agua sube por un costado del balde. El contra experimento mental de Mach merma nuestra confianza en el de Newton. El espacio absoluto explicaba los fenómenos de los experimentos mentales de Newton, pero ya no estamos tan seguros sobre los fenómenos mismos (al menos, esta es la intención de Mach).

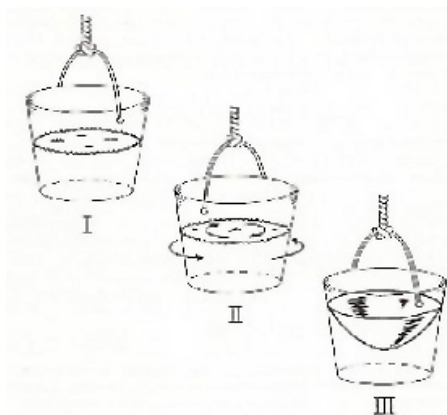


Figura 4: Las etapas del experimento del balde

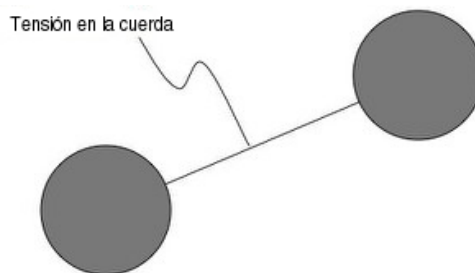


Figura 3: Dos esferas unidas por una cuerda en un espacio que de otra manera estaría vacío

Frank Jackson creó un experimento mental muy discutido que apuntaba a mostrar que el fisicismo, esto es, la doctrina de que todos los hechos son hechos físicos, es falso. En el experimento mental, Mary es una científico brillante que, desde su nacimiento, ha sido confinada en una pieza; allí solo tiene experiencias en blanco y negro. Dentro de la pieza, aprende todos los hechos físicos sobre la percepción. Un día, abandona el laboratorio y experimenta los colores por primera vez; aprende cómo es experimentar el rojo. Según Jackson, es claro que aprende algo nuevo. Puesto que sabía de antemano todos los hechos físicos, debe haber experimentado algo no-físico cuando experimentó el color. Por lo tanto, el fisicismo está necesariamente equivocado.

Dennett replicó a este experimento mental con uno propio. Comienza como el de Jackson, pero cuando Mary deja el laboratorio, ella dice: “Ah, la percepción del color es tal como pensaba que sería”. Tal como Mach, Dennett niega el fenómeno del experimento mental original. Y como el de Mach, su experimento mental resulta efectivo para neutralizar el de Jackson en tanto parece similarmente plausible.

Para ser efectivos, los contra-experimentos mentales no necesitan ser muy plausibles. En una corte, el jurado va a condenar si se establece la culpa “más allá de la duda razonable”. Una estrategia de defensa común consiste en presentar la evidencia de una forma alternativa que tenga suficiente plausibilidad como para dejar en duda el caso del abogado acusador. Eso es suficiente para socavarlo. Un buen contra-experimento mental tiene que hacer solamente eso para ser efectivo.

Los experimentos mentales también pueden ser constructivos. Hay muchas maneras en que un experimento mental puede dar fundamento eficaz a una teoría. Una de estas consiste en proporcionar un cierto tipo de ilustración que haga que las afirmaciones de la teoría sean claras y evidentes. En tales casos los experimentos mentales sirven como una suerte de ayuda heurística. Podría haber ya un resultado establecido, pero un experimento mental puede conducir una sensación de entendimiento muy satisfactoria. Newton presentó

un ejemplo maravilloso que muestra cómo la Luna mantiene su órbita de la misma manera en que un objeto cae a la Tierra. Lo ilustró mediante la imagen de un cañón que dispara una bala más y más lejos. En el límite, la Tierra se curva tan rápido como cae la bala, con el resultado eventual de que la bala terminará cayendo en el punto desde el que fue lanzada. Si nada se lo impide, dará vueltas una y otra vez. Eso es lo que hace la Luna. Podríamos llegar a la misma conclusión mediante cálculo, pero el experimento mental de Newton proporciona ese entendimiento aparente [*illusive*]. *Es un maravilloso ejemplo del “efecto ajá”*.

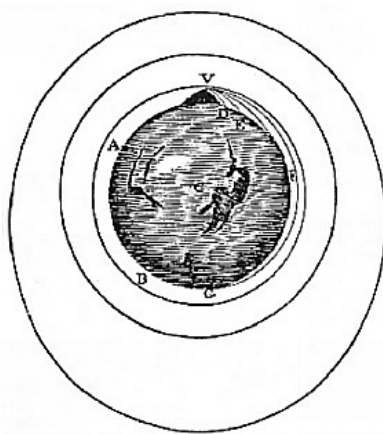


Figura 5: El disparo que se oye en todo el mundo

El ascensor de Einstein mostró que la luz se curva en un campo gravitatorio; el demonio de Maxwell que la entropía podía descender; el violinista de Thompson que el aborto podía ser moralmente permisible incluso si el feto tiene derecho a la vida; el balde de Newton que el espacio es una cosa independiente; las personas de Parfit que se dividen mostraban que la supervivencia es una noción más importante que la de identidad cuando se considera la naturaleza de la persona. Digo que “mostraron” tal y tal cosa, pero podría ser mejor decir “pretenden mostrar”, ya que algunos de esos experimentos son

bastante discutibles. Lo que tienen en común es que intentan concluir algo positivo. Al contrario de los experimentos mentales destructivos, no tratan de demoler una teoría existente, aunque eventualmente podrían hacerlo.

4. Algunas posturas recientes sobre los experimentos mentales

El ensayo de Thomas Kuhn “A function for thought experiments” emplea muchos de los conceptos (aunque no la terminología) de su bien conocido libro *La Estructura de las Revoluciones Científicas*. Según él, un experimento mental correctamente concebido puede conducir a una crisis o al menos crear una anomalía en la teoría dominante, y contribuir así a un cambio de paradigma. Los experimentos mentales nos pueden enseñar algo nuevo sobre el mundo, incluso si no tenemos nuevos datos empíricos, ayudándonos a re-conceptualizar el mundo de una mejor manera. Tamar Gendler ha desarrollado recientemente esta postura en varios aspectos importantes.

En los años recientes se ha visto un súbito aumento del interés en los experimentos mentales. Las posturas de Brown (1991) y Norton (1991, 1996) representan los extremos del racionalismo platónico y del empirismo clásico, respectivamente. Norton afirma que cualquier experimento mental es en realidad un argumento (posiblemente disfrazado); comienza con premisas fundadas en la experiencia y sigue reglas de inferencia deductivas o inductivas para llegar a su conclusión. Los rasgos pintorescos de un experimento mental, que le dan un sabor experimental, pueden ser psicológicamente útiles, pero en último término son redundantes. Así, sostiene Norton, nunca vamos más allá de las premisas empíricas de una manera que un empirista objetaría (para críticas, véase Bishop 1999; Brown 1991, 2004a, 2004b; Haggqvist 1996; Gendler 1998, 2004; Nersessian 1993; y Sorensen 1992a y 1992b; y para una defensa, véase Norton 1991, 1996, 2004a, y 2004b).

A diferencia de lo anterior, Brown sostiene que en algunos casos especiales bien podemos ir más allá de los datos antiguos para

obtener conocimiento *a priori* de la naturaleza (ver también Koyré 1968). Galileo mostró que todos los cuerpos caen a la misma velocidad con un experimento mental brillante que comenzaba destruyendo la teoría aristotélica, entonces dominante. Esta mantenía que los cuerpos pesados caían más rápido que los livianos ($H > L$). Pero considérese la figura número 6, en que una pesada bala de cañón (H) y una ligera bala de mosquete (L) son atadas para formar un objeto compuesto (H+L); el último debería caer más rápido que la bala de cañón sola. Pero el objeto compuesto debería caer más lentamente también, ya que la parte liviana actuará como freno sobre la parte más pesada. Ahora tenemos una contradicción ($H+L > H$ y $H > H+L$). Ese es el fin de la teoría de Aristóteles. Pero hay, además, un beneficio extra, ya que la explicación correcta se hace evidente: caen a la misma velocidad ($H = L = H+L$).

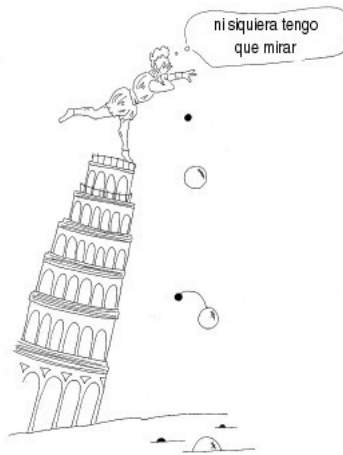


Figura 6: Ni siquiera tengo que mirar

Podría decirse que este es conocimiento *a priori* (aunque falible) de la naturaleza, ya que no hay información nueva involucrada, ni la conclusión se deriva de datos antiguos, ni es alguna forma de verdad lógica. Esta postura sobre los experimentos mentales puede desarrollarse más vinculando la epistemología de lo *a priori* con

teorías recientes sobre las leyes de la naturaleza que mantienen que las leyes son relaciones entre entidades abstractas existentes de manera objetiva. Es, por tanto, una postura bastante platónica, similar a las teorías platónicas sobre la matemática, como la propuesta por Gödel (para más detalles, ver Brown 1991).

Las dos posturas que acaban de esbozarse ocupan los extremos opuestos de un espectro de posiciones sobre los experimentos mentales, al menos dentro de la filosofía de la ciencia. Una lista de algunas de las posturas prometedoras incluiría la de Sorensen (un poco en el espíritu de Mach), quien sostiene que los experimentos mentales son un caso límite de los experimentos ordinarios; pueden alcanzar sus objetivos, dice, sin ser ejecutados (el libro de Sorensen es valioso también por su extensa discusión de experimentos mentales en una amplia gama de disciplinas). Otras posturas promisorias son las de Gooding (quien enfatiza la similar naturaleza procedural de los experimentos mentales y los experimentos reales), Miscevic y Nersessian (quienes vinculan los experimentos mentales a “modelos mentales”), y varias de las que hay en Horowitz y Massey 1991. Aparte de estas propuestas, una muestra de discusiones recientes excelentes incluye lo siguiente: Arthur 1999; Gendler 1998, 2000, 2002a, 2004; Haggqvist 1996; Humphreys 1994; McAllister 1996, 2004; y muchos más. La literatura sobre los experimentos mentales en la ciencia sigue creciendo rápidamente (véase la Bibliografía para más ejemplos).

Fuera de la filosofía de la ciencia, los filósofos siguen debatiendo los méritos de experimentos mentales particulares como los de Searle, Thompson, Jackson, etc. A un nivel más general, hay un debate sobre la utilidad de ejemplos tremendamente rebuscados. De todas maneras, ¿qué tan confiables son nuestras intuiciones en esos casos? El mismísimo tema de la intuición ha sido tópico de reciente discusión. Un grupo pequeño pero significativo de filósofos apoya su uso mientras que otros minimizan su confiabilidad y significatividad (véase DePaul y Ramsey 1998 para encontrar ejemplos de artículos sobre este tema.) La relación entre la concebibilidad [*concebibility*] y la posibilidad es otro

tema que ha estado en el aire recientemente, y que tiene mucho que ver con los experimentos mentales (véase Gendler y Hawthorne 2002). Se ha empezado a explorar la relación entre los experimentos mentales y la ficción literaria (véase Swirski 2007).

Referencias

Le debo agradecimiento a Tamar Gendler, en quien me he apoyado mucho al construir la siguiente bibliografía. Puede encontrarse mucho más en las bibliografías de Sorensen 1992a, Gendler 2000, y Kühne 2005.

- Arthur, R. (1999): “On thought experiments as a priori science”. *International Studies in the Philosophy of Science* 13/3, 215-229.
- Bishop, M. (1998): “An epistemological role for thought experiments”. En: N. Shanks (ed.) *Idealization IX: Idealization en Contemporary Physics*. Amsterdam: Rodopi.
- _____ (1999): “Why thought experiments are not arguments”. *Philosophy of Science* 66, 534-41.
- Bokulich, A. (2001): “Rethinking thought experiments”. *Perspectives on Science* 9/3, 285-307.
- Brendel, E. (2004): “Intuition pumps and the proper use of thought experiments”. *Dialectica* 58/1, 88-108.
- Brown, J. (1991): *Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.
- _____ (1993): “Why empiricism won’t work”. *Proceedings of the Philosophy of Science Association* 2, 271-279.
- _____ (2004a): “Why thought experiments transcend experience”. En: C. Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*. Malden, MA: Blackwell.
- _____ (2004b): “Peeking into Plato’s heaven”. *Philosophy of Science* vol. 71, 1126-1138.
- Bunzl, M. (1996): “The logic of thought experiments”. *Synthese* 106/2 (Fall), 227-240.
- Buzzoni, M. (2004): *Esperimento ed esperimento mentale*.

Milano: FrankoAngeli.

- Cargile, J. (1987): “Definitions and counterexamples”. *Philosophy* 62, 179-193.
- Cohen, M. (2005): *Wittgenstein’s Beetle and Other Classic Thought Experiments*. Oxford: Blackwell.
- Cohnitz, D. (2006): *Gedankenexperimente in der Philosophie*. Paderborn: Verlag GmbH.
- Cooper, R. (2005): “Thought experiments”. *Metaphilosophy* 36, 3, 328.
- Dancy, J. (1985): “The role of imaginary cases in ethics”. *Pacific Philosophical Quarterly* 66 (January-April), 141-153.
- Dennett, D. (1991): *Consciousness Explained*. New York: Little Brown.
- _____ (2005): *Sweet Dreams*. Cambridge, MA: MIT Press.
- DePaul, M. y Ramsey, W. (2002): *Rethinking Intuition: The Psychology of Intuition & Its Role in Philosophical Inquiry*. New York: Rowan and Littlefield.
- Duhem, P. (1954): *Aim and Structure of Physical Theory*. Princeton: Princeton University Press.
- Gendler, T. (1998): “Galileo and the indispensability of scientific thought experiment”. *The British Journal for the Philosophy of Science* 49/3 (Sept), 397-424.
- _____ (2000): *Thought Experiment: On the Powers and Limits of Imaginary Cases*. NY: Garland Press (ahora Routledge).
- _____ (2002a): “Personal Identity and Thought-Experiments”. *Philosophical Quarterly* 52/206, 34-54.
- _____ (2002b): “Thought experiment”. *Encyclopedia of Cognitive Science*. NY/London: Nature/Routledge.
- _____ (2004): “Thought experiments rethought —and re-perceived”. *Philosophy of Science* 71, 1152-1164.
- _____ (2005): “Thought experiments in science”. *Encyclopedia of Philosophy*. New York: MacMillan.
- Gendler, T. y Hawthorne, J. (2002): *Conceivability and*

Possibility. Oxford: Oxford University Press.

- Genz, H. (1999): *Gedankenexperimente*. Weinheim: Wiley-VCH.
- Gooding, D. (1993): "What is experimental about thought experiments?". En: D. Hull, M. Forbes, and K. Okruhlik (eds.) *PSA 1992*, vol. 2, East Lansing, MI: Philosophy of Science Association.
- _____ (1992): "The cognitive turn, or, why do thought experiments work?". En: R. Giere (ed.) *Cognitive Models of Science*. Minneapolis: University of Minnesota Press.
- _____ (1994): "Imaginary science". *British Journal for the Philosophy of Science* 45/4 (December), 1029-1045.
- Hacking, I. (1993): "Do thought experiments have a life of their own?". En: D. Hull, M. Forbes, and K. Okruhlik (eds.) *PSA 1992*, vol. 2, East Lansing, MI: Philosophy of Science Association. [Véase en este volume p. 36].
- Haggqvist, S. (1996): *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell International.
- Horowitz, T. y Massey, G. (1991): *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Humphries, P. (1994): "Seven theses on thought experiments". En: J. Earman *et al.*, (eds.) *Philosophical Problems of the Internal and External World*. Pittsburgh: University of Pittsburgh Press.
- Ierodiakonou, K. (2005): "Ancient thought experiments: A first approach". *Ancient Philosophy* 25, 125-140.
- Irvine, A. (1991): "Thought experiments in scientific reasoning". En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Jackson, F. (1982): "Epiphenomenal qualia". *Philosophical Quarterly* 32, 27-36.
- Jackson, M. (1992): "The *gedankenexperiment* method of ethics". *The Journal of Value Inquiry* 26, 525-535.

- Janis, A. (1991): “Can thought experiments fail?” En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- King, P. (1991): “Mediaeval thought-experiments: The metamethodology of mediaeval science”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Klassen, S. (2006): “The science thought experiment: How might it be used profitably in the classroom?” *Interchange* 37/1, 77-96.
- Koyré, A. (1968): *Metaphysics and Measurement*. London: Chapman and Hall.
- Kuhn, T. (1964): “A function for thought experiments”. Reimpreso en: *The Essential Tension*. Chicago: University of Chicago Press.
- Kühne, U. (2005): *Die Methode des Gedankenexperiments*. Frankfurt: Suhrkamp.
- Kujundzic, N. (1992): “How does the laboratory of the mind work?” *Dialogue* 32/3 (Summer), 573-578.
- Kujundzic, N. (1995): “Thought experiments: Architecture and economy of thought”. *The Journal of the British Society for Phenomenology* 26/1 (January), 86-93.
- Laymon, R. (1991): “Thought experiments of Stevin, Mach and Gouy: Thought experiments as ideal limits and as semantic domains”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Lennox, J. (1991): “Darwinian thought experiments: A function for just-so stories”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Lichtenberg, G. (1983): *Schriften und Briefe: Sudelbücher, Fragmente, Fabeln, Verse* (Erster Band). Frankfurt: Insel Verlag.

- Mach, E. (1960): *The Science of Mechanics*. LaSalle, IL: Open Court.
- _____ (1976): “On thought experiments”. En: *Knowledge and Error*. Dordrecht: Reidel.
- Massey, G. (1991): “Backdoor analyticity”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- McAllister, J. (1996): “The evidential significance of thought experiments in science”. *Studies in History and Philosophy of Science*, 27/2, 233-250.
- _____ (2004): “Thought experiments and the belief in phenomena”. *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science* 71, 1164-1175.
- _____ (2005): “The virtual laboratory: Thought experiments in seventeenth-century mechanics”. En: H. Schramm, L. Schwarte y J. Lazardzig (eds.) *Collection, Laboratory, Theater: Scenes of Knowledge in the 17th Century*. New York: Walter de Gruyter.
- Miscevic, N. (1992): “Mental models and thought experiments”. *International Studies in the Philosophy of Science* 6/3, 215-226.
- _____ (1997): “Categorial and essentialist intuitions: A naturalist perspective”. *Acta Analytica* 12/19, 21-39.
- Nersessian, N. (1992): “How do scientists think? Capturing the dynamics of conceptual change in science”. En: R. Giere (ed.) *Cognitive Models of Science*. Minneapolis: University of Minnesota Press.
- _____ (1993): “In the theoretician’s laboratory: Thought experimenting as mental modeling”. En: D. Hull, M. Forbes, and K. Okruhlik (eds.) *PSA 1992*, vol. 2, East Lansing, MI: Philosophy of Science Association.
- Norton, J. (1991): “Thought experiments in Einstein’s work”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in*

- Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- _____ (1996): “Are thought experiments just what you always thought?” *Canadian Journal of Philosophy* 26, 333-366.
 - Norton, J. (2004a): “On thought experiments: Is there more to the argument?” *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science* 71, 1139-1151. [Véase en este volumen p. 22].
 - _____ (2004b): “Why thought experiments do not transcend empiricism”. En: C. Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell.
 - Parfit, D. (1984/1987): *Reasons and Persons*. Oxford: Clarendon Press.
 - Rescher, N. (2005): *What If? Thought Experimentation in Philosophy*. New Brunswick, NJ: Transaction Publishers.
 - Schildknecht, C. (1990): *Philosophische Masken: Literarische Formen der Philosophie bei Platon, Descartes, Wolff und Lichtenberg*. Stuttgart: Metzler.
 - Sorensen, R. (1992a): *Thought Experiments*. Oxford: Oxford University Press.
 - _____ (1992b): “Thought experiments and the epistemology of laws”. *Canadian Journal of Philosophy* 22/1 (March), 15-44.
 - Stinner, A. (1990): “Philosophy, thought experiments, and large context problems in the secondary physics course”. *International Journal of Science Education* 12/3, 157-244.
 - Swirski, P. (2007): *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution and Game Theory*. London & New York: Routledge.
 - Thompson, J. (1971): “A defense of abortion”. *Philosophy and Public Affairs* 1/1 (Fall), 47-66.
 - Tittle, P. (2005): *What If... Collected Thought Experiments in Philosophy*. New York: Pearson Longman.
 - Urbaniec, J. (1988): “In search of a philosophical experiment”.

Metaphilosophy 19 (July-August), 294-306.

- Wilkes, K. (1988): *Real People: Personal Identity without Thought Experiments*. Oxford: Clarendon Press.
- Witt-Hansen, J. (1976): "H.C. Orsted, Immanuel Kant, and the thought experiment". *Danish Yearbook of Philosophy* 13, 48-65.
- Yablo, S. (1993): "Is conceivability a guide to possibility?" *Philosophy and Phenomenological Research* 53/1 (March), 1-42.

Sobre los Experimentos Mentales: ¿Hay algo que añadir al argumento?²

John D. Norton

Los experimentos mentales en ciencia son simplemente argumentación pintoresca. Fundamento esta postura de varias formas, incluyendo la afirmación que se sigue del hecho de que los experimentos mentales pueden errar, pero aún pueden usarse de manera confiable. Esta postura se defiende contra alternativas propuestas por mis colegas expositores [*co-symposiasts*].

1. Introducción

Un científico —un Galileo, Newton, Darwin o Einstein— nos presenta un problema complicado. Quedamos perplejos. En pocas palabras de prosa simple, imagina un experimento, puramente mental. Seguimos, replicando en nuestra mente los cuerpos del experimento que caen o sus baldes que dan vueltas, y nuestra incertidumbre se evapora. Sabemos la resolución, y de algún modo intuimos que la supimos todo el tiempo. Ese momento de darse cuenta es exquisito; y es difícil resistir la sensación de que algo de profundo momento epistémico acaba de acontecer.

² “Thought experiments: Is there more to the argument?” En: Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association. *Philosophy of Science* 71, 1139-1151. Traducido desde la versión disponible en línea en <http://philsci-archive.pitt.edu/archive/00000719/>, con permiso del autor.

Mi propósito es dar fuerza a aquellos que quieran resistir. Mi forma de ver los experimentos mentales es bastante deflacionaria. Afirmo que son simplemente argumentación ordinaria, vestida de una vívida forma narrativa o pictórica. Como resultado, epistémicamente no pueden hacer más que la argumentación ordinaria. No dudo que esta vestidura pictórica les dé poderes retóricos especiales, pero esos no son mi preocupación. De manera más precisa, mi preocupación es lo que denomino:

El problema epistemológico de los experimentos mentales en la ciencia. Se supone que los experimentos mentales nos dan conocimiento del mundo natural. ¿De dónde proviene este conocimiento?

Debido a que sostengo que los experimentos mentales son simplemente argumentos pintorescos, mi solución al problema es que este conocimiento viene de las premisas introducidas explícita o tácitamente en el experimento mental. Tal conocimiento es transformado, usualmente de forma tácita, por medio de argumentación deductiva o inductiva para entregar el resultado final.

En la sección 3, voy a revisar el argumento que puede plantearse para esta visión deflacionaria de los experimentos mentales. Voy a elaborar un desarrollo reciente del argumento que depende de la noción de que los experimentos mentales pueden errar, pero incluso así ser usados confiablemente. En la forma de la “tesis de la confiabilidad”, voy a proponer que esto es posible solo si la experimentación mental es gobernada por una lógica muy generalizada; y consideraciones evolutivas sugieren que esas son las lógicas que nos son familiares en la literatura sobre inducción y deducción. Como preámbulo de la discusión sobre la confiabilidad, en la sección 2 ilustraré con un ejemplo de cómo los experimentos mentales pueden errar.

En la sección 4 compararé mi postura con la de mis colegas expositores. Tanto James Brown como Tamar Szabó Gendler proponen que los experimentos mentales tienen más poderes

epistémicos que la mera argumentación —dicen que hay algo más que añadir al argumento. En las secciones 4.1-4.3 voy a discutir que esos poderes extra, si existen, serían epistémicamente irrelevantes, ya que no pueden ser usados confiablemente. James McAllister propone que se adjudican poderes epistémicos a los experimentos mentales cuando se ven satisfechas ciertas presunciones estrechas, históricamente contingentes. En la sección 4.4 explicaré que el puro hecho de que una ciencia busque leyes es probablemente suficiente, según mi postura, para que permita la experimentación mental; y eso a su vez nos asegura la admisibilidad casi universal de los experimentos mentales en ciencia.

2. Un experimento mental erróneo

Aunque hay consenso general de que los experimentos mentales pueden entregar resultados útiles en ciencia, no son infalibles. Los experimentos mentales pueden dar resultados falsos, y a menudo lo hacen. Esto puede verse muy claramente en la existencia de pares de experimentos mentales que dan resultados contradictorios. He descrito algunos de esos en mi (Norton, próxima publicación), donde los llamo pares experimento—anti-experimento mental. Podemos tener experimentos mentales que muestren que el mundo es finito o infinito; que hay un espacio absoluto, o no; que la geometría de un disco que rota es Euclidiana o no; y que un rotor infinito pero en reposo proporciona elevación o no. Aquí daré otro ejemplo de error. Es una variante de un experimento mental bien conocido discutido en Brown (1991, pp. 38-40) que trata de la contracción de Lorentz en la longitud de los cuerpos en movimiento en la relatividad especial³.

Imagínese una lámina lisa de metal de la que cortásemos cuidadosamente una barra larga que calce precisamente en la ranura creada, como en la figura 1.

³ La diferencia crucial de la versión de Brown es que todos los movimientos en el mío son inerciales, así que no hace falta considerar la aceleración y la rigidez.

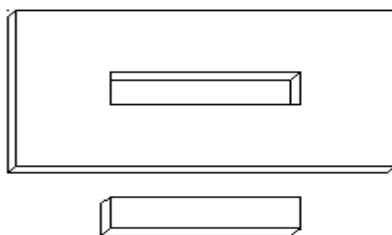


Figura 1: Una barra cortada de una lámina deja una ranura del mismo tamaño

Transportamos la lámina con la ranura y la barra a una región del espacio alejada de la gravitación, donde flotan en reposo. Ponemos la barra sobre la lámina y a un lado. Ponemos entonces la barra en movimiento uniforme hacia la ranura de manera tal que la barra permanezca alineada con la ranura, y paralela a ella (véase la figura 2). Viendo el proceso desde la lámina y la ranura, nos preguntamos si la barra aún calzará en la ranura. De acuerdo con la relatividad especial, la barra se encoge en la dirección de su movimiento. Debido a que parte del movimiento se orienta con la ranura, la barra se volverá más corta que la ranura y pasará fácilmente por ella, como se muestra en la figura 2. En un punto estará enteramente encerrada por la ranura.

Ahora vemos exactamente el mismo proceso desde la barra. De acuerdo con este punto de vista, la barra permanece en reposo, pero la ranura se mueve uniformemente hacia aquella. Dejando de lado el hecho de que la barra y la ranura cambian de lugar en el experimento, los dos casos son exactamente el mismo. En reposo, la barra y la ranura son objetos de idéntica forma y tamaño; todo lo que cambiamos es que movemos la ranura en vez de la barra. Así que esta vez la ranura se ha contraído y es más corta que la barra. Por lo tanto, la barra no puede pasar por la ranura.

Podemos ver que no puede pasar a partir de la simetría de las dos vistas. En la vista de la ranura, llega un momento en que la barra

contraída está completamente encerrada en la ranura, que no se ha contraído. Por simetría, en la vista de la barra debe llegar un momento en que la ranura contraída se encerraría completamente en la barra que no se ha contraído —o lo haría si el material de la lámina que rodea la ranura pudiera penetrar el material de la barra. La inevitable colisión ocurre cuando el material de la lámina intenta penetrar el de la barra.

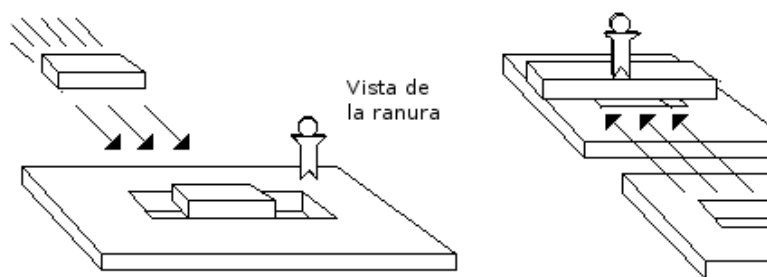


Figura 2: Barra y ranura en movimiento uniforme y oblicuo

Claramente algo ha salido muy mal, pues son dos vistas del mismo proceso. O bien la barra pasa por la ranura o no. No puede ser que resulten ambas. Hay una respuesta común y natural a experimentos mentales como este en la relatividad especial. Hay solo una dependencia de la longitud en el movimiento que no producirá problemas en un experimento mental de barra-ranura y en otros parecidos. Esta es que, después de todo, la longitud de una barra no altera su movimiento. En ese único caso, si la barra en reposo calza exactamente en la ranura cuando ambas están en reposo, siempre calzarán, ya sea que veamos una barra que se mueve acercándose a una ranura en reposo o a una ranura que se mueve acercándose a una barra en reposo. Ese es nuestro resultado: la contracción relativista no es real después de todo. Las barras y ranuras que se mueven realmente no se encogen.

Como sea que uno lo enfrente, este experimento presenta un error. En efecto, Einstein comenzó a luchar con este intento de eliminar la contracción de Lorentz en una fecha tan temprana como 1911 (Einstein 1911). ¿Cómo podríamos caracterizar el error? Podríamos pensar que la vista de la barra y la vista de la ranura son dos experimentos mentales relacionados. Debido a que están en desacuerdo sobre el resultado, al menos uno debe estar equivocado. Son un par experimento—anti-experimento mental, ya que entregan resultados contradictorios. O, si uno cree que la relatividad especial estándar proporciona una explicación consistente de los cuerpos en movimiento, el resultado contradictorio del experimento mental combinado no puede estar en lo correcto.

3. ¿Por qué argumentos?

3.1 Dos partes de la tesis justificadas

Creo que los experimentos en ciencia son simplemente argumentación pintoresca. En otros lugares (Norton 1991, 1996, y próxima publicación) he tratado de plantear mi argumento dividiendo esta afirmación en dos partes. Brevemente:

Primero, propongo que los experimentos mentales en ciencia siempre pueden reconstruirse como argumentos basados en suposiciones explícitas o tácitas que entregan el mismo resultado. Un fundamento para lo anterior es el empirismo. Si los experimentos mentales nos van a enseñar algo sobre el mundo, el empirismo nos dice que sólo pueden hacerlo disponiendo de nuestra experiencia del mundo. El resultado de un experimento mental debe ser la reformulación de esa experiencia por medio de un proceso que preserve la verdad o su probabilidad, esto es, mediante argumentación deductiva o inductiva. Un segundo fundamento, independiente, es que no he encontrado ningún experimento mental en ciencia que no pueda reconstruirse como un argumento. Los ejemplos incluyen los que la literatura ofrece como resistentes a tal reconstrucción (véase Norton 1996).

Segundo, propongo que llevar realmente a cabo un experimento mental consiste en la puesta en marcha de un argumento. Por supuesto es muy tentador no recoger esta segunda parte de la tesis y mantener alguna posición híbrida, como que los experimentos mentales explotan un misterioso poder mental cuyos productos pueden validarse separadamente por argumentación. Esa posición más débil me parece timorata. Equivale a no aceptar el significado real de un hecho notable: el alcance epistémico de un experimento mental coincide exactamente con el de un argumento. Imaginemos un caso análogo. Consultamos un oráculo, pero descubrimos que las predicciones verdaderas de este coinciden con lo que podría aprenderse de la experiencia ordinaria. El oráculo vaticina: “Cuidado. ¡Viene el fuego!” Pero tras los recientes temblores y flujos de lava, ¿no esperaban ya todos que el volcán estuviera a punto de hacer erupción?! La triste verdad es que este oráculo no tiene poderes especiales, sino que simplemente infiere predicciones a partir de la experiencia ordinaria, sea consciente o inconscientemente. Postulo el mismo veredicto para los experimentos mentales. Los experimentos mentales en ciencia tienen el mismo alcance epistémico que los argumentos simplemente porque son argumentos.

3.2 La tesis de la confiabilidad

En Norton (próxima publicación) busco dar sustento a mi postura de que los experimentos mentales son argumentos proponiendo:

La tesis de la confiabilidad. Si los experimentos mentales pueden ser usados de manera confiable epistémicamente, deben ser argumentos (en un sentido muy amplio) que justifiquen sus resultados, o deben poder reconstruirse como argumentos así.

Tal como explicaré más abajo, la tesis recoge una noción de argumentación mucho más amplia que la usualmente citada en los textos de lógica.

Brevemente, la tesis procede del hecho de que los experimentos mentales pueden entregar resultados erróneos. No obstante, pensamos que podemos tener confianza en el producto de un experimento mental si está adecuadamente formulado. Pero, ¿cómo vamos a saber si el experimento mental está formulado de manera adecuada? Debe haber algún indicio de verdad. Este indicio no es algo externo al experimento mental. Podría ser reconfortante saber que, por ejemplo, Einstein fue el autor del experimento. Pero en principio debería sostenerse por sí mismo. Deberíamos ser capaces de saber si el experimento mental está formulado adecuadamente simplemente al leer su texto. Así, el indicio debe ser algo que podamos reconocer en el experimento mental mismo. No puede ser que el experimento pueda simplemente encontrarse en una lista aprobada; eso sería un indicio externo. Debe haber alguna sistematización—algún rasgo común, alguna forma identificable, alguna estructura común—en los experimentos mentales admisibles cuya presencia podamos percibir. Esa forma o estructura no necesita definir todos los detalles del experimento mental. En efecto, no podría. El experimento del balde de Newton claramente no se ve afectado por la elección de un balde de madera o cuero. De modo que un buen experimento mental tiene dos partes: una estructura o forma que asegura su corrección y una porción que podemos alterar sin afectar su estructura.

Esto es, un buen experimento mental se construye a partir de una plantilla [*template*] en la que estamos libres de insertar material particular de nuestra elección. Encontramos un ejemplo familiar del uso de esas plantillas en lógica: los esquemas de una lógica son las plantillas y las sentencias o términos que se insertan en ellas son el contenido variable. Así, diré que una lógica generalizada es lo que gobierna cualquier exposición que use plantillas con contenido variable. El rango de las plantillas concebibles es enorme. Aquellas que se emplean en los experimentos mentales, sin embargo, deben ser más limitadas. Deben ser de tal manera que manifiestamente den lugar a las aspiraciones epistémicas del experimento mental. Los

esquemas lógicos satisfacen este requisito si preservan la verdad o su probabilidad. Quizás también deberíamos esperar que las plantillas de los experimentos mentales sean lo suficientemente simples como para permitir que su uso sea manejable [*tractable*].

¿Cómo sería la lógica generalizada de un buen experimento mental, si no es una lógica familiar? Una posibilidad viene del uso de la ciencia cognitiva para analizar los experimentos mentales (ver Nersessian 1992 y Palmieri próxima publicación) Uno describe los experimentos mentales como la manipulación de un modelo mental. Al centro de los modelos mentales hay lo que equivale a las plantillas. Por ejemplo, lo siguiente sería una plantilla que codifica el arreglo espacial de cuatro cuerpos A, B, C y D en un cuadrado:

<i>A</i>	<i>B</i>
<i>C</i>	<i>D</i>

Los resultados de la actividad cognitiva que emplea esta plantilla son permitidos mientras esta plantilla codifique correctamente las propiedades del espacio.

Como se indica en Norton (próxima publicación), no pienso que la modelación mental [*mental modeling*] pueda ajustarse a todos los experimentos mentales. Algunos, muy notablemente en las ciencias físicas, dependen de derivaciones explícitas de resultados matemáticos dentro de una teoría física. Aquellas derivaciones son, inequívocamente, argumentos. Por lo demás, aún no he visto un experimento mental en ciencia (en oposición a otras instancias cognitivas) que no pueda reconstruirse como un argumento. Así que sigo creyendo que los experimentos mentales son argumentos que explotan las lógicas deductivas e inductivas familiares.

En parte, mi confianza en las últimas deriva de una confianza en el ingenio de los lógicos tanto deductivos como inductivos, cuya profesión tiene como objeto extraer los esquemas usados en la argumentación exitosa. Dado que la actividad de la experimentación

mental ha sido prominente e importante por largo tiempo, mi postura es simplemente un voto de confianza en que los lógicos han tenido éxito al seleccionar las lógicas que emplean. Así que no mantengo que los experimentos mentales son argumentos a causa de una creencia en la auto-disciplina de los experimentadores mentales para restringirse a un canon estrecho de la lógica. Más bien confío en que la lógica evoluciona para recoger cualquier forma argumentativa buena, nueva, que pueda emerger de los esfuerzos creativos de los experimentadores mentales.

4. Las epistemologías de los experimentos mentales comparadas

Mis colegas expositores ofrecen diferentes epistemologías de los experimentos mentales. Dos son más optimistas que la mía sobre los poderes epistémicos de los experimentos mentales; y una tercera es más pesimista. Antes de dirigirme a la tercera, usaré el experimento mental de la barra y la ranura como un caso probatorio [*a test case*] de por qué no comparto el optimismo de las primeras dos. Espero que una epistemología viable pueda dar cuenta de por qué este experimento mental yerra y cómo los experimentos mentales pueden ser usados confiablemente aún si tales errores son posibles. Mostraré que mi postura argumental puede hacerlo, mientras las más optimistas no pueden.

4.1 La postura argumental

El error del experimento mental de la barra y la ranura se diagnostica y corrige rápidamente al examinarlo como un argumento. Simplemente, hallamos que esconde una suposición falsa. Una vez que esa suposición es expuesta y corregida, la contradicción con la relatividad especial se evapora. La suposición falsa se manifiesta de manera muy vívida en la figura 2. Presumíamos que la barra y la ranura estaban paralelas cuando se las veía desde la vista de la ranura. El análisis asume tácita pero erróneamente que permanecen paralelas cuando adoptamos la vista de la barra. En la nueva vista,

no permanecen paralelas. El cambio de vista es llevado a cabo por una transformación de Lorentz, y resulta que esta transformación rota cuerpos que están orientados de forma oblicua hacia la dirección del movimiento, tales como la barra y la ranura. En la vista de la barra corregida, como se muestra en la figura 3, la barra y ranura rotadas ya no son paralelas y las rotaciones permiten que la ranura pase la barra, incluso cuando la barra es más larga que la ranura. Con la corrección, ambas vistas dan el mismo resultado; la barra pasa (véase el Apéndice para una descripción simple de cómo la transformación de Lorentz lleva a la rotación de cuerpos oblicuos).

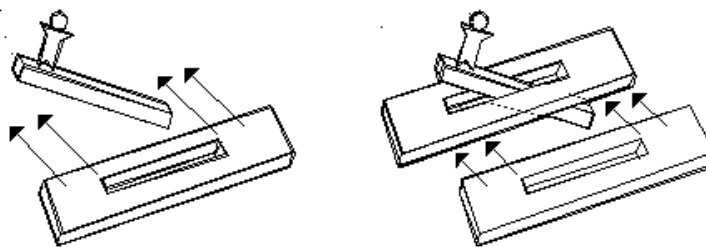


Figura 3: Vista de la barra corregida

El argumento del experimento mental original puede resumirse:

1. Una barra que se mueve hacia la ranura como en la figura 2 (vista de la ranura) se contrae y pasa (suposición de la relatividad especial).
2. El proceso que se ve desde la barra es el mismo, pero con la barra y la ranura cambiados de lugar (suposición de la simetría).
3. Bajo la simetría de 2, si la barra pasa en la vista de la ranura, no puede pasar en la vista de la barra (establecido en el texto del experimento mental).
4. La ranura no pasa la barra (de 1, 2 y 3).

5. Contradicción. La barra pasa (de 1) y no pasa (de 4).

Un segundo argumento revela lo que hay que desechar para eliminar la contradicción.

6. Hay una cinemática modificada que da resultados consistentes para ambas vistas (suposición).
7. El único caso que da resultados consistentes para ambas vistas es aquel en que no hay dependencia de la longitud en el movimiento (suposición).
8. No hay contracción de la longitud con el movimiento (de 6 y 7).

Lo que vemos ahora es que no hay simetría entre ambas vistas. Si la barra y la ranura son paralelas en una vista, no lo serán en la otra. Y esta diferencia es esencial para que ambas vistas den el mismo resultado. Esto es, para eliminar la contradicción, deberíamos haber desechado la premisa de simetría 2, no la contracción de Lorentz. El argumento falla a causa de la premisa falsa 2.

4.2 El platonismo de Brown

En otros lugares (Norton 1993, 1996, y próxima publicación) he descrito tanto mi admiración por la audacia de la teoría platónica de James Brown (1991, 1992, 1993, y próxima publicación) de ciertos experimentos mentales como mis resquemores acerca de la misma. Los últimos se basan en mi escepticismo sobre la realidad de un mundo platónico de leyes a las que se podría acceder por medio de experimentos mentales, o por otros medios. Ciertamente, no creo que una epistemología viable de los experimentos mentales la requiera. El problema más inmediato es cómo los experimentos mentales pueden usarse confiablemente si se supone que dan vistazos a este mundo platónico.

Brown sostiene que sólo algunos experimentos mentales especialmente favorecidos nos dan acceso directo al mundo platónico.

Esos son los que llama “experimentos mentales platónicos” (1991, pp. 43-45). Se distinguen por ser simultáneamente destructivos y constructivos; destruyen una postura y simultáneamente fundamentan otra. El experimento mental de la barra y la ranura es un caso de estos. Destruye la cinemática de la relatividad especial, en tanto muestra que la contracción de Lorentz no es sostenible. Al mismo tiempo, establece su reemplazo, que no hay contracción de los cuerpos que dependa del movimiento. Aun así, yerra. Sin que lo sepamos, nuestra visión platónica estaba obnubilada. ¿Cómo se supone que sepamos cuáles son los buenos experimentos mentales platónicos?

La respuesta de Brown (1991, pp. 65-66; 1993) consiste en sugerir que la visión ordinaria puede errar también sin que queramos descartar toda la experiencia visual como ilusión, y que podemos mantener alguna confianza en la experiencia visual incluso teniendo un entendimiento rudimentario de su mecanismo. Mi respuesta (Norton próxima publicación) es que no tenemos ni siquiera la más rudimentaria explicación sobre la naturaleza de la percepción platónica⁴. Podríamos tratar de corregir sus productos notando que el resultado del experimento mental de la barra y la ranura contradice la relatividad especial, pero, ¿cómo podemos estar tan seguros de que este no sea el experimento mental que selle el destino de la relatividad especial después de todo? Mi respuesta es simple. Reconocemos que los experimentos mentales son meramente argumentos, y este en particular alberga una suposición falsa. Brown podría querer apuntar a la misma suposición falsa para explicar el fallo de la visión platónica, pero al hacer esto simplemente replica el análisis basado en argumentos que acaba de darse. Nada en el diagnóstico depende de que los experimentos mentales sean nada más que argumentos.

⁴ De hecho, hasta donde sé nada impide que la argumentación misma sea el vehículo para acceder al mundo platónico. Si ese fuera el vehículo, no habría manera de distinguir en principio el enfoque de Brown y el mío. En ese caso afirmaría que el mío es preferible debido a la simplicidad.

4.3 El constructivismo de Szabó Gendler

Tamar Szabó Gendler (1998) ha desarrollado una teoría de los experimentos mentales que viene de Kuhn (1964). La noción principal es que la función de un experimento mental es revelar lagunas en los sistemas conceptuales e indicar cómo tienen que alterarse. Estoy muy de acuerdo con su análisis. Esa función de los experimentos mentales es interesante e importante. Donde diferimos, sin embargo, es en su insistencia en que servir esta función muestra que los experimentos mentales no pueden ser argumentos. Soy mucho más optimista sobre el poder de la argumentación para reconfigurar los esquemas conceptuales. El argumento de la paradoja de Russell condujo al abandono de la concepción ingenua de los conjuntos. Las pruebas de consistencia relativa del siglo XIX ubicaron a la geometría euclidiana y no-euclidiana al mismo nivel en lo que se refiere a la consistencia. Eso erradicó la milenaria creencia en que hay algo lógicamente defectuoso en una geometría que viole los postulados de Euclides. La prueba de la incompletitud [*incompleteness*] de la aritmética de Gödel nos mostró que un concepto de verdad como derivabilidad es inadecuado. Esto no es un mal desempeño para la argumentación pura.

Szabó Gendler parece permitir que la mera argumentación pueda revelar una contradicción. Pero ella sugiere que no puede decirnos qué ajustar en nuestro esquema conceptual a fin de eliminarla. Con esto estoy nuevamente de acuerdo. La decisión de qué es lo refutado por una *reductio ad absurdum* parece en gran parte depender de la manera en que se arma el argumento, y en las predilecciones e intereses mayores del científico. Sin embargo, aún no veo que los experimentos mentales puedan funcionar mejor en esta tarea. El contexto del experimento mental podría sugerir poderosamente el objetivo de la *reductio*, y podría estar completamente equivocado. Ese es el caso, precisamente, con el experimento mental de la barra y la ranura. Llegamos a una contradicción, y entonces sabemos que nos equivocamos en alguna parte. La manera en que se arma el experimento mental casi irresistiblemente nos conduce al objetivo

equivocado, la contracción de Lorentz. El verdadero culpable, la suposición de simetría, escapa porque es tan natural, aun cuando se basa en suposiciones bien ocultas sobre la simultaneidad absoluta, que son inadmisibles en la relatividad especial.

Si los experimentos mentales tienen algún poder extra, de revelar al culpable correcto cuando emerge una contradicción, ¿cuál es la base epistémica de este poder? Szabó Gendler ha mencionado la idea de Mach del conocimiento tácito asimilado de la experiencia, y habla de una “participación constructiva” de parte del lector (pp. 414-415). ¿Cómo vamos a saber si el poder de la participación constructiva se ha ejercido bien? ¿Cómo vamos a distinguirlo de los casos como el de la barra y la ranura, en que el experimento mental prueba ser enormemente engañoso en este sentido? Brevemente, si los experimentos mentales tienen esos poderes extra: ¿cómo pueden ejercerse confiablemente?

4.4 McAllister y la Inercia Evidencial de los Experimentos Mentales en Ciencia

Mientras Brown y Szabó Gendler son considerablemente más optimistas que yo sobre el poder epistémico de los experimentos mentales, mi colega expositor James McAllister (1996) ha desarrollado un pesimismo innecesario. Señala correctamente que los experimentos mentales solo pueden tener poder persuasivo si juegan según las reglas epistémicas correctas. La lectura mágica de entrañas animales, por ejemplo, no juega según esas reglas y por tanto no tiene poder epistémico en ciencia. Sin embargo, piensa que la licencia epistémica de los experimentos mentales es estrecha y dependiente del factor tiempo. El ejemplo que usa habitualmente es una “doctrina galileana de los fenómenos” que licencia la experimentación mental sobre fenómenos libres de accidentes cuyo comportamiento se ha idealizado como lo suficientemente regular para estar sujeto a leyes. El ejemplo familiar es el de los cuerpos cayendo libremente sin las complicaciones accidentales de la resistencia del aire. Los aristotélicos

se ocupan de las ocurrencias reales con todos sus accidentes, y por tanto, según la postura de McAllister, dejan de lado los experimentos mentales sobre cuerpos que caen.

Creo que todo lo que McAllister puede mostrar es que los aristotélicos dejan de lado experimentos mentales *particulares* que podrían depender de una noción que encuentran ilícita. No muestra que desechen la experimentación mental en general. Y no puede. Aparecen experimentos mentales en el *corpus* de Aristóteles. Algunos son del movimiento cósmico e incluso tratan con la caída de los cuerpos⁵. Aristóteles (*De Caelo* Libro II, Cap. 14, 297a13-30) se imagina la tierra en formación y presenta cómo el movimiento de los cuerpos hacia el centro daría como resultado una Tierra de forma esférica. En efecto, en un experimento mental de ambición sobrecogedora, Aristóteles continúa (297a33-297b13), e imagina que un peso varias veces el de la Tierra se añade a una mitad de la esfera de la tierra, trazando luego los movimientos resultantes.

El acento de McAllister en la doctrina galileana de los fenómenos como el fundamento del experimento mental es demasiado acotado⁶. Lo que es peor es que su insistencia (1996, p. 242) en que el poder de evidencia de los experimentos mentales es conferido “por el esfuerzo persuasivo de los científicos” confunde las cuestiones mezclando

⁵ Le agradezco a James Lennox haberme sugerido dos experimentos mentales pertinentes a la locomoción de los animales en la biología de Aristóteles. En: *Animales en Progreso* 8, 708b4-10 y 709a1-8.

⁶ McAllister (1996, sección 6) también ha propuesto que no puede haber experimentos mentales que aclaren la naturaleza de las transiciones discontinuas de estado cuántico —“saltos”, porque su carácter indeterminista no admite fenómenos galileanos. No me resulta claro que su carácter indeterminista no admita fenómenos galileanos; y, si lo hace, no es claro que esto sea suficiente para impedir la experimentación mental; y no es claro que no haya tales experimentos mentales. Bohm (1951, p. 107) imagina un pulso de luz generado por un obturador que es luego absorbido por un átomo en un salto cuántico. El experimento mental muestra que la energía ΔE y el tiempo Δt del salto son gobernados por la relación de incertidumbre $\Delta E \cdot \Delta t \geq h/2\Delta$.

[*conflating*] hechos sociológicos sobre los científicos con lo que es lícito en sus teorías o epistemologías. McAllister alude a “alguna razón alternativa” (p. 248). Yo creo que hay disponibles en abundancia. Los experimentos mentales son argumentación pintoresca de naturaleza hipotética o contrafáctica. Esencialmente, todo lo que se necesita es que la ciencia admita razonamientos hipotéticos o contrafácticos para que admita la experimentación mental. No hay seguridad de que un campo de investigación admitirá tal razonamiento. En muchos círculos históricos, las historias contrafácticas (como las que escribo en Norton, próxima publicación (a)) simplemente se desechan como ilícitas; se nos dice que la academia histórica no apoya los contrafácticos. Sin embargo, es difícil imaginar una ciencia que no admita tal razonamiento. Todo lo que se tiene que hacer es buscar leyes o incluso simples regularidades proyectables; esas dan fundamento a hipotéticos y contrafácticos; y esos a su vez podrían dar sostén a experimentos mentales.

Eso no significa que los científicos deban realizar experimentos mentales. Podrían simplemente no ser útiles. Pero si la ciencia se apoya en contrafácticos, estos son admisibles. Por ejemplo, los sistemas de clasificación botánicos se sustentan en contrafácticos, y por tanto admiten experimentos mentales, aunque parecen haber pocos. Linnaeus, fundador de la botánica moderna, señala (Linnaeus 1786, p. 58) que la diversidad moderna de las especies de vegetales resulta de la hibridización [*hybridization*], no de los cambios del suelo, pues “de otro modo las plantas regresarían a su forma original, si se las devolviera a su situación original”. ¿Qué pasa con una visión muy empobrecida que ve que la ciencia meramente cataloga eventos ocurrentes, sin ir nunca más allá de eso? ¿Podría esa visión impedir la experimentación mental? Curiosamente, el positivismo del mismo Mach estaba cerca de esta manera de ver la ciencia, pero ni siquiera él la desarrolló al punto de impedir los experimentos mentales. Sus cuidadosas reflexiones sobre el tópico fundaron la literatura moderna sobre los experimentos mentales.

Apéndice: ¿Por qué una Transformación de Lorentz rota cuerpos oblicuos?

El resultado puede parecer enigmático, pero es en realidad bien familiar una vez que se lo ve. La relatividad especial nos dice que los cuerpos que han sido puestos en movimiento se contraen en la dirección del movimiento; y que se expanden en esta dirección cuando se los lleva a reposo. Aplíquese esto a un cuadrado con movimiento en su diagonal. Lo que tenemos son los cambios de la figura 4.

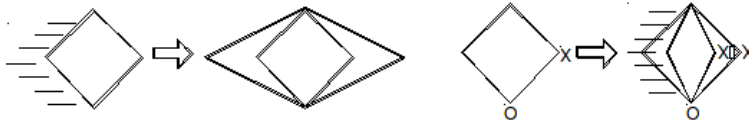


Figura 4: Efecto de la Transformación de Lorentz. El cuadrado en movimiento se expande cuando se detiene, y se contrae cuando se pone en movimiento

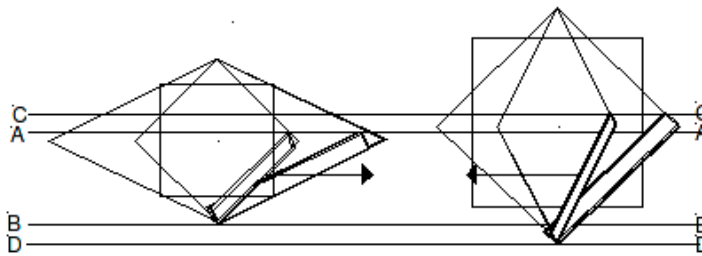


Figura 5: Cómo rota la transformación de Lorentz la barra y la ranura.

Considérese el efecto de la transformación en el lado OX del cuadrado. Es rotado en torno a O a la nueva posición OX'. Las mezclas de rotación y cambio de longitud de la figura 4 son las mismas que

surgen en el experimento mental de la barra y la ranura. Para ver esto, imagínese que la barra y la ranura son los lados de los cuadrados, como se muestra en la figura 5, y que la vista se rota en el espacio de modo que el movimiento relativo es horizontal. La figura muestra una sección a través de la barra y la ranura. Los elementos sombreados son la vista de la ranura; los otros, la vista de la barra.

Con un poco de paciencia, puede recuperarse un análisis cualitativo completo a partir de la figura. Los extremos de la barra son forzados a yacer en las líneas guía AA y BB cuando cambiamos entre la vista de la barra y la vista de la ranura; de forma similar, los extremos de la ranura son forzados a yacer en las líneas guía CC y DD. Se sigue prontamente que la barra pasa por la ranura en la vista de la ranura si y solo si pasa a través en la vista de la barra.

Referencias

- Alper, J., Bridger, M., Earman, J. y Norton, J. (2000): “What is a newtonian system? The failure of energy conservation and determinism in supertasks”. *Synthese* 124, 281-293.
- Bohm, D. (1951): *Quantum Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Brown, J. (1991): *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.
- _____ (1992): “Why empiricism won’t work”. En: D. Hull, M. Forbes, y K. Okruhlik, K. (eds.) *PSA 1992: Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*. Volumen 2. East Lansing, MI: Philosophy of Science Association.
- _____ (1993): “Author’s response”. (to Norton, 1993). *Metascience* 3 (new series), 38-40.
- _____ (en prensa): En: C. Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell.
- Einstein, A. (1911): “Zum ehrenfestschen paradoxon”. *Physikalische Zeitschrift* 12, 509-10.

- Gendler, T. (1998): “Galileo and the indispensability of scientific thought experiments”. *British Journal for the Philosophy of Science* 49, 397-424.
- Horowitz, T. y Massey, G. (1991): *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Hull, D., Forbes, M. y Okruhlik, K. (1992): *PSA 1992: Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*. Volumen 2. East Lansing, MI: Philosophy of Science Association.
- Kuhn, T. (1964): “A function for thought experiments”. Reimpreso en: *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.
- McAllister, J. (1996): “The evidential significance of thought experiment in science”. *Studies in History and Philosophy of Science* 27, 233-50.
- Nersessian, N. (1992): “In the theoretician’s laboratory: Thought experimenting as mental modeling”. En: D. Hull, M. Forbes, y K. Okruhlik (eds.) *PSA 1992: Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*. Volumen 2. East Lansing, MI: Philosophy of Science Association 1993.
- Norton, J. (1991): “Thought experiments in Einstein’s work”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- _____ (1993): “Seeing the laws of nature”. (Review of Brown, 1991). *Metascience* 3 (new series), 33-38.
- _____ (1996): “Are thought experiments just what you thought?” *Canadian Journal of Philosophy* 26, 333-66.
- _____ (en prensa): “Why thought experiments do not transcend empiricism”. En: C. Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell.
- _____ (en prensa (a)): “The n-Stein family”. En: J. Renn (ed.) *Revisiting the Foundations of Relativistic Physics*.

Dordrecht: Kluwer.

- Palmieri, P. (en prensa): “Mental models in Galileo’s early mathematization of nature”. *Studies in History and Philosophy of Science*.
- Linnaeus, C. (1786): *Dissertation on the Sexes of Plants*. (Smith, J. trad.) London: Impreso por el autor, y vendido por George Nicol.

¿Tienen los experimentos mentales vida propia?
Comentarios sobre James Brown, Nancy Nersessian y David
Gooding⁷
Ian Hacking

Creo que los experimentos mentales no tienen vida propia. La corriente de opinión va en sentido contrario. Es evidente si consideramos las amplias citas que han hecho nuestros tres conferencistas. Incluyen, por ejemplo, la sobresaliente colección de ensayos, editados por Tamara Horowitz y Gerald Massey (1991), con su maravillosa introducción editorial que efectivamente dice qué es lo que hay en cada capítulo. Están los capítulos de David Gooding (1990, caps. 8-10) sobre las relaciones entre los experimentos mentales y reales de Faraday. Está el libro de James Brown (1991). Es lamentable que John Norton tuviera que regresar por asuntos personales a Australia, y que no le fuera posible moderar este encuentro porque sus posturas se oponen absolutamente a las de Brown. El “empirismo” del título de hoy de Brown no es otro que el de John Norton (1991). Está también el libro por publicarse pronto de Roy Sorensen, para el que él ha escrito un excelente prolegómeno popular (1991). Además, hay numerosos ensayos por publicarse que no he visto, y que han sido citados por nuestros conferencistas. Finalmente, esta sesión por

⁷ “Do thought experiments have a life of their own?” En: *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1992, Volumen Dos: Symposia and Invited Papers (1992), pp. 302-308. Traducido y reimpresso con permiso de su autor.

sí misma confirma que la discusión de los experimentos mentales está pasando por un auge muy importante.

Este entusiasmo por los experimentos mentales es aún más notable porque nuestros antepasados no han escrito mucho sobre el tema. Como Nersessian nos recuerda, nuestros predecesores filosóficos —en contraposición a los científicos que usaban experimentos mentales en su argumentación y reflexión— son Mach y Koyré. Podríamos además mencionar a Karl Popper, cuyos comentarios en *La Lógica del Descubrimiento* (1959, apéndice XI) fueron elogiados particularmente por Koyré. Y por supuesto tenemos el famoso ensayo de Kuhn (1977 [1964]). Él nos dice que casi lo llamó “La” Función de los Experimentos Mentales—no simplemente “Una” Función, como aparece en la versión final del título.

Hemos escuchado tres enfoques muy distintos hoy. La gente ha quedado choqueada con el platonismo de Brown. Él usa una “p” minúscula, y su platonismo es falibilista [*fallibilist*]. Para mí, el mensaje importante que deja su ensayo podría ser que hay distintos tipos de experimentos mentales, para lo que se argumenta en extenso en el capítulo 2 de su libro. A veces Brown y yo diferimos en cuestiones de nomenclatura. Me resisto a llamar al demonio de Maxwell parte de un experimento mental. Es parte de una fantasía. Aquí estoy de acuerdo con Nersessian, quien me ha observado que es difícil ver qué hay de experimental en el demonio. Quizás es tan sólo un dispositivo retórico para reforzar el análisis estadístico de Maxwell. El demonio no prueba, desde mi punto de vista, la posibilidad de cosa alguna. Tal vez muestra un aspecto de una situación de mecánica estadística. Brown me dice que la gente de buena gana llama al demonio un experimento mental. Ciertamente, eso le ayuda. Afirma que el análisis de Norton es inadecuado, y yo ciertamente no veo que el demonio tenga cabida en el marco conceptual de este. Pero, ¿es un experimento mental? Nersessian dice que los experimentos mentales son típicamente tan convincentes que no hay necesidad de llevar a cabo el experimento. El problema con el demonio de Maxwell como experimento es que

no se puede llevar a cabo de ninguna manera, sin importar cuánta fantasía e idealización uno consienta. Tampoco se puede llevar a cabo otra fantasía notada por Gooding —la de Faraday que evalúa si las cargas eléctricas sentidas por unos ácaros en una esfera están cargadas positiva o negativamente (¿Serán los ácaros ancestros del demonio de Maxwell?). Quizás Gooding diría que el demonio de Maxwell no es parte de un *experimento mental* porque uno no puede ponerse en el lugar de un demonio, o en el de los ácaros de Faraday. No niego, por supuesto, que hay algo como un *continuum* aquí, y comprendo el deseo de contar al demonio como parte de la letanía de experimentos mentales.

En todo caso, la posición debe verse en parte como un contraste a la del ausente Norton, quien sostiene que cualquier experimento mental es sólido mientras pueda plantearse como una forma deductiva válida (Norton 1991). Brown concede esto de algunos experimentos mentales, pero no de todos. En particular, quiere insistir en experimentos mentales que se acercan al entendimiento intuitivo en matemática, en vez de al chequeo combinatorio de una prueba. En mi opinión, le concede demasiado a Norton cediéndole la palabra “argumento”. Hay muchos tipos de argumento regulados por la lógica simbólica. Norton afirma que todos los experimentos mentales pueden plantearse en la forma de deducciones. El ideal es que cada paso sea una instancia de una regla de inferencia. Así es como hemos llegado a definir la palabra “argumento” en lógica elemental. Un argumento es una secuencia de oraciones; es válido si cada miembro de la secuencia es una premisa o se sigue de las oraciones precedentes por la aplicación de una regla de inferencia. Es sólido [*sound*] si las premisas son todas verdaderas. Esa es una poderosa idea logística. Captura una de las tendencias en el pensamiento matemático. Esa tendencia es, para ponerla de una forma quizás demasiado tosca, algorítmica y árabe. Los ejemplos que Brown prefiere son geométricos y griegos. No hay razón para que él conceda que todos los argumentos que usan experimentos imaginarios, si merecen ser convincentes, deban poder

convertirse en pruebas válidas. A lo más tiene que conceder que si no pueden convertirse en pruebas válidas, se ha visto que son falibles. Y Brown es un platonista abiertamente falibilista.

Los enfoques de Nancy Nersessian y David Gooding a primera vista parecen tan diferentes uno del otro como del de Brown, y sin embargo los tres son compatibles —y anti-Nortonianos. Todos ven a los experimentos mentales como algo más que inferencias ocultas. El enfoque de Nersessian viene del ala de la modelación mental [*mental modelling*] de la ciencia cognitiva. Su eslogan es “razonamiento simulado basado en modelos”. El lenguaje de David Gooding sólo se ve diferente al de Nersessian. Su palabra clave es “corporización” [*embodiment*]. Podríamos ver a Nersessian y Gooding combinados tras una lectura del reciente libro de Mark Johnson, *The Body in the Mind* (1987). Nersessian ha confirmado mediante correspondencia que considera que la corporización es un rasgo central de su análisis. Noté que ambos conferencistas han enfatizado la palabra “narrativa”, y Brown la ha usado también. Los tres tienen una actitud esencialmente no-proposicional con respecto a los experimentos mentales. Las historias vienen en oraciones, pero para estos filósofos, una historia se cuenta para ponerse en una situación imaginada. Los tres podrían contestarle a Norton que transformar un experimento mental en una inferencia es transformarlo en algo distinto a un experimento mental. ¡Desde esta perspectiva, Norton no tiene teoría alguna sobre los experimentos mentales! Tiene una teoría acerca de lo que puede construirse a partir de experimentos mentales. Eso a su vez podría proporcionar un criterio de corrección para los experimentos mentales, pero no es el experimento mismo.

Así, los tres conferencistas siguen rutas que difieren de forma importante de la de Norton. Podría ser verdad que hay una derivación de cada conclusión alcanzada en un experimento mental válido. Ciertamente el mismo trabajo de Norton es una manera maravillosa de revelar las deducciones de las historias de Einstein. Y tenemos que lamentar una vez más que Norton no esté aquí para responder contra

este frente común que he identificado en los tres ensayos que acabo de escuchar.

Retornemos a los primeros principios. ¿Cuál es el problema con los experimentos mentales? Mach, quien parece haber inventado el término *Gedankenexperiment*, aparentemente no vio problema alguno. Koyré, quien recogió la idea bajo el nombre de *expérience imaginaire*, tampoco se aproblemó con ellos. Koyré pensaba que los experimentos imaginarios eran abstracciones ideales, y se deleitaba con ellas. En cambio, Mach creía que hay algo así como un continuum entre los experimentos reales y los experimentos mentales. David Gooding es un Machiano. En efecto, nos muestra un *continuum* en el trabajo de Faraday, una cadena de experimentos que culmina en un experimento imposible, un experimento mental. Pero incluso así, podemos ver un problema con los experimentos mentales. Kuhn lo enunció con particular éxito.

Si tenemos que ver con un experimento mental real, los datos empíricos en que se basa han de ser bien conocidos y aceptados generalmente antes de que el experimento fuera incluso concebido. ¿Cómo puede un experimento mental, entonces, dependiendo exclusivamente de datos familiares, llevar a conocimiento nuevo, a una comprensión nueva de la naturaleza? (1977 [1964], p. 241).

Esta afirmación se ha vuelto casi clásica, citada por ejemplo al comienzo del ensayo de Mark Wilson sobre experimentos mentales (1991). Y sin embargo hay algo que no está del todo bien con ella. Hace que nos preguntemos, ¿cómo diantre podríamos descubrir algo sobre el mundo “simplemente pensando”? Pero detengámonos un momento. ¡Lo hacemos todo el tiempo! ¿Qué es la matemática aplicada, después de todo, sino la deducción de lo que no conocemos a partir de lo que conocemos? Hay por supuesto una antigua tradición que afirma que la deducción simplemente nos hace conscientes de lo que estaba contenido en las premisas. Muy bien, pero aun así, nos lleva a ser

conscientes de algo nuevo para nosotros, a una nueva comprensión del mundo. Volveré a Kuhn porque pienso que ha dejado una fuerte impronta en cómo tratar con nuestro tópico. Los experimentos mentales son valiosos cuando sacan a relucir, de una manera sucinta, una tensión conceptual entre dos maneras de pensar, y nos fuerzan a asumirla. Los experimentos mentales no son simplemente una manera de sondear confusiones conceptuales. Es más bien una tensión entre dos maneras de describir el mundo, ambas internamente consistentes, pero una de alguna manera más simple que la otra. La manera más simple es forzada a someterse a la más rigurosa. (Wilson tiene un ejemplo de la más rigurosa sometiéndose a la simple).

Sin cuestionar de ninguna manera ese aspecto de la intelección de Kuhn, me gustaría añadir una manera bastante pedestre de entender su cita. Nos lleva junto al camino de Brown, y a comparar los experimentos mentales con el razonamiento matemático intuitivo. ¿Cuál es el problema filosófico en cuestión? Hay un problema famoso cuyo planteamiento más elocuente nos llega desde hace casi 2500 años. Es el problema del niño esclavo en el *Menón*, a quien se le pide construir un cuadrado del doble del tamaño de un cuadrado dado. ¿Es eso matemática? Podría ser mejor llamarlo geometría racional, de forma semejante a la mecánica racional. Platón estaba desarrollando la idea de que el conocimiento matemático pertenece a un plano más elevado que el conocimiento de las cosas materiales, pero su imagen no era la nuestra más simplista de una distinción entre lo analítico y lo sintético, usada en el siglo veinte por positivistas lógicos y otros para combatir el platonismo. Consideremos a Platón en sus propios términos, sin imponerle nuestras distinciones entre lo matemático y lo empírico.

Al hacer eso, es mucho más natural llamar experimento mental a la historia de Platón que al demonio de Maxwell. El hecho de que lo llamemos matemática no me desincentiva. El argumento presentado en el *Menón* se ajusta a la lista de Nersessian de cinco rasgos característicos de los experimentos mentales:

- Es una narrativa —¡Y qué narrativa!
- Seguimos su desarrollo como lo haríamos en el mundo real.
- El experimento funciona. Tanto Gooding como Nersessian notan que se tiende a no mostrar los ajustes que llevan a un experimento mental pulido. Platón nos muestra los ajustes, y los fallos iniciales.
- El experimento mental funciona mediante abstracciones. En cierto punto se hace que Sócrates diga que las figuras dibujadas en la arena podrían ser más grandes o pequeñas.
- El experimento mental es tan convincente que no hay necesidad de realizar el experimento. En efecto alguien que comenzara a medir, después de ver la prueba final del niño, no habría entendido el argumento (o habría estado en el milenio equivocado, preguntándose sobre geometría no-euclidiana).

El ejemplo dado en el *Menón* es parte de un argumento filosófico. Llevó a Sócrates a la teoría filosófica más sorprendente. El niño recuerda de un alma preexistente, y es ayudado por el proceso dialéctico. La teoría de la *anámnesis* es un recordatorio poderoso de cuán tremendamente puede un pensador ser conmovido por la experiencia de la prueba —la mismísima experiencia a la que el platonismo de Brown nos lleva constantemente.

Pero si ese es el problema (“la dureza [*hardness*] del *debe* lógico”), entonces no se trata de un problema peculiar de los experimentos mentales. Es un problema sobre el conocimiento matemático. Apenas sorprende que Brown nos dé una imagen para aprehender una identidad aritmética, y la llame experimento mental. Nótese que su

ejemplo-imagen aritmético obviamente no satisface el quinto rasgo de Nersessian. No entiendo bien qué sería realizar el experimento, para la proposición general. Puedo tal vez ver que funciona para cualquier numeral, pero no veo que funcione para todo numeral. Pero no tengo duda sobre la coherencia [*cogency*] del objetivo general de Brown. Siempre es una tarea esencial recordarnos cuán extraña es la matemática.

No obstante, hallo un problema con Brown que encuentro también en los otros dos conferencistas. Las *explicaciones* de cómo funcionan los experimentos ofrecidas por nuestros tres expositores, me sirven, sorprendentemente, de bien poco. Hemos sido inoculados contra la variedad de platonismo de Brown por demasiado tiempo. La mayoría de nosotros ve que su idea de familiaridad inmediata con los universales, objetos abstractos, simplemente no nos sirve explicativamente. Pero extrañamente ocurre lo mismo con las ideas presentadas por nuestros otros conferencistas.

Antes de insistir con esa postura, debería reconocer que el hablar de Gooding de corporización es tremendamente importante. Si vemos la imagen del plano inclinado de Stevinus (dibujado por Brown), y pensamos mediante el experimento, no queremos una sola figura, como en la exposición de Brown. Necesitamos al menos una imagen de antes y otra de después del plano y el collar, y necesitamos a alguien recortando con unas tijeras el collar —simultáneamente— a ambos extremos del bloque de madera. Nos vemos a nosotros mismos yendo ahí y cortándolo. En la vida real, tendríamos una enorme dificultad al tratar de hacer ambos cortes simultáneamente —podría de hecho ir al punto más bajo del collar y hacer tan solo un corte. Pero de cualquier forma hay acción corporizada. Así como las imágenes de los experimentos reales omiten usualmente a los experimentadores y lo que hacen literalmente, a menudo olvidamos que es la sensación corpórea del experimento mental lo que nos convence. Norton acierta al producir las derivaciones formales que surgen de los experimentos mentales de Einstein. Pero Gooding acierta al poner énfasis en el

hombre en el elevador. De acuerdo con su postura, creo, ese hombre hace que la historia sea un experimento mental en lugar de una prueba; incluso que el hombre esté drogado, como en una versión del experimento, no deja de ser relevante.

Habiendo defendido la idea de corporización de Gooding, debo repetir que encuentro que los tres conferencistas hacen bien poco por mí cuando examino un experimento mental. Voy a tomar como ejemplo uno especialmente fácil. En marzo de 1686 Leibniz publicó un pequeño libretto anticartesiano en el *Acta Eruditorum*. Tenía un título mordaz: “Una Breve Demostración de un Error Notable de Descartes y Otros Concerniente a una Ley Natural.” Trataba de conceptos desconcertantes como los de *momentum*, energía cinética y trabajo. Esta era una escena Kuhniana par excellence, en la que estaba en cuestión la clarificación conceptual. Hay cierta semejanza entre el argumento de Leibniz y los famosos objetos en caída libre que Galileo ató. Pero el razonamiento es parte de un proceso de sofisticación que condujo a los *Essaie de Dynamique* (1693) y *Specimen Dynamicum* (1695) de Leibniz. En vez de considerar el primer experimento de Leibniz, consideremos un complemento a él de un periodo posterior, que se conecta con la idea de trabajo. El proyecto era concluir algo que casi damos por sentado. “Ha de mostrarse que el poder (*potentiae*) requerido para elevar una libra por dos pies es la misma que el poder requerido para elevar dos libras por un pie” (Leibniz 1969, p. 298).

“Postulo”, escribió Leibniz, “que se me permite asumir varias conexiones de los cuerpos pesados unos con otros, y su ulterior separación, e introducir cualquier otro cambio que no produce un cambio en la fuerza. También nos hacemos de hilos, ejes, levas y otros mecanismos que carecen de peso y resistencia”. Algo de abstracción.

Leibniz probó entonces que un cuerpo de una libra, cayendo dos pies, tendrá exactamente el “poder” (*potentiae* que el traductor pone aquí como “fuerza”, aunque una página antes ponía “poder”) requerido para elevar dos libras una por pie. Y hace esto dividiendo el de dos libras

en dos, y entonces preparando cada mitad sucesivamente, dejando caer en sucesión el peso de una libra un pie, hasta un total de dos pies.

Eso, sugiero, es un modelo [*paragon*] de experimento mental. Tal como lo considero, ninguno de nuestros conferencistas presenta algo que me sirva para evaluar esto. Sí, hay narrativa. Sí, hay corporización. ¿Razonamiento simulado basado en modelos? Parece más fácil pensar en el experimento, o tal vez dibujar una figura, que descender al nivel de los modelos mentales. Y en relación con los universales de Brown con los que tenemos familiaridad, pues eso tampoco me sirve. Este es un tipo de objeción muy poco filosófica a mis colegas, pero una que la persona corriente siente con bastante convicción. Ciertamente está en el meollo de la resistencia general al platonismo de Brown. Y si la ciencia cognitiva no estuviera tan de moda, pienso que muchos de nosotros sentiríamos la misma resistencia a los modelos conceptuales Nersessianos. Son presentados como si fueran explicativos, pero de hecho no explican nada.

Me emociona más, de hecho, leer a Wittgenstein; por ejemplo, una observación en la última parte de sus observaciones sobre las matemáticas. Dibuja una imagen de un mecanismo, una suerte de pistón en el que una rueda se hace girar por una barra fija a ella en un punto A, y que pasa a través de un portaobjetos en un punto B.

Mientras el punto A describe un círculo, B describe una figura ‘ocho’. Ahora escribimos esto como una proposición de cinemática.

Cuando opero el mecanismo su movimiento me demuestra la proposición; del mismo modo que lo haría una construcción en papel. La proposición corresponde e.g. a una imagen del mecanismo con las rutas de los puntos A y B dibujados. De este modo, es en cierto sentido una imagen de ese movimiento. Mantiene lo que la *prueba* me muestra. O —aquello de lo que me persuade (Wittgenstein 1978, p. 434).

La palabra “imagen” de Wittgenstein me sirve mucho más que la palabra de moda, “narrativa”. Es esencial para recordar un pensamiento suyo muy anterior:

El carácter experimental desaparece cuando uno ve el proceso como una imagen memorable (p. 68).

No niego que un modelador mental podría en un futuro distante proporcionar una explicación sobre cómo funcionan las imágenes de Wittgenstein. Mientras tanto, se encontrarán muchas intelecciones aforísticas sobre imágenes y experimentos en las *Observaciones sobre los Fundamentos de la Matemática*. Recomiendo consultar el índice a la tercera edición (1978).

Wittgenstein es, entre otras cosas, un gran autor cómico, de cierta tradición austriaca que nosotros los poco estilosos filósofos de la ciencia tendemos a ignorar. Y cuando examinamos experimentos mentales, especialmente aquellos que nos muestra Brown, podríamos experimentar un toque iconoclasta. ¿Qué me recuerdan esas imágenes? Me recuerdan a chistes, esto es, a cosas muy irónicas y bien trabajadas con un final gracioso. Me recuerdan también a las ilusiones ópticas. Es decir, son experiencias que son difíciles o imposibles de exorcizar. Son rápidas y, como las ilusiones ópticas o las paradojas de probabilidad, siempre te pillan, sin importar lo bien preparado que estés.

Hace más de una década escribí que los experimentos tienen una vida propia. Quería en parte expresar el hecho de que los experimentos son orgánicos, se desarrollan, cambian, y aun así mantienen cierto desarrollo a largo plazo que nos hace hablar sobre experimentos que se repiten y que se replican —véase el ensayo de Hans Radder en esta conferencia (Ver PSA 1992, vol. 1). Pienso que los experimentos tienen una vida: maduran, evolucionan, se adaptan, y no son tan solo reciclados, sino reacondicionados [*retooled*]. Pero los experimentos mentales son bastante estáticos, en gran parte inmutables. Ese es otro aspecto en que son como pruebas matemáticas, pero las buenas pruebas tienen buenas ideas que pueden usarse una y otra vez en

nuevos contextos lo que no es el caso, en general, con los experimentos mentales. Tienen tan sólo una tensión que exponer. Por supuesto, hay partidas falsas, y la exposición se vuelve más precisa con el tiempo. Y aquí vuelve a primer plano el acierto del ensayo de Kuhn. La razón por la que la gente disputa con experimentos mentales, y los usa para exponer ideas y argumentos demoledores [*put-down arguments*], es que pueden revelar tensiones entre una visión del mundo y otra. Pueden hacer que una persona deje de describir el mundo de cierta forma. Pueden reemplazar una imagen por otra. Ese es su trabajo, ahora como en el futuro.

Kuhn, en la discusión posterior a nuestras presentaciones, instó en contra mía a que los experimentos mentales tienen vida propia. Pero estaba señalando algo consistente con lo mío. Insistió en que un experimento mental se repite constantemente, y que no se interfiere con él. La gente vuelve a pensar mediante el experimento, incluso de generación en generación. Pero lo que piensan es algo que fue pensado una vez, y sigue usando los mismos diagramas y diálogo —del mismo modo que la exposición de Platón de Sócrates y el niño esclavo nos lleva a trazar los mismos diagramas, de generación en generación, una vez en la arena, ahora en un papel o en una transparencia. Pero, ¿es esto vida? ¿No es como actuar un papel en una obra teatral? Olivier dio nueva vida a Otelo, pero sin importar lo profundo del poder de Otelo o Lear, esos personajes no tienen vida propia, a pesar de Tom Stoppard. Una vez que el experimento mental es escrito de manera perfecta, es un icono. Los íconos, reitero, no tienen vida propia.

Referencias

- Brown, J. (1991): *The Laboratory of the Mind: Thought Experiments in Natural Science*. London and New York: Routledge.
- Gooding, D. (1990): *Experiment and the Making of Meaning*. Dordrecht: Kluwer.
- Horowitz, T. y Massey, G. (1991): *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Johnson, M. (1987): *The Body in the Mind: the Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.
- Kuhn, T. (1977 [1964]): “A function for thought experiments”. Reimpreso en: *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.
- Leibniz, G. (1969): *Philosophical Papers and Letters*. L. E. Loemker (traductor y ed.). Dordrecht: Reidel.
- Norton, J. (1991): “Thought experiments in Einstein’s work”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Popper, K. (1959): *The Logic of Scientific Discovery*. London: Hutchinson.
- Sorensen, R. (1991): “Thought experiments”. *American Scientist* 79, 149-63.
- Wilson, M. (1991): “Reflections on strings”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Wittgenstein, L. (1978): *Remarks on the Foundations of Mathematics*. Oxford: Blackwell.

Experimentos mentales en el razonamiento científico⁸

Andrew D. Irvine

1. Introducción

Considere la siguiente definición ordinaria de experimento mental⁹: un experimento mental es un caso de razonamiento que intenta extraer una conclusión acerca de cómo es el mundo o de cómo podría ser a través del postulado de una situación hipotética, e incluso contrafáctica. Puesto de manera breve, un experimento mental es un tipo de razonamiento hipotético cuyos supuestos antecedentes acerca de la naturaleza del mundo o del ámbito fenoménico podrían ser falsos, pero que llevan a conclusiones acerca de la naturaleza del mundo o de las cosas que nos rodean.

Como intento inicial, una definición como ésta tiene sus virtudes, pero no es difícil concluir que también tiene sus desventajas. Después de todo, es muy improbable que uno quisiese sostener, como tal definición permite, que *cualquier* ejemplo de razonamiento hipotético que proporciona información acerca del mundo o acerca de lo que nos rodea es un experimento mental. Ciertamente algunos casos de razonamiento hipotético, cualquiera que sean sus propósitos, no son

⁸ “Thought experiments in scientific reasoning”. En: T. Horowitz y G. Massey (1991, eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield, pp. 149-166. Traducido y reimpresso con permiso de su autor.

⁹ Un experimento mental es a veces denominado como un *gedankenexperiment*. El término parece haber sido creado por Ernst Mach. Véase Mach (1883), en donde discute diversos ejemplos, e.g., pp. 32-41, 159-62, y Mach (1897).

suficientemente científicos para ser estimados como experimentos mentales. Después de todo, si los experimentos mentales en efecto generan conocimiento científico, parecería plausible sostener que el tipo de métodos sistemáticos y rigurosos de investigación que caracterizan a otras prácticas científicas deberían también aplicarse en el caso de los experimentos mentales. Si ello es así, se necesita precisar una explicación más detallada de qué caracteriza a los experimentos mentales genuinos de otras formas menos científicas de razonamiento.¹⁰

Teniendo presente lo anterior, el propósito de lo que viene es examinar la naturaleza y el rol de varias formas de razonamiento usadas en las ciencias, que frecuentemente se agrupan bajo el rótulo de “experimento mental”. En particular, la tarea consistirá en distinguir los experimentos mentales en ciencia de otras formas igualmente frecuentes de razonamiento hipotético. De este modo, ¿qué es un experimento mental? ¿Cuál es su relación —similitudes y diferencias— con un experimento físico? ¿Cómo se pondera su resultado? ¿Qué relación tienen los experimentos mentales con las teorías científicas? Y así sucesivamente.

En el transcurso de la deliberación de tales interrogantes, se abogará a favor de la siguiente tesis: los experimentos mentales, en primer lugar, no son nada más que argumentos acerca de eventos particulares o situaciones de una naturaleza hipotética (y a menudo contrafáctica) que llevan a conclusiones acerca de la naturaleza del mundo que nos rodea. Después de todo, los experimentos mentales no tienen como objetivo mostrar nuevos datos empíricos y, debido a que no se pretende

¹⁰ Los debates acerca de la distinción entre ciencia y no-ciencia tienen un rol central dentro de la mayor parte de la filosofía del siglo XX. Primero como resultado del movimiento positivista lógico y más tarde a través de los escritos de Popper y sus contemporáneos. Para los objetivos actuales, no se debe ahondar en el debate acerca de *cuáles* características distinguen la ciencia de la no ciencia. Solo se debe asumir que las ciencias científicas y no científicas son distintas, y que las características más relevantes de las primeras, cualquiera que sean estas en detalle, pueden determinarse de una manera razonablemente amplia. Para la contribución presente al debate histórico que ayuda a indicar por qué la distinción entre ciencia y no ciencia se ha desdibujado recientemente, véase Store (1982).

que generen alguna información nueva acerca del mundo, la única manera alternativa de obtener dicha información debe necesariamente consistir en reconsiderar datos antiguos mediante argumentación. Este punto, al menos, no debería ser controvertido.¹¹ Sin embargo, tal como los experimentos físicos, los experimentos mentales también tienen una relación preponderante con un conjunto de observaciones anteriores y con marco teórico bien desarrollado. Es este requisito adicional lo que distingue a los experimentos mentales de otras formas de razonamiento menos científicas y menos experimentales. El resultado es que, tal como no toda observación acerca del mundo puede decirse que sirva para constituir un experimento físico, no todo caso de razonamiento hipotético o contrafáctico puede decirse que se constituye como un experimento mental.

Este paralelo entre experimentos físicos y experimentos mentales es bastante importante. Por ejemplo, tal como con los experimentos físicos, muchas (aunque no necesariamente todas) las suposiciones de un experimento mental deben ser apoyadas por observaciones empíricas confirmadas independientemente; el experimento mental debe realizarse dentro del contexto de un marco teórico bien constituido; debe tener algún tipo de asidero en cómo uno responde preguntas específicas dentro de contextos teóricos o prácticos; deben existir variables independientemente aisladas para que uno sea capaz de establecer correlaciones entre las variaciones de estas variables y las del resultado del experimento; y así sucesivamente. Además, debe notarse que, tal como un experimento físico tiene a menudo repercusiones para la teoría de base en relación con su confirmación, la falsación, etc., así también ocurrirá con un experimento mental.

¹¹ Desgraciadamente, aunque (por las razones expuestas) este punto no debería ser controvertido, lo es. Para una perspectiva opuesta, en la cual los experimentos mentales son vistos como ventanas a lo no observado véase Brown (1986). También en Koyré (1968), que Brown cita de manera aprobatoria pero no exenta de crítica. La dificultad con todas estas teorías, por supuesto, es al dar una explicación epistemológica satisfactoria y naturalista del nuevo conocimiento que supuestamente emerge en tales casos.

Por supuesto, la analogía no es exacta: los experimentos mentales, a diferencia de los experimentos físicos, no involucran intervenciones reales en el ámbito físico. Podría darse el caso de que algunas otras características de los experimentos físicos (tales como dificultades relacionadas con el tiempo y el ámbito físico) podrían ser factores ocasionales en el caso de los experimentos mentales. No obstante, sin este tipo de características comunes los experimentos mentales nunca podrían distinguirse, incluso en principio, de otras formas de razonamiento igualmente frecuentes.

Al reconocerse lo anterior, resulta útil considerar otro punto. A veces se sostiene que la razón por la que la definición inicial de experimento es demasiado inespecífica es que simplemente permite bajo el rótulo “experimento mental” casos de razonamiento en los cuales todas las premisas podrían resultar verdaderas. En este sentido, lo que caracteriza a un experimento mental de otras formas de razonamiento hipotético es que el primero debe necesariamente tener algún elemento contrafáctico. Mal que mal, los ejemplos típicos de experimentos mentales a menudo se basan en suposiciones contrafácticas. Los experimentos mentales ocasionalmente proponen cosas tales como superficies perfectas sin fricción, vacíos absolutos, cuerpos perfectamente elásticos, o una economía de mercado ideal.¹²

Sin embargo, este requisito más estricto de contrafactualidad [*counterfactuality*] en la definición de un experimento mental puede considerarse como innecesaria, especialmente cuando se toma en cuenta que: i) muchos experimentos mentales están pensados

¹² Como ejemplo de alguien que sostiene que los experimentos mentales se caracterizan (al menos en parte) por su naturaleza contrafáctica véase Brown (1986). Él afirma que los experimentos mentales deben distinguirse de experimentos *puramente imaginados* dado que los primeros, y no los segundos, son “tecnológicamente, físicamente o conceptualmente posibles” (p. 3) Mediante una *ad hominem* contra Brown puede destacarse que él mismo es incapaz de sostener este requisito pues él también reconoce el famoso experimento mental de Galileo acerca de los objetos que caen libremente (el que no es tecnológica, ni física, ni conceptualmente imposible, incluso en el siglo XVII) como un paradigma de experimento (pp. 9 y ss.).

como antecediendo experimentos reales en los cuales las premisas originales del experimento mental son en efecto verdaderas; y ii) muchos experimentos mentales solo incluyen premisas que se saben verdaderas o acerca de las cuales uno se siente agnóstico. En este último caso, tales premisas podrían ciertamente resultar ser contrafácticas pero, tal como sabemos en ese momento, podrían haber resultado verdaderas [*actualized*].¹³ Ambos casos sugieren que el requisito de contrafactualidad necesaria debería descartarse.

2. Los experimentos mentales en la física

Permitásememe considerar ahora si mi tesis en relación con la naturaleza de los experimentos mentales puede defenderse. Debo comenzar caracterizando un experimento físico “típico”, la clase de experimento usado ocasionalmente en una disciplina que innegablemente ha empleado tanto experimentos físicos como mentales a través de toda su historia: la física.

Típicamente,¹⁴ un experimento físico puede definirse como una

¹³ Por ejemplo, considérese el caso de un físico hipotético que sabe de la invención del ciclotrón moderno. Trabajando extremadamente solo, podría razonar de la siguiente forma: “Hoy entendemos bastante bien el proceso de la desintegración radioactiva, al menos en teoría, como se presenta en elementos tales como el radio-226, que eventualmente se desintegra a plomo-206. Ahora bien, si pudiésemos introducir algún mecanismo capaz de acelerar cierta partícula subatómica a altas velocidades, seríamos capaces de inducir el mismo resultado. Al bombardear un elemento con neutrones a alta velocidad, desestabilizaríamos su núcleo y así iniciaríamos un proceso mediante al cual uno de los neutrones se convierte en un protón a través de declinación beta. Sin duda, el resultado sería la creación de otro elemento ligeramente más pesado”. Aunque él no está consciente, su experimento mental ha sido realizado efectivamente. Ciertamente, este solo hecho no determina si su instancia de razonamiento puede contar como un experimento mental.

¹⁴ Por supuesto no todo experimento físico tiene cada una de las siguientes características. (Por ejemplo, las observaciones relacionadas con los experimentos en astronomía podrían no ser repetibles. Véase la sección IV abajo.) De este modo, en el caso de los experimentos mentales y en el caso de los experimentos físicos es improbable que exista un conjunto de características específicas que son en su conjunto tanto necesarias como suficientes para una definición. Sin embargo,

intervención diseñada en la naturaleza o en el laboratorio cuyas consecuencias observacionales sirven para testear alguna hipótesis previamente desarrollada a la luz de observaciones previas, y dentro de un contexto teórico particular. De manera típica también, un experimento se lleva a cabo bajo condiciones controladas y reproducibles. Estas condiciones se asocian con eventos particulares o con situaciones desde las cuales pueden extraerse conclusiones más generales acerca de la naturaleza del mundo. Las observaciones efectuadas a menudo llevan a suplir o a revisar el contexto teórico original.

Además, debe agregarse que los experimentos físicos usualmente permiten identificar y distinguir una cantidad bien definida de variables (o antecedentes) independientes. Esto ocurre para que puedan hacerse observaciones relacionadas entre las variaciones en las condiciones originales del experimento y un conjunto adicional de variables dependientes. Es decir, existe una suposición en cada experimento mental de que las condiciones que de hecho causan los efectos observados pueden individuarse a través de factores variables independientemente, de modo tal que, al alterar uno o más de estos factores, y manteniendo otros constantes, se pueden determinar distintas relaciones causales. En resumen, los experimentos reales típicamente incluyen procedimientos para testear hipótesis en los cuales las condiciones antecedentes del experimento se regulan o controlan para descubrir sus relaciones con el resultado del experimento.

Ahora bien, si en verdad existen similitudes reales entre los experimentos físicos, por una parte, y los experimentos mentales, por otra, uno debería esperar que muchas de las mismas características mencionadas arriba se incorporasen a un experimento mental típico. A modo de caso de prueba, consideraré el siguiente experimento mental asociado con la teoría especial de la relatividad:¹⁵ permítaseme asumir hipotéticamente que estoy en una pieza de vidrio que se mueve a una

debería ser claro que en ambos casos hay ejemplos *típicos* o *paradigmáticos* que podrían caracterizarse.

¹⁵ Para el ejemplo más familiar de Einstein acerca de los trenes véase Einstein (1917), pp. 25-27.

velocidad alta y uniforme hacia un observador externo. En el centro de la pieza existe una ampolleta que parpadea periódicamente. En el momento exacto en que uno y la pieza alcanzan al observador externo la ampolleta resplandece. ¿Existe alguna diferencia entre lo que uno ve y lo que el observador ve?

La respuesta de Einstein a esta pregunta es obviamente afirmativa. Dado que la luz se aleja de la ampolleta a una velocidad constante, y dado que todas las paredes de la pieza son equidistantes de la fuente de la luz, se infiere que uno verá la luz tanto en la pared de enfrente y en la trasera al mismo tiempo. No obstante, para el observador externo existe una diferencia fundamental. Desde su punto de vista fuera de la pieza, la pared de enfrente se mueve hacia delante, alejándose de la luz que se aproxima. Al mismo tiempo la pared trasera se mueve hacia delante para encontrarse con ésta. De este modo, desde el punto de vista del observador externo, la luz alcanzará la pared trasera antes que alcance la pared delantera. Lo que fue simultáneo desde la posición de uno dentro de la pieza será entonces distinto. La conclusión a la que uno llega es que el mismo evento aparecerá diferente dependiendo del marco de referencia del observador.

Este experimento mental se parece a un experimento físico en varios aspectos cruciales. Varias de sus premisas (tales como enunciar la constante de la velocidad de la luz) son suposiciones hechas a partir de observaciones previas, o como resultado de un contexto teórico particular. Existe una hipótesis (en este caso con respecto a la simultaneidad de los eventos) que es testeada y cuya corroboración o falsación tendrá implicancias para la teoría de base. Es también posible aislar variables antecedentes particulares (tales como la velocidad de la pieza relativa al observador) y variarlas en repeticiones del experimento para determinar los efectos distorsionados (si los hubiera) del resultado.

Con estas cosas en mente es posible también ver que muchos de los parámetros del experimento han sido determinados solamente por sus contextos teóricos. Por ejemplo, dado que la Teoría Especial de

la Relatividad se aplica solamente a sistemas de coordenadas que se mueven de modo uniforme en relación con sí mismos, es esencial que la pieza imaginada se conciba moviéndose uniformemente en función del observador externo si el experimento tiene la aplicación que se le había atribuido. Del mismo modo que un experimento físico, este experimento mental trata primordialmente con eventos particulares para ayudar a llegar a una conclusión más general acerca de la naturaleza del mundo que lo rodea a uno. Las teorías pasadas y presentes podrían tener que revisarse a la luz de la conclusión del experimento.

Donde el experimento difiere de un típico experimento físico es en el caso de lo que podrían llamarse sus características *formales* en contraste con sus características científicas. Por ejemplo, mientras que un experimento físico típicamente depende de una intervención real en la naturaleza, el experimento mental toma la forma de un argumento basado en premisas hipotéticas. Un experimento físico usualmente descansará en los instrumentos y métodos de medida; en cambio, un experimento mental generalmente enfatizará diferencias que sean de una naturaleza primordialmente cualitativa. Sin embargo, lo que debe tomarse en cuenta es que las diferencias en las características formales no afectan el rol científico básico del experimento mental. Tal como los experimentos físicos, los experimentos mentales todavía se llevan a cabo dentro de un contexto teórico particular, todavía generan evidencia acerca de la naturaleza del mundo que lo rodea a uno, y todavía se ajustan a los requerimientos inherentes a alguna forma de método científico.

3. Los experimentos mentales en la Antigüedad

A diferencia de las observaciones hechas más arriba, consideraré una hipótesis adicional, a saber, que fueron los presocráticos quienes introdujeron el uso de los experimentos mentales en el razonamiento acerca de la naturaleza y, al hacerlo, fueron los que introdujeron un instrumento eficiente y versátil que probaría ser esencial para el desarrollo posterior de las ciencias.¹⁶

¹⁶ Una tesis de este mismo tipo es defendida por Rescher.

Si la hipótesis considerada en la sección anterior es correcta y los experimentos mentales, tal como los experimentos físicos, solo pueden llevarse a cabo dentro del contexto de un método científico razonablemente complejo, uno debería sentirse inmediatamente incómodo con una afirmación como esa. Puesto de una manera simple, esta incomodidad aparecerá una vez que uno recuerde el hecho de que, a pesar de su importancia en la historia del pensamiento occidental, los presocráticos sólo representaron un primer paso en el desarrollo de la ciencia, tal como se concibe en el presente. Es improbable que aquellos pioneros de lo que fue (a lo más) una ciencia primitiva, o (lo que es más probable) una pre-ciencia sofisticada, pudieran haber introducido un instrumento científico de la complejidad de un experimento mental.

Por supuesto, de haber sido así, se debe agregar que ello en ningún modo afectaría las contribuciones intelectuales logradas por estos pensadores. Tal conclusión no negaría el hecho de que los presocráticos fueron los pioneros en el desarrollo de los experimentos mentales, tal como fueron los pioneros en las ciencias naturales y en las ciencias del razonar en general. Ellos introdujeron en occidente una curiosidad intelectual acerca del mundo natural en conjunción con la conciencia de la importancia de la observación y del método crítico que todavía se emula hoy en día. Según la aguda reflexión de un autor, los griegos presocráticos “se atrevieron a mirar la naturaleza misma” (Kline 1980, p. 10) y fue esto, a diferencia de cualquier otra teoría particular o metodología exacta, lo que constituyó su mayor logro.

Sin embargo, tal como no puede decirse que las primeras incursiones en la naturaleza del mundo por parte de los presocráticos pudieron haber constituido la realización o la puesta en marcha de la ciencia, tampoco puede decirse (al menos en la mayoría de los casos) que el razonamiento hipotético presocrático pudo haber constituido nada más que un paso inicial hacia un desarrollo e implementación de los experimentos mentales. Tal como la ciencia y su método tardaron siglos en convertirse en una forma madura y reconocible, así también ocurrió con los experimentos mentales. Puesto de una manera concisa, aunque los presocráticos eran adeptos al uso de razonamiento

hipotético, *reductios* y conjeturas explicativas, no se sigue que el uso de estas formas de razonamiento deben en sí mismo considerarse suficientes para probar la existencia de los experimentos mentales. Es mucho más probable que el uso de razonamiento hipotético constituyó en sí y de suyo solo un paso inicial pero necesario en el desarrollo de lo que más tarde se convertiría con propiedad en experimento mental.

En defensa de la conclusión precedente, compararé dos ejemplos del uso de razonamiento hipotético en el pensamiento helénico temprano. Uno es el ejemplo de Tales en relación con el mundo natural. El otro es un ejemplo de Euclides muy posterior acerca de la infinitud de los primos.

Consideraré en primera instancia el ejemplo de Tales.¹⁷ Él introdujo la suposición hipotética de que la tierra es como un tronco flotante con el objeto de explicar su posición constate. Su idea es que si uno asume que la tierra es como un tronco que flota, esta suposición a su vez lo ayudará a uno a explicar la creencia común de que la tierra tiene una posición estable. Esta conjetura entonces ganará apoyo debido a su rol explicativo respecto de la creencia original y confirmada independientemente.

El uso de Tales y de los otros presocráticos de argumentos que involucran conjeturas explicativas es importante en varios sentidos. Por ejemplo, indica un interés fundamental en, y que estaban conscientes de, el mundo natural. Además, muestra que, al menos en cierta medida, los presocráticos podrían haber anticipado la moderna teoría de la abducción. De este modo, indudablemente, Tales y sus contemporáneos se involucraron con formas de razonamiento hipotético acerca del mundo natural. Ahora bien, la pregunta que debe formularse es esta: ¿Acaso la conjunción de estos dos factores significa que la conjetura de Tales debería comprenderse como un experimento mental?

De acuerdo con C. S. Pierce, la abducción constituye, junto con la

¹⁷ Rescher cita esta conjetura de Tales como un ejemplo de los experimentos mentales presocráticos. Véase en Horowitz and Massey (1991) pp. 6-7.

deducción y la inducción, una tercera forma general de razonamiento. Según la perspectiva de Pierce, la abducción trata con la selección y la evaluación de hipótesis sobre la base de lo que puede denominarse su contenido explicativo. Considere un hecho conocido F . Según Pierce, en la medida que una hipótesis H puede decirse que explica F (en contraste con la simple implicación de F , o como proveyendo apoyo inductivo para F), puede decirse que este hecho podría más o menos constituir evidencia a favor de la hipótesis H . Puede decirse entonces que H está apoyada por la abducción. En tales casos, la explicación debe consistir en algo más que en simplemente haber deducido o inducido F de H , ya que, por ejemplo, aunque F se siga deductivamente de, por ejemplo, p & $\neg p$, esta conjunción en ningún modo puede decirse *que explica* F . En los mismos términos, aunque el avistamiento de humo podría dar una buena razón para creer que hay fuego, el avistamiento de humo no explicaría la presencia de fuego en el modo que la presencia de fuego explicaría el avistamiento de humo.¹⁸ En síntesis, las explicaciones deben distinguirse de las buenas razones.

Ahora, aunque es claro que la conjetura explicativa de Tales era acerca del mundo natural e involucraba razonamiento de una forma que recuerda la teoría de la abducción desarrollada por Pierce, la pregunta sigue en pie: ¿Es esta instancia de razonamiento suficiente para ser considerada como un experimento mental? Tal como este se constituye, no parece probable. Sin un contexto observacional y teórico más definido en el cual el razonamiento hipotético ocurra, resulta implausible categorizar la conjetura de Tales como un *experimento*, sea como un experimento mental o de otra índole. La conjetura de Tales simplemente figura como una instancia de razonamiento por analogía. Tal como se constituye, no posee algo reconociblemente experimental para considerarse un experimento mental.

Lo que estamos obligados a concluir es que, aunque varios tipos de razonamiento usados por los presocráticos representan un

¹⁸ El ejemplo es de Armstrong quien en efecto usa la abducción en sus argumentos en contra de una teoría de la regularidad de las leyes de la naturaleza.

importante primer paso hacia el desarrollo y uso de los experimentos mentales en ciencia, esto es todo lo que aquellos representan, a saber, un primer paso. Tal como este grado de sofisticación en general en la experimentación no pudo encontrarse en los primeros intentos de los presocráticos en la ciencia, tampoco aparece en sus intentos iniciales de razonamiento acerca del mundo natural.

A guisa de comparación, consideraré la famosa prueba de Euclides de la existencia de una infinidad de números primos.¹⁹ La prueba es una *reductio*. Considérese primero la serie de los números primos conocidos ordenados de modo ascendente, esto es,

2, 3, 5, 7, 11, 13, 17, 19, 23, ...

Para probar que esta serie nunca llega a un final, primero supondré que lo hace. Si esto es así, debería existir un número P , tal que P es el número primo más grande y que

2, 3, 5, 7, 11, 13, 17, 19, 23, ... P

es la serie completa.

Entonces, consideraré que el número Q definido por la siguiente fórmula

$$Q = (2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17 \times 19 \times 23 \dots P) + 1$$

Por supuesto, Q no puede ser un número primo porque por definición es mayor que cualquier primo que termina la serie que finaliza con P . Pero si no es primo, debe ser divisible por un factor distinto que 1 y sí mismo y por lo tanto divisible por algún otro primo. Evidentemente, Q no es divisible por cualquiera de los primos que completan la serie, pues en cada caso quedará un remanente de 1. Así, contrariamente a la suposición original, debe existir un número primo adicional mayor que P . Y debido a que P es un número al azar, debe

¹⁹ Lo que sigue es un parafraseo de la prueba de la proposición 20 del libro IX de *Los Elementos* de Euclides. Véase Heath (1956), volumen 2, pp. 412 y ss.

existir un número infinito de primos.

Ahora bien, ¿qué se debe decir respecto de esta muestra de razonamiento hipotético? A diferencia del caso de la conjetura de Tales, este argumento es mucho más parecido a un experimento mental en varios sentidos. Usa razonamiento hipotético para testear una hipótesis específica (acerca de la cantidad de números primos) que se originó a partir de un contexto teórico particular y bien definido. Las suposiciones hechas pueden verse como originadas en, y ser justificadas por, el mismo contexto. Aparecen variables aislables independientemente, y el resultado de este razonamiento, tendrá ramificaciones, no sólo con respecto a la teoría de las matemáticas, sino además en relación con muchas creencias asociadas con lo finito y lo infinito. Este ejemplo de razonamiento hipotético parece ser más parecido a un experimento mental en muchos más aspectos que el caso del ejemplo de Tales.

Sin embargo, también debe agregarse que el ejemplo de razonamiento hipotético mencionado arriba parece, *prima facie*, tener que ver muy poco con el mundo físico. Si esta instancia de razonamiento debe de hecho clasificarse como un experimento mental, será claramente un experimento mental en las matemáticas y no en las ciencias naturales. A diferencia de los experimentos mentales en dichas ciencias, ese experimento mental tiene muy poco que ver con observaciones pasadas. No es claro del todo, además, (al menos no sin tener una explicación metafísica adecuada²⁰) si este experimento mental podría testearse por medio de experimentos físicos relacionados, y si pudiese, cuáles serían dichos tests.

²⁰ Resulta que creo que puede darse una explicación de ese tipo y que, en consecuencia, las matemáticas se comprenden mejor si se piensa que son más similares a las ciencias naturales de lo que tradicionalmente se asume. Según tal tipo de explicación, las matemáticas pueden verse como distintas de la física, la química, y otras disciplinas, pero tan solo como una cuestión de grado, y no como una diferencia de naturaleza. Sin embargo, dentro del contexto del presente argumento, nada depende mucho de si esa explicación resulta exitosa.

Lo que estas últimas reflexiones indican es que la característica definitoria de los experimentos mentales resultará ser, en parte, dependiente de su área de aplicación. Así, podría resultar que dentro de algunos cuerpos de conocimiento organizado varias de las características que uno atribuiría a los experimentos mentales no estarán presentes en otras disciplinas. Por ejemplo, en algunos casos podría existir un contexto teórico satisfactorio así como un conjunto definido de variables antecedente. Pero podrá también resultar que algunas características científicas, tales como las que normalmente requieren de un alto grado de dependencia de las observaciones pasadas, variarán. En esos casos esta variabilidad será mejor comprendida como una función de la clase de restricciones inherentes al método experimental de la disciplina misma, en vez de ser un resultado de una característica intrínseca de los experimentos mentales en general. Con todo, en tales casos, la característica esencial del experimento mental aún sería que está fuertemente vinculado con los intereses, prácticas y metodología de la rama de la ciencia en cuestión.

4. Los experimentos mentales en la astronomía

Como ejemplo de la variabilidad de las características científicas particulares dentro de las disciplinas, consideraré el caso de los experimentos en la astronomía. Los experimentos físicos en la astronomía difieren de los experimentos físicos ordinarios en varios aspectos: no se realizan en las condiciones controladas de un laboratorio, no involucran intervenciones en la naturaleza, y muy a menudo sus consecuencias observacionales no son reproducibles (al menos no dentro de lo que podría estimarse en un período de tiempo razonable). A pesar de esto y de la ausencia en la astronomía de las características “típicas” asociadas con los experimentos físicos de, por ejemplo, la física o la química, importantes experimentos físicos y mentales se llevan a cabo dentro de esta disciplina.

A modo de ejemplo, podría hacerse mención del siguiente experimento mental del siglo XVIII que dio lugar a la ahora famosa

paradoja de Olbers.²¹ Asumiré hipotéticamente que todos los objetos que emiten luz o fuentes de ondas de radio, tales como las galaxias, se distribuyen uniformemente a través del espacio.²² También asumiré inicialmente que el universo es infinito en tamaño. En pro de la simplicidad, podría incluso asumir que es euclidiano. Hechas estas suposiciones, intentaré entonces calcular la cantidad de radiación total que hay en el cielo de una noche.

Ahora, dado que la intensidad de la luz que nos alcanza desde cualquier galaxia a una distancia r desde la tierra, disminuirá según un factor de $1/r^2$ podría parecer que solo las galaxias más cercanas contribuirán de manera significativa a la luz del cielo de una noche. Sin embargo, las fuentes de luz de distancias más lejanas podrían jugar de todas formas un rol importante por acumulación. Esto resulta así porque se ha asumido una distribución uniforme de las galaxias. En esas condiciones, el número de galaxias a una distancia r se incrementará a un factor de r^2 . Es decir, si se duplica la distancia, solo un cuarto de la luz alcanzará la tierra desde cualquier galaxia particular, pero dado que el número de galaxias a esa nueva distancia será proporcional al cuadrado de la nueva distancia, los dos factores se anulan mutuamente.²³ Así, la contribución de luz recibida desde las

²¹ El problema fue originalmente propuesto por Edmund Halley en 1720 y reelaborado más formalmente por el astrónomo suizo Chéseaux en 1743. No obstante, no fue sino hasta la década de 1820 que recibió su reformulación más famosa en las *ruinations* de Heinrich Wilhelm Matthäus Olbers (1758-1840). Para una discusión científica más detallada de este problema, véase Sciama (1971), Cap. 8, pp. 98-127 (especialmente pp. 112 y ss.); Hoyle (1955), Cap. 18, pp. 304-321; y Wesson, Vale y Stabell (1987).

²² Más específicamente, se asume que el universo es tanto homogéneo como isotrópico. Por lo tanto, la densidad promedio de las galaxias, y así la luminosidad promedio, puede asumirse ser la misma en todas partes.

²³ Esto es, por una parte, la intensidad de la radiación recibida de cualquier fuente particular es inversamente proporcional al cuadrado de la distancia; pero, por otra parte, el número de las fuentes de la radiación es directamente proporcional al cuadrado de la distancia. Así, el número de las galaxias se incrementa como una función de r en una forma tal que compensa exactamente el debilitamiento de la radiación a causa del incremento de la distancia.

galaxias a la distancia r será independiente de r misma. La disminución en intensidad desde cualquier otra galaxia, se compensará exactamente con el incremento de la radiación a causa del aumento del número de fuentes a esa distancia. Dado que también se ha asumido que r puede llevarse al infinito, la conclusión ineludible es que, en tanto que se extiende r , el aumento en la radiación sólo se incrementará proporcionalmente hasta que el cielo completo pareciese cubierto por un gran sol.

Tal como antes, el experimento mental incluye muchas de las características científicas requeridas por cualquier experimento, sea mental o físico. Tan importante también, tiene ramificaciones para la teoría de base en la cual la paradoja fue desarrollada. Por ejemplo, durante el siglo XIX la solución más comúnmente propuesta para la paradoja resultante consistía en asumir que el universo no estaba poblado de modo uniforme, y que las galaxias estaban entonces localizadas en algunas áreas del universo y no en otras. Hoy, sin embargo, estos intentos de solución han sido desechados. Las observaciones hechas a distancias de decenas de megaparsecs²⁴ no han dado indicación alguna del tipo de localización que se requeriría para evitar la paradoja de esa manera.

A diferencia de lo propuesto en el siglo XIX, la solución contemporánea más satisfactoria de la paradoja enfatiza dos factores por separado. En conjunción estos dos factores explican cómo se suavizan las consecuencias de la contribución hecha por las distancias lejanas respecto de la luz de fondo observable extragaláctica. El primero de estos dos factores es que el universo, incluyendo todas las distancias de radiación distantes, ha existido por un período finito de tiempo. Debido a que el tiempo requerido por las emisiones de galaxias lejanas para alcanzar la tierra es en muchos casos mayor que la edad de esas galaxias, resulta que el tiempo finito de duración del

²⁴ Un parsec es igual a 3.2615 años luz, o el desplazamiento aparente de un objeto cuyo movimiento en el cielo a causa del movimiento de la tierra alrededor del sol (o paralax) es un segundo de arco. Un megaparsec es igual a un millón de parsecs.

universo se convierte en un factor crucial en cualquier cálculo de la radiación de fondo.²⁵

El segundo factor es que ahora generalmente se acepta que el universo está en expansión. Tal expansión a su vez causa un debilitamiento de la radiación recibida de galaxias distantes. Debido a que se piensa que todas las galaxias se alejan unas de otras a una velocidad proporcional a la distancia que las separa, resulta que mientras más distante está una galaxia de la tierra, mayor será el debilitamiento de la radiación que se recibe desde la misma. Una estimación es que dadas estas condiciones la luz no podrá alcanzar la tierra desde distancias que sean superiores a 2000 megaparsecs.²⁶

En la actualidad, la teoría de un universo en expansión de edad finita ha recibido confirmación por parte de dos fuentes independientes. Primero, hay confirmación a partir de observaciones del denominado corrimiento al rojo, observaciones del corrimiento de las líneas de espectro hacia las ondas más largas. El corrimiento resulta de un cambio en la frecuencia ocasionado por el movimiento relativo entre el observador y la fuente; en resumen, por la expansión del universo. Segundo, la confirmación viene como resultado de los cálculos de la edad del universo sobre la base de la declinación de la tasa de radioactividad (por ejemplo, los isótopos U-235 y U-238), los cuales producen cifras comparables a aquellas determinadas por el corrimiento al rojo.

5. Acerca del valor de los experimentos mentales para la ciencia

Dadas las consideraciones precedentes, ¿qué puede decirse ahora en general en relación con la experimentación mental en el ámbito de la física? Mi conclusión inicial tiene dos partes. Primero, puede defenderse la intuición de que debería hacerse una distinción entre el razonamiento hipotético en general y los experimentos mentales en particular. En disciplinas tan diversas como la astronomía y la física

²⁵ Para más detalles véase Wesson, Valle y Stabell (1987).

²⁶ Hoyle (1955), p. 314.

puede encontrarse una situación que permita argumentar que los experimentos mentales pueden y deberían distinguirse de ejemplos de razonamiento hipotético simple. Segundo, si el análisis precedente es acertado debería haber alguna forma de indicar por qué sucede que no toda instancia de razonamiento hipotético y contrafáctico será un experimento mental. Los experimentos mentales pueden entenderse como argumentos acerca de eventos hipotéticos particulares o como estados de las cosas [*states of affairs*] con una relación privilegiada tanto con las observaciones empíricas previas como con una teoría de base particular. La relación privilegiada entre el experimento mental y su contexto observacional/teórico requiere que el experimento mental exhiba al menos varias, sino todas, las características científicas propias de los experimentos en general. Como una lista representativa (aunque no exhaustiva probablemente) de tales características, se han mencionado las siguientes:

- Un experimento mental debe ser relevante para el testeo de alguna hipótesis (o para contestar algún conjunto de preguntas) que se originó dentro un contexto observacional/teórico particular.
- Muchas aunque no todas de las suposiciones dentro de un experimento mental deben apoyarse por observación empírica independientemente confirmada. En síntesis, al menos algunas características del experimento mental deben basarse en el mundo observable si es que tiene relevancia para la investigación científica ordinaria.
- El experimento mental debe plantearse con suficientes detalles (en un ambiente suficientemente controlado, por llamarlo así) para que sea capaz de ser repetido.
- Debe ser posible identificar un número de variables independientes (o antecedentes) dentro

de un experimento mental a fin de determinar correlaciones entre las variaciones de estas variables y un conjunto adicional de variables dependientes usadas para caracterizar el resultado del experimento.

- El resultado del experimento mental debería tener repercusiones para la teoría original de base. El razonamiento involucrado respecto del evento hipotético particular o la situación en cuestión debería generar evidencia a favor o en contra de alguna conclusión general acerca del mundo o lo que lo rodea a uno. Esto a su vez conducirá a complementar o a revisar el contexto teórico original del experimento mental.

De esta manera, así como no de toda observación o manipulación del mundo alrededor de uno puede aseverarse que constituye un experimento físico, no toda instancia de razonamiento hipotético acerca del mundo natural será considerada como un experimento mental genuino. Solo aquellas instancias de razonamiento hipotético o contrafáctico que poseen sus parámetros en gran medida determinados por un cuerpo de cuestiones teóricas u observacionales relevantes contarán como experimentos mentales genuinos. La existencia y utilidad de cualquier experimento (tanto mental como físico) figurará sólo dentro de un contexto de un método científico bien desarrollado y, aunque los primeros diferirán de los segundos con respecto a sus características formales, los experimentos mentales genuinos deben compartir con los experimentos físicos al menos algunas de aquellas características físicas que ayudan a definir las disciplinas particulares y la empresa científica en general. Si no se acepta esta tesis, cualquier distinción que uno pudiese hacer entre el razonamiento científico en general y la experimentación mental en particular corre riesgo de colapsar.

En vista de estas conclusiones, puede decirse algo más para

explicar el valor de los experimentos mentales dentro de las ciencias naturales. En particular, algo puede decirse acerca de las ventajas y de las limitaciones de emplear experimentos mentales. En cierto sentido, las ventajas de los experimentos mentales son claras: estos no requieren de ningún aparato y, como resultado, no poseen las limitaciones prácticas de los laboratorios. Han demostrado también, en variadas circunstancias y bajo diversas condiciones teóricas, ser poderosas herramientas para el avance de la ciencia.

Además, los experimentos mentales tienen la ventaja de ser capaces de llegar a conclusiones acerca de ambientes hipotéticos ideales (superficies sin fricción, cuerpos perfectamente elásticos, situaciones de cero absoluto, y similares), y por consiguiente ayudan a identificar y distinguir entre las variaciones no accidentales de variables y de parámetros fijos. Al omitir la consideración de las variables particulares (tales como el roce de una superficie) la dependencia o independencia de las otras variables podría ser mucho más clara. Una que vez que esto ocurre, las leyes subyacentes pueden aplicarse útilmente a situaciones más complejas del mundo real.

El análisis precedente también ayuda a indicar cómo es que tales razonamientos hipotéticos e incluso contrafácticos acerca de los límites ideales pueden ser informativos sobre las leyes físicas del mundo real. La insistencia en que los experimentos mentales poseen una relación de privilegio en relación con las observaciones pasadas garantiza que estén basadas en el mundo real. En tales casos, postular un límite ideal (o alguna otra premisa contrafáctica) es análogo a ignorar en qué, de modo correcto o incorrecto, se creen ser los factores extraños dentro de un experimento, factores que se estiman irrelevantes para los propósitos o problemas particulares a considerar. De acuerdo con esta explicación, los experimentos mentales pueden verse como simples predicciones hipotéticas o contrafácticas, como predicciones de lo que sería el caso si ciertas condiciones (hipotéticas o contrafácticas) se diesen. No obstante, en ambientes más prácticos y mundanos, serán más provechosamente caracterizados como heurísticas teóricas, como

casos guía que en mayor o menor grado (dependiendo de algunas suposiciones hechas) se aproximarán a las aplicaciones del mundo real.

A diferencia de tales ventajas, las limitaciones de los experimentos mentales pueden ser menos evidentes. La más importante de estas es que de un experimento mental, por definición, no pueden obtenerse nuevos datos empíricos. En consecuencia, se corre el riesgo de ser desviado por un experimento mental cuando la conclusión de este va más allá de la información dada por las premisas o cuando esa información es en sí injustificada. Los experimentos mentales están restringidos por el nivel en que uno puede extraer conclusiones injustificadas desde la evidencia disponible.²⁷ Debe tenerse presente que la razón por la que se considera que los experimentos mentales son útiles es debido a que se cree que *si* se llevasen a cabo, los resultados alcanzados serían los predichos. Este es el caso incluso con respecto a las aseveraciones contrafácticas tales como aquellas relacionadas con superficies sin roce y similares. La aseveración en estos casos es que *si* fuese posible construir una superficie sin roce, entonces tal y tal cosa resultaría. Al final no existe un test mejor que el real.

La razón por la que los experimentos mentales pueden ser falaces tan a menudo es que detrás de casi todo experimento mental yace un gran número de suposiciones auxiliares, suposiciones que se asumen verdaderas, pero que, si fuesen falsas, revocarían el resultado en cuestión. A pesar de sus ventajas, los experimentos mentales no pueden reemplazar la observación y la experimentación verdadera. Para probar esto solo se debe enfatizar que, si un experimento mental y uno real estuviesen en conflicto, será casi siempre el experimento mental, y no el experimento físico, el que tendrá que ser revisado.

Como un ejemplo de los tipos de suposiciones auxiliares presentes dentro de un experimento, considérese el famoso experimento mental

²⁷ Incluso Koyré (1960), quien afirma que “la buena física se hace *a priori*” (p. 88), admite que incluso el experimento de Galileo puede a veces llevar a “afirmaciones que la realidad persistentemente refuta” (p. 82).

de Galileo acerca de los objetos que caen, que es presentado como parte del diálogo durante el primer día de los *Discorsi*. La teoría aristotélica contemporánea en boga entonces era que las “velocidades naturales” de los objetos que caen eran una función de su peso. El experimento mental de Galileo socava [*vitiare*] esta teoría, facilitando el camino para una en la que la aceleración natural de un cuerpo es independiente de su masa. El siguiente diálogo es entre Salviati y Simplicio:²⁸

Salviati: Pero, incluso sin experimento adicional, es posible probar de manera clara, por medio de un argumento breve pero contundente, que un cuerpo más pesado no se mueve más rápidamente que un cuerpo más ligero... Pero dime, Simplicio, si tú admites que cada cuerpo que cae alcanza una velocidad definida, fija por naturaleza, una velocidad que no puede incrementarse o disminuir excepto por el uso de la fuerza [*violenza*] o la resistencia.

Simplicio: No hay duda alguna de que uno y el mismo cuerpo moviéndose en un medio definido tiene una velocidad fija que está determinada por la naturaleza y que no puede incrementarse excepto por la adición de *momentum* [*impeto*] o disminuida excepto por la misma resistencia que lo retarda.

Salviati: Si entonces tomamos dos objetos cuyas velocidades naturales difieren, es claro que al unir ambos, el más rápido será retardado por el más lento, y que el más lento será de algún modo

²⁸ Salviati, el portavoz de Galileo, parece haber sido bautizado después de un amigo de Galileo cuya memoria este quiso perpetuar. Es también tentador pensar que Galileo nombró a Simplicio después de que el comentarista de Aristóteles Simplicius, pero véase Koyré (1960), p. 46 n. 1, para una discusión de este punto.

acelerado por el más rápido. ¿No estás tú acaso de acuerdo conmigo respecto de esta opinión?

Simplicio: Estás indudablemente en lo correcto.

Salviati: Pero si esto es verdadero, y si la piedra más grande se mueve con una velocidad de, por ejemplo, ocho, mientras que la más pequeña se mueve con una velocidad de cuatro, entonces cuando estén unidas el sistema se moverá con una velocidad de menos de ocho; pero las dos piedras al estar unidas conforman una unidad más grande que la que antes se movía con una velocidad de ocho. Por lo tanto, el cuerpo más pesado se mueve con menos velocidad que el más liviano; un efecto que es contrario a tu suposición. Así, tú ves cómo, de tu suposición de que el cuerpo más pesado se mueve más rápido que el liviano, infiero que el cuerpo más pesado se mueve más lentamente (Galileo 1638, pp. 62 y ss.)

Aquí las suposiciones auxiliares tácitas incluyen algunas tales como que los cuerpos que caen pueden ser de hecho exitosamente unidos durante una caída de la manera que Salviati describe. Ellos también incluyen la suposición de que dos piedras, cuando están unidas, conforman una unidad, y así tiene una velocidad natural que es comparable en principio a la de las otras piedras. Si se alteran esas suposiciones, incluso de manera leve, se seguirá un resultado distinto.

Como ejemplo, se puede proponer que, después de hacer consideraciones adicionales, Simplicio no está proclive a asumir la premisa empírica de que dos cuerpos quedarán unidos durante una caída. No está del todo confiado de que en el caso de que dos objetos estén unidos estos no puedan separarse en algún momento de su caída. Incluso podría proponer que esto es resultado del hecho de que

el más rápido es retardado por el más lento, y de que el más lento es acelerado por el más rápido. Si los nudos de la cuerda y similares se sueltan regularmente durante la caída, una explicación tal podría ser que la teoría pre-Galileana era correcta. La discusión más detallada entre Salviati y Simplicio acerca de consideraciones similares muestra que Galileo se percataba de este hecho (véase Galileo 1638, pp. 63 y ss.). La solidez del experimento mental depende crucialmente de la relevancia de las suposiciones empíricas implícitas.

La lección, por supuesto, es que los experimentos mentales, a pesar de su poder y versatilidad, son siempre falibles. Se corre el riesgo de ser desviado por estos en la medida que se emplean para cualquier otra cosa que no sea explorar las consecuencias de los enunciados dentro de un contexto teórico y observacional.²⁹

²⁹ Este ensayo comenzó a fraguarse mientras fui invitado al Centro de Filosofía de la Ciencia de la University of Pittsburg, y continuó más tarde en la University of Toronto. Durante este período, recibí ayuda financiera del Consejo para la Investigación de las Ciencias Sociales y las Humanidades de Canadá. Estoy agradecido de John Norton, Nicholas Rescher y Jim Brown por las interesantes discusiones acerca de este tópico, y especialmente a John Forge por sus perspicaces comentarios a un borrador previo de este ensayo. Finalmente, me gustaría expresar mi gratitud hacia Gerald Massey y Tamara Horowitz por su invitación, primero, a participar en su taller de 1986 sobre experimentos mentales y, más tarde, para contribuir al presente volumen.

Referencias

- Armstrong, D. (1983): *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Brown, J. (1986): “Thought experiments since the scientific revolution”. *International Studies in the Philosophy of Science* vol. 1, 1-15.
- Einstein, A. (1917): *Relativity*. London: Methuen.
- Galileo [Galileo Galilei] (1638): *Dialogues Concerning Two New Sciences*. New York: Macmillan.
- Heath, T. (1956): *The Thirteen Books of Euclid’s Elements* (3 volúmenes). New York: Dover Publications.
- Horowitz, T. y Massey, G. (1991): *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Hoyle, F. (1955): *Frontiers of Astronomy*. New York: Harper & Brothers.
- Kline, M. (1980): *Mathematics*. New York: Oxford University Press.
- Koyré, A. (1960): “Galileo’s treatise *De Motu Graviorum*: the use and abuse of imaginary experiment”. En: *Metaphysics and Measurement*. London: Chapman and Hall.
- Kuhn, T. (1964): “A function for thought experiments”. Reimpreso en: *The Essential Tension*. Chicago: University of Chicago Press.
- Mach, E. (1883): *The Science of Mechanics* (6a edición, traducido por T. J. McCormack). LaSalle, IL: Open Court.
- _____ (1897): “On thought experiments”. En: *Knowledge and Error*. Dordrecht: Reidel.
- Rescher, N. (1987): “Thought experimentation in presocratic philosophy”. En: T. Horowitz y G. Massey (eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.
- Sciama, D. (1971): *Modern Cosmology*. Cambridge: Cambridge University Press.

- Stove, D. (1982): *Popper and After*. Oxford: Pergamon Press.
- Wesson, P., Valle, K. y Stabell, R. (1987): “The extragalactic background light and a definite resolution of Olber’s paradox”. *The Astrophysical journal* vol. 317, 601-6.

Acerca de algunos experimentos mentales viciados³⁰

Richard M. Gale

Los experimentos mentales desempeñan un papel destacado en la práctica de la filosofía analítica, por ser uno de los métodos principales mediante los cuales los análisis se someten a evaluación. Desafortunadamente, muchas veces se les invoca de manera errónea. Sus defensores los publicitan como si demostraran un cierto resultado cuando en realidad muestran algo totalmente diferente. Debido a que no se han descrito de una manera correcta, su resultado positivo no destaca. Una vez que uno capta el resultado verdadero de estos experimentos mentales viciados, uno puede replantearlos y ponerlos al servicio de algo que valga la pena. Mi objetivo es proponer las bases para esta reconceptualización a través de un escrutinio [*botanization*] de diferentes tipos de experimentos mentales. Se espera que esto haga que los experimentos mentales viciados se diferencien de sus congéneres, y que nos den una base conceptual para replantearlos.

Para entender el objetivo de un experimento mental, uno debe comenzar con la forma en la cual el análisis filosófico ha sido tradicionalmente concebido —realmente mal concebido. Se supone que quien lleva a cabo un análisis filosófico enuncia las condiciones necesarias y suficientes para la correcta aplicación del concepto *C* en cualquier caso, tanto real como meramente posible. En aras de

³⁰ “On some pernicious thought experiments”. En: T. Horowitz y G. Massey (1991, eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield, pp. 297-303. Traducido y reimpresso con permiso de su autor.

este objetivo, el analista debe primero localizar alguna palabra p en su lenguaje que exprese o signifique C . Entonces, articula las reglas que prescriben las condiciones de acuerdo con las que es correcto o incorrecto usar p , reglas que tienen una fuerza normativa en la medida que las personas que se involucran en esta práctica gobernada por reglas están dispuestos a ofrecer y aceptar corrección cuando una jugada se desvía de tales reglas. Dicho analista describe, en otras palabras, las reglas de una clase particular de práctica humana regida por reglas normativamente—un juego de lenguaje.

Tan pronto como se logra esto, el Hombre Contraejemplo hace lo suyo. A menudo esto consiste en llevar a cabo un experimento mental, esto es, una descripción de una situación meramente posible o contrafáctica, en la cual se supone que el análisis colapsa. Hay distintas maneras en las que un análisis puede fallar, y probará ser útil clasificar los experimentos mentales en función de dichas maneras.

Una manera obvia es que el análisis enfrente un claro contraejemplo en la situación propuesta: las reglas enunciadas en el experimento requieren que uno use (o se abstenga de usar) p en esa situación, pero uno quiere (o no) usar p . Los experimentos mentales pueden ser instrumentos útiles para presentar tales contra-ejemplos. En algunos casos ahorran un montón de tiempo y esfuerzo. Para refutar el viejo análisis de que el hombre es un bípedo implume no fue necesario ir y desplumar un gallo. Es suficiente afirmar: “imagine que existe un gallo desplumado...” En algunos casos crear la situación imaginada es práctica o incluso físicamente imposible. No tengo cuentas que arreglar [*I have no bone to pick*] con tales experimentos mentales. Podríamos denominarlos *experimentos mentales claramente contraejemplificadores* en función de sus resultados.

Otro tipo de experimento mental auténtico le presenta a uno un caso límite o cuyo resultado es ambiguo. Las reglas enunciadas para el uso de p no determinan qué cosa es correcta afirmar en la situación imaginada. Que existan tales situaciones posibles para p no debería

sorprendernos. Es ya la historia más grande alguna vez contada que no hay un conjunto de reglas que puedan ser decisivas para cada situación posible. Las conclusiones de tales *experimentos mentales de resultado ambiguo* deberían tener un efecto recuperador en la práctica del análisis. El analista ya no puede jactarse de hacer explícitas las reglas para el uso correcto de *p* en cada situación real y posible. El análisis se debe restringir a los casos claramente inequívocos.

Desafortunadamente, muchos casos de experimentos mentales cuyo resultado es ambiguo son falsamente anunciados como situaciones claras de contraejemplos. Esto a menudo sucede cuando *C* es de criterio múltiple.³¹ El concepto actual de identidad personal a través del tiempo, por ejemplo, requiere tanto de continuidad espacio-temporal como de continuidad psicológica. Y un caso en el cual solo uno de estos factores es satisfecho es una situación límite o con un resultado ambiguo. Locke piensa que su experimento mental del príncipe y el zapatero que intercambian cuerpos mediante el cambalache de sus rasgos psicológicos es un claro contraejemplo al requisito de la continuidad psicológica; pero, según mis intuiciones, es solo un problema sin solución.

La mayoría de los experimentos mentales con resultado ambiguo resultan crucialmente indeterminados en tanto que no nos dicen si el caso anómalo es excepcional o sucede frecuentemente, por ejemplo, en la mayoría de las ocasiones. Si es extravagante deja intacta la viabilidad de nuestro juego de lenguaje actual, puesto que las cosas no siempre resultan bien. Es suficiente que lo hagan la mayoría de las veces. Pero si un caso ambiguo es la regla en vez de la excepción el resultado del experimento es radicalmente distinto. Al ver la diferencia en su resultado, el experimento mental viciado será exorcizado.

La diferencia entre un mundo en el cual un caso ambiguo es una

³¹ Para una interesante discusión de los conceptos de múltiples criterios véase Nicholas Rescher (1973, capítulo VI), *The Primacy of Practice*. Rescher muestra cómo la viabilidad de tales conceptos depende de la presuposición empírica de que, en general, no estamos enfrentados con casos ambiguos.

excepción y uno en el cual este es la regla, es la diferencia entre un mundo en el que participamos en el juego de lenguaje con p normativamente gobernado por reglas y uno en el cual no lo hacemos porque sería fútil hacerlo. La descripción del último como un mundo que contiene casos ambiguos es un error grave, pues en tal mundo uno no participa en el juego de lenguaje en absoluto. Por supuesto, uno podría jugar algún juego de lenguaje nuevo que tenga una semejanza de familia con el juego original de lenguaje; pero esa es una situación diferente. Si el mismo concepto C entra tanto en el juego de lenguaje nuevo como en el original no es algo que pueda determinarse, debido a que carecemos de criterios para la identidad de un concepto no solo a través del tiempo, sino a través de mundos posibles.

La razón por la que nuestro juego de lenguaje actual que involucra a p no sería frecuentemente jugado en el mundo en el cual los casos ambiguos ocurren es que es fútil, en general, participar en una práctica normativamente gobernada por reglas cuando en la mayoría de los casos resulta poco claro si una jugada es permitida o no. Imagine cómo sería jugar ajedrez si la mayoría del tiempo no estuviera claro si una jugada es válida o no. Sería como jugar a los dados con las caras de estos borradas. Los más poderosos preferirían este nuevo “juego”, porque les daría la oportunidad de amedrentar a los oponentes más débiles haciéndolos aceptar que, sobre la base de sus recuerdos, cualquier cosa que dijese es la verdadera marca de los dados. En realidad este “juego” está más cercano a lo que uno llama extorsión o robo a plena luz del día.

La distinción de estos dos tipos de experimentos mentales ambiguos descansa en una diferencia importante, pero a veces ignorada, entre las reglas de un juego de lenguaje y las presuposiciones para participar en este juego. Una presuposición para participar en un juego es que no ocurren casos ambiguos relativos a las reglas de un juego la mayor parte del tiempo. Pero hay presuposiciones más sutiles para participar en un juego de lenguaje que tienen que ver con la forma de vida involucrada en el juego, y con las condiciones empíricas bajo las

cuales este es viable. Tales presuposiciones involucran las condiciones bajo las cuales los propósitos o fines por los cuales uno participa en un juego pueden realizarse. Espero que un par de ejemplos ayuden a clarificar la noción del objetivo o valor de un juego de lenguaje.

Considere un juego de lenguaje que involucra bautizar a gente con nombres propios que se usan para identificarlos y re-identificarlos. Las reglas para este juego de lenguaje de la “identidad personal” explicitan que para que una persona P1 sea numéricamente una y la misma que una persona P2 debe ser el caso que P1 es tanto física como psicológicamente continuo con P2. Participar en este juego de lenguaje presupone ciertas formas en las cuales las personas interactúan, junto con los valores que se adjudican a tales interacciones. Pero estas interacciones tienen los valores que se les imputan solo en cierta clase de circunstancias empíricas. Para responder la fructífera pregunta de Wittgenstein de por qué se nombra a la gente pero no a las piezas de una cuchillería uno debe apreciar la significativa diferencia entre la manera en que la gente se relaciona con otra gente y la manera en que lo hace con los tenedores (cuchillos, etc.), diferencias que están bien separadas del hecho de que las personas, a diferencia de los tenedores, responden cuando se les llama por su nombre. Para todas las intenciones y propósitos humanos no hay nada que elegir entre un tenedor del juego y otro. Como reza el viejo refrán, todos parecen iguales (sea cual sea el valor estético que tengan, este es común a todos), y cumplen la misma función de una manera igualmente eficiente y, de esta forma, son completamente intercambiables. Uno se refiere a estos con el término *Buberiano* “esos” [*its*].

Cuán diferente es con las personas. Estas se merecen ser individuadas y rastreadas a través del tiempo, pues son nuestros *tús* [*thous*] —nuestros amigos, amantes, compatriotas, colegas, enemigos, y similares. La canción aparentemente tautológica de Mr. Roger “Solo tú eres tú” en verdad enfatiza la fascinante tautología que, como una cuestión de orden empírico, cada persona es única y, más aún, única en formas que son humanamente relevantes. Cada persona materializa de manera única

algún conjunto de propiedades. Incluso cuando dos personas comparten alguna propiedad genérica o propiedad determinable, usualmente tienen alguna especie diferente o materialización de ésta; uno tiene *Oidad*⁴⁵⁹ y la otra *Oidad*³⁵⁸. Nuestro lenguaje no es suficientemente rico para especificar estas sutiles pero importantes diferencias, pero somos capaces de intuirlos: le dan a cada persona un aura especial. Es esta unicidad humanamente importante lo que hace a la gente especial. Solo María González *María Gonzal-iza* [énfasis del traductor]. Si Ud. quiere experimentar ese conjunto único de propiedades, Ud. debe ir donde esa persona de carne y hueso. Es esto lo que explica nuestro apego a la carne y la importancia que se le otorga a la continuidad espacio-temporal del cuerpo en la reidentificación de las personas. Las propiedades que nos agradan de María González y que la hacen querer relacionarse con ella como una colega, compatriota, amante, y similares, son expresadas únicamente a través de ese cuerpo.

Las emociones muy fuertes y las actitudes del amor, de la adoración, respeto, y del afecto forman parte de nuestra participación en el juego de la identidad personal, pues implican las formas de vida implícitas a la participación en este juego. Estas formas de vida involucran el tipo especial de relación que las personas tienen entre sí y el tipo de obligaciones y deberes que generan. A través de estas relaciones uno se compromete y la vida se hace significativa. Tan seriamente tomamos a las otras personas que estamos dispuestos a conferirles algún tipo especial de dignidad al hacerlos moralmente responsables de sus acciones y, por tanto, invitarlos a ser recíprocos al hacer que ellos nos consideren responsables. Nuestro sentido de la dignidad como personas se basa en nuestro deseo de ser partícipes del juego moral de la responsabilidad.

Es una presuposición empírica para las formas de vida que informan al juego de la identidad personal el ser viable que las personas sean únicas en formas que tienen relevancia humana. En la literatura reciente acerca de la identidad personal numerosos experimentos mentales de ciencia ficción parecen proponer mundos contrafácticos

en los cuales nosotros poseemos la tecnología para duplicar personas a nuestra voluntad, intercambiar sus partes, reemplazarlos con personas artificiales, y así sucesivamente. A veces estos experimentos pretenden mostrar que estamos demasiado atrapados en nuestra carne, que hacemos demasiada alharaca respecto de la continuidad espacio-temporal, especialmente cuando hacemos de esta una condición necesaria de la identidad personal a través del tiempo. Se nos pregunta: “¿Por qué habría de importar si la mujer que vuelve de una estadía en el hospital es espacio-temporalmente continua con la amada esposa anterior, en vez de ser un clon que no puede distinguirse de ella y que satisface las mismas funciones?”

Lo viciado de estos experimentos mentales de ciencia ficción es que nos transportan junto con nuestros juegos actuales de lenguaje y sus formas de vida, a un mundo contrafáctico. Y la gente afirma que en tal mundo uno enfrenta contraejemplos a los análisis de la identidad personal o al menos a los casos equívocos. Tal gente no logra darse cuenta que en este mundo uno no querría jugar nuestro juego original de lenguaje de la identidad personal, debido a que no habría objetivo o utilidad en hacerlo: las presuposiciones empíricas para hacerlo no figuran. De este modo, en este mundo no enfrentamos contraejemplos o casos ambiguos, puesto que no existe la práctica normativamente gobernada por reglas de identificar o reidentificar a las personas. La pregunta de si la mujer que regresa del hospital es realmente mi esposa ni siquiera se plantea. Más aún, no tendría ninguna emoción involucrada en cómo se puede responder tal pregunta, pues la forma de vida que involucra la participación en el juego de lenguaje no tendría asidero alguno en ese mundo imaginado.

Wittgenstein, cuyas ideas estoy desarrollando aquí, lleva a cabo un experimento mental en el que concibe una situación contrafáctica en que no existiría objeto alguno para seguir participando en el juego de la identidad personal, pues la presuposición empírica para hacerlo está ausente:

Imagine que, por ejemplo, todos los cuerpos humanos que existen fuesen idénticos en apariencia, que, por otra parte, conjuntos distintos de rasgos pareciesen cambiar del cuerpo que habitan... Bajo tales circunstancias, aunque sería posible darle nombre a los cuerpos, tal vez estaríamos tan inclinados a hacerlo como lo somos con las sillas de nuestro juego de comedor. Por otra parte, sería útil darle nombres a los conjuntos de rasgos, y el uso de estos nombres correspondería *grosso modo* con los nombres propios de nuestro lenguaje (Wittgenstein 1958, pp. 61-62).

He estado tratando de explicar por qué no seríamos “proclives” a participar en este juego, por qué no sería “útil” hacerlo. En este caso imaginario, así como en el caso de arriba del ejemplo médico de ciencia ficción, probablemente reemplazaríamos nuestros juegos de identidad personal originales con uno en el cual diésemos nombres a conjuntos de características y funciones. No diríamos “Tráigame a Napoleón” sino “Tráigame un Napoleón” (mientras Ud. está de pie). El experimento mental de Wittgenstein, a diferencia de nuestro caso de ciencia ficción, no es viciado pues se introduce adecuadamente como la descripción de un mundo en el cual no participaríamos del juego de la identidad personal, ya que no habría ningún objeto en hacerlo, no siendo la forma de vida fomentada por el juego de lenguaje original viable o deseable en aquellas circunstancias empíricas.

Otro interesante caso para mi distinción entre las reglas de un juego de lenguaje y las presuposiciones empíricas para participar en el juego es que el criterio de identidad de las así denominadas obras de arte autográficas, esto es, las obras de arte que otorgan la posibilidad de una falsificación o de una copia de un trabajo específico. (Es importante distinguir la imitación de un trabajo artístico específico de un caso de un trabajo artístico fraudulentamente publicitado o

puesto en escena: fue consignada la puesta en escena de las sonatas de Beethoven por Dame Myra Hess pero fue tocada por un travestido Liberace; afirmó que era una copia autografiada del manuscrito de Dickens *Una Historia de Dos Ciudades*, pero era apócrifa). El criterio de identidad de una obra de arte autográfica es la continuidad espacio-temporal con el objeto creado por el artista. Parte de participar en el juego de lenguaje de la identidad del arte autográfico involucra que tenemos diferentes creencias, actitudes, emociones acerca del original, que no tenemos de la copia. Las mismísimas reglas del juego de lenguaje requieren que hagamos una odiosa [*invidious*] distinción estética entre el original y la copia. Después de todo el original es una obra auténtica y la copia no lo es —una importante diferencia estética, indudablemente. En general, apreciamos el original de maneras en que no lo hacemos para la copia, tal como apreciamos a alguien que es espacial y temporalmente continuo con mi esposa anterior de maneras en que no lo hacemos con alguien que no lo es, por ejemplo, un clon.

Tan pronto se genera una explicación de nuestro juego de lenguaje de identidad del arte autográfico, el Hombre Contraejemplo se asoma a la ventana —una que quedó abierta de modo conveniente desde el último film de Drácula— y nos aporrea con esta especie de experimento mental de ciencia ficción:

Imagine que poseemos una máquina de duplicación super-duper. Podríamos poner *La Mona Lisa* en la máquina y reproducirla molécula por molécula, tantas copias como uno deseara. Ciertamente, en una situación tal, sería un caso de sentimentalismo mal entendido seguir insistiendo en la continuidad espacio-temporal como una condición necesaria para que la pintura sea (o una ejemplificación de) *La Mona Lisa*.

Nuevamente, tenemos un experimento mental mal planteado cuyo resultado es bastante diferente del deseado. En vez de describir un

mundo posible en el cual nuestro requisito espacio-temporal actual de continuidad para la identidad de una pintura enfrenta un contraejemplo o un caso ambiguo, nos da un mundo en el cual no participaríamos en el juego de lenguaje de la identidad del arte autográfico. Ello se explica porque la presuposición empírica para hacerlo está ausente. Nuestro requisito actual de la continuidad espacio-temporal para la identidad de una pintura se basa en el hecho contingente de que las pinturas son únicas en formas que son estéticamente importantes y no duplicables. Sí. Ud. desea ver el conjunto especial de propiedades de *La Mona Lisa* y tener la experiencia tipo única de *La Mona Lisa* debe ir al retrato colgado allá, tal como Ud. debe ir al montón de carne y hueso de allá si quiere tener la experiencia única tipo de María González. Otra vez, se debe enfatizar que el experimento mental de arriba no sería perjudicial, sino para servir un propósito útil, si este se introdujese para mostrar un mundo posible en el cual no participásemos más en el juego de lenguaje original.

Dado que, como cuestión de hecho, vivimos en un mundo en el cual las pinturas y la gente son únicas y no duplicables, estamos justificados en preocuparnos tanto como lo hacemos por las pinturas y la carne, así como en jugar los juegos de lenguaje en que participamos.

Referencias

- Rescher, N. (1973): *The Primacy of Practice*. Oxford: Basil Blackwell.
- Wittgenstein, L. (1958): *The Blue and the Brown Books*. New York: Harper and Brothers.

¿Pueden fallar los experimentos mentales?³²

Allen I. Janis

Debido a que no todo el mundo está de acuerdo con respecto a qué es un experimento mental, comenzaré diciendo a qué me refiero con la expresión. En primer lugar, me restringiré al contexto de la física. Tal vez las ideas presentadas aquí pueden aplicarse a otras áreas, pero dejaré este juicio a quienes las conocen mejor. Hay dos características de un experimento mental que serán relevantes para la discusión que sigue, y que tomo como características definitorias. Primero, un experimento mental es la descripción de un procedimiento experimental así como de su resultado o posibles resultados. (Al referirme a un procedimiento experimental, no quiero decir que el experimento podría en efecto llevarse a cabo. Las razones de esto pueden ir desde dificultades de orden técnico, que podrían en principio superarse, a dificultades de principio; por ejemplo, puede ser deseable suponer en un experimento mental que el universo fuese totalmente distinto de la manera en que realmente es). Segundo, el resultado o posibles resultados deben ser deducidos mediante razonamiento consistente con un marco teórico dado.

Por ejemplo, citaré la conocida modificación de Bohm del experimento mental Einstein-Podolsky-Rosen (EPR).³³ Uno se

³² “Can thought experiments fail?” En: T. Horowitz y G. Massey (1991, eds.) *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield, pp. 113-118. Traducido y reimpresso con permiso de su autor.

³³ Puede encontrarse una discusión más acabada de este experimento mental en Bohm (1951).

imagina una fuente que emite dos partículas, cada una con un espín [*spin*] de un medio (en las unidades métricas Standard de la constante Planck dividida por 2π), en un estado cuyo *momentum* angular total y su *momentum* angular orbital son ambos cero. Las partículas viajan en ambas direcciones, de manera tal que simplemente con esperar lo suficiente éstas pueden estar tan separadas como uno desee. Uno entonces imagina que se mide el componente de la carga de cada una de las partículas en una dirección dada, por ejemplo, la dirección Z. De acuerdo con las leyes de la Mecánica Cuántica, estas medidas mostrarán que el componente Z del espín de una de las partículas será $+1/2$, y que el de la otra será $-1/2$, a pesar de la dirección que se elija como la dirección Z.

Comentaré más abajo el significado del experimento EPR. Mi objetivo aquí es sólo ilustrar las condiciones que exijo de los experimentos mentales. Este ejemplo satisface las dos condiciones que he señalado. Primero, es una descripción de un experimento y de sus resultados posibles; y, segundo, los posibles resultados son deducibles desde el marco teórico de la Mecánica Cuántica, cuya exploración es sin duda el objetivo del experimento mental.

Es sabido que los experimentos reales a veces fallan. El problema que quiero explorar en este trabajo es si los experimentos mentales pueden fallar también. Como cuestión preliminar, distinguiré tres sentidos distintos en los cuales se dice que falla un experimento real.

Primero, un experimento puede fallar porque no puede ser ejecutado hasta el final. Por ejemplo, podría existir un fallo del equipo. O tal vez puede existir alguna influencia externa que interfiera con el funcionamiento adecuado del experimento (o del ejecutante del experimento). En este sentido de fallar, el experimento (o intento de experimento) simplemente no logra producir resultado alguno. ¡Ejemplos de este tipo de fallo abundan, pero nunca, o casi nunca, son publicados!

Segundo, un experimento puede fallar cuando da un resultado

que es incorrecto. Esto también podría ser causado por fallas en los equipos (probablemente no sean detectadas en esta situación). O tal vez el experimento fue mal diseñado de modo que los resultados obtenidos son alterados por influencias que escapan a nuestro control. Aunque se obtiene algún resultado definido en este tipo de fallo, el resultado no es una representación adecuada de lo que el experimento debería haber mostrado. Un ejemplo conocido de este tipo de fallo es dado por las mediciones de Kaufmann, a principios del siglo XX, de la dependencia de la masa del electrón sobre la velocidad. Los resultados obtenidos por Kaufmann eran inconsistentes con la Teoría Especial de la Relatividad, o al menos parecían serlo. No obstante, Planck examinó los experimentos y concluyó que los valores que Kaufmann había asumido para los campos eléctricos en su equipo [*apparatus*] eran incorrectos debido a la ionización producida por los residuos de gas que escapaban de dicho equipo. El re-análisis de Planck de los datos apoyaba las predicciones de la Teoría de la Relatividad.³⁴

Tercero, un experimento puede fallar cuando produce resultados que, aunque correctos, no logran responder las preguntas que motivaron su ejecución. Un ejemplo de este tipo de fallo es el de los experimentos que intentaron establecer si el protón era estable. Estos experimentos no produjeron evidencia concreta del decaimiento del protón, pero mostraron que si lo hubiese, su tiempo de duración era probablemente de al menos 10^{32} años.³⁵ En la época en que estos experimentos comenzaron a realizarse, sin embargo, las teorías entonces en boga indicaban un período de vida más corto que hubiera sido detectable mediante ellos. Aunque generalmente se acepta que estos experimentos dan un resultado correcto, no habían logrado establecer la supuesta inestabilidad del protón.

Uno podría argüir que este tercer sentido de fallo no es en absoluto

³⁴ Los experimentos de Kaufmann y el análisis que Planck hizo de ellos son descritos en Miller (1981), por ejemplo.

³⁵ Se hace una revisión de esos experimentos en LoSecco (1985).

un fallo, ya que se obtienen resultados correctos de todas formas. Si, en mi ejemplo, la motivación del experimentador hubiese sido fijar un nivel inferior de duración, entonces el experimento no se habría ajustado a mi definición de fallo en el tercer sentido. Establecer esta duración inferior, no obstante, no arroja luz sobre la pregunta de si el protón decae, y así al menos algunas partes interesadas habrían sostenido indudablemente que en un sentido el experimento hubiera fallado. Aunque concedo que uno puede razonablemente considerar que experimentos de este tipo son éxitos en vez de fracasos, continuaré refiriéndome a estos como fracasos en el tercer sentido (al menos para propósitos de este análisis).

Refinaré mi definición de fallo en el tercer sentido mediante la comparación de la búsqueda del decaimiento del protón con el examen de Millikan a la ecuación fotoeléctrica de Einstein.³⁶ Millikan emprendió este trabajo pensando que Einstein estaba equivocado en sus posturas respecto a la naturaleza de la luz, y esperaba que el experimento refutaría el análisis de Einstein. De hecho, tales suposiciones estaban absolutamente de acuerdo sobre las predicciones de Einstein. En consecuencia, estos experimentos no cumplían con los objetivos que habían motivado originalmente su realización. La distinción que haría entre los experimentos de Millikan y aquellos que intentaron medir el decaimiento del protón es que los resultados de los experimentos de Millikan lograban establecer la corrección de la ecuación fotoeléctrica de Einstein, mientras que la única conclusión que puede inferirse de los experimentos del protón es que, *si* el protón decae, su duración debe ser mayor que un valor específico. Por lo tanto, si se piensa un experimento para testear ciertas ideas y de éste se puede concluir (al menos en la medida que se puede concluir algo de los experimentos) que aquellas ideas son verdaderas o falsas, entonces, a pesar de las expectativas del experimentador o de las motivaciones que podrían haberlo llevado a su ejecución, consideraría

³⁶ Por ejemplo, puede encontrarse una presentación del trabajo de Millikan sobre el efecto fotoeléctrico y sus creencias sobre la naturaleza de la luz en Kargon (1982).

que el experimento ha sido un éxito. Por otra parte, si el experimento está pensado para probar la existencia de cierto efecto y termina fijando ciertos límites (más allá del rango dado por el experimento) respecto del tamaño del efecto, consideraría que el experimento ha fallado en el tercer sentido.

Mi opinión [*contention*] es que los experimentos mentales pueden también fallar en cada una de estas tres maneras. Las consideraré una a una.

Un experimento mental puede fallar en el primer sentido si el experimentador (a quien de aquí en adelante me refero como EM) carece del conocimiento o de la habilidad para llevar a cabo el análisis hasta llegar a la conclusión. Tal como en el caso de los experimentos reales, esto puede ocurrir por equipo deficiente (esto es, capacidad intelectual [*brain power*] inadecuada o falta de entrenamiento), o por influencias externas (por ejemplo, un accidente que produzca impedimento, o por la influencia del alcohol o las drogas). En este tipo de fallo no hay resultado alguno que emerja del proceso de pensar el experimento. Sin embargo, tal como ocurre análogamente con un experimento real, puede tener sentido afirmar que ha habido un experimento mental que falló, en vez de decir que no ha habido ningún experimento mental. El experimento mental surge entonces cuando EM piensa que el análisis conceptual en conjunto con algunas suposiciones lo conducirá a ciertas clarificaciones o conclusiones; el experimento comienza cuando EM empieza con este análisis y falla cuando EM es incapaz de llevar el experimento a una conclusión. Tal como en el caso del fallo de un experimento real, sospecho que abundan los experimentos mentales fallidos en este sentido, pero una rara vez sabe de ellos.

Un experimento mental puede fallar en el segundo sentido si EM llega a una conclusión incorrecta. Un experimento mental famoso en este sentido es el propuesto por Einstein en la Conferencia Solvay en los años 30. Einstein describe un aparato que, en principio, parecería

permitir medidas que estarían en conflicto con la manera en que el principio de incertidumbre relaciona la energía y el tiempo. Bohr, sin embargo, encontró que el análisis de Einstein era incorrecto. Irónicamente, fue un aspecto de la Teoría General de la Relatividad lo que Einstein había soslayado; cuando tal teoría se tomaba en consideración, el análisis era en efecto consistente con el principio de incertidumbre.³⁷

Un experimento mental puede fallar en el tercer sentido cuando su análisis lleva a una solución correcta, pero no a una que satisfaga el objetivo que motivó su realización. Argumentaré que el experimento mental EPR es un ejemplo de falla en este sentido, pero para hacerlo es necesario hacer algunos comentarios adicionales acerca del experimento.

En función de lo que ya se ha explicado acerca del experimento EPR, resulta claro que si se toma una medida del componente Z del espín de una de las partículas separadas, entonces el componente Z del espín de la otra partícula puede predecirse con certeza. Del mismo modo, si el componente X del espín de la primera partícula se mide en vez del componente Z, entonces el componente X del espín de la segunda partícula podría predecirse con certeza. El argumento EPR³⁸ consistía en que las medidas del espín de la primera partícula no influyen en la segunda partícula de ninguna manera, y en que si un atributo físico de un sistema puede predecirse con exactitud sin interferir en el sistema, entonces tal atributo corresponde a un elemento de la realidad física. Más aún, EPR mostraba que, debido a que el componente X o el Z de la carga de la segunda partícula pueden ser anticipados con certeza, dependiendo de cuál medida

³⁷ Una explicación de este experimento mental y de la respuesta de Bohr a él se da en Bohr (1951).

³⁸ Además de la discusión en Bohm (1951), citada más arriba, véase también el ensayo original de Einstein, Podolsky y Rosen (1935). Aunque este ensayo se refiere a una forma distinta de este experimento mental, las ideas que se discuten en EPR en conexión con el experimento se aplican a ambas versiones.

uno elija para aplicar en la primera partícula, ambos componentes de la carga de la segunda partícula corresponden a elementos de la realidad física. Debido a que según la interpretación convencional de la Mecánica Cuántica (Copenhague), estos componentes de carga no pueden simultáneamente tener valores definidos, EPR mostró como conclusión que la descripción de la realidad dada por la Mecánica Cuántica es incompleta.

En este punto, me gustaría separar entre el experimento mental mismo y las conclusiones que uno podría obtener de los resultados del experimento mental. Me parece que en este ejemplo la separación debería venir justo después de que uno deduce cuáles serían los resultados de las mediciones del componente de la carga; las conclusiones acerca de la naturaleza de la Mecánica Cuántica están del otro lado de la separación. Entonces parecería ser indudable que los resultados del experimento mental son correctos. Pero no ha habido consenso general de que las conclusiones del EPR acerca de la naturaleza de la Mecánica Cuántica sean correctas.³⁹ Así, el experimento mental no logró mostrar un fundamento sólido para concluir que la Mecánica Cuántica es incompleta. Sin embargo, en vista de que este objetivo era la motivación para llevar a cabo el experimento mental, este falló en el tercer sentido.

Explicaré por qué considero que el experimento mental EPR es más parecido a los experimentos del decaimiento del protón, que señalé como un ejemplo de fallo de experimentos reales en el tercer sentido, que a los experimentos fotoeléctricos de Millikan, el cual consideré como un experimento real exitoso. Para ser análogo al experimento de Millikan, un experimento mental tendría que llevar a la conclusión de que ciertas ideas son correctas, independientemente de las expectativas del experimentador. El experimento EPR, no obstante, no muestra que la Mecánica Cuántica es completa, sino que

³⁹ Bohr (1935a) escribió casi inmediatamente una breve nota en *Nature* en la que expuso la idea básica de su respuesta al análisis de EPR, y prontamente lo siguió con un análisis más detallado en *The Physical Review* (Bohr 1935b).

simplemente no concluye que la Mecánica Cuántica es incompleta. Tal como los experimentos del decaimiento del protón, el experimento EPR no muestra ni la verdad ni la falsedad de las ideas que motivaron su realización.

Me parece que el tercer sentido de fallo, tanto para los experimentos reales como para los mentales, es el más interesante. Como ya he mencionado, los fallos en el primer sentido rara vez son publicados. Aunque los fallos en el segundo sentido a veces son publicados, es inusual que los errores cometidos sean suficientemente interesantes para que los experimentos sean dignos de discusión una vez que estos errores han sido descubiertos. Sin embargo, el tercer sentido de fallar a menudo lleva a un entendimiento más cabal de la situación que motivó la realización del experimento. En el ejemplo que di para el tercer sentido de fallo de un experimento real, los argumentos teóricos que llevaban a la predicción del decaimiento del protón fueron refinados enseguida para otorgar una mejor comprensión de los procesos involucrados (y, por supuesto, para dar un tiempo de decaimiento consistente con los límites experimentales); no obstante, no está claro si lo que surgió de allí es la comprensión adecuada, pues la tesis del decaimiento del protón aún no ha sido corroborada experimentalmente. Y en mi ejemplo del tercer sentido en el que un experimento mental puede fallar, los argumentos que el experimento EPR originó acerca de la naturaleza de la Mecánica Cuántica todavía están en discusión. Sospecho que muchos otros experimentos mentales fallidos en el tercer sentido podrían requerirse todavía antes de que se llegue a una comprensión cabal de la naturaleza de la Mecánica Cuántica.

Referencias

- Bohm, D. (1951): *Quantum Theory*. New York: Prentice-Hall.
- Bohr, N. (1935a): “Quantum mechanics and physical reality”. *Nature* vol. 136, 65.
- _____ (1935b): “Can quantum-mechanical description of physical reality be considered complete?” *Physical Review* vol. 48, 696-702.
- _____ (1951): “Discussion with Einstein on epistemological problems in atomic physics”. En: P. Schilpp (ed.) *Albert Einstein: Philosopher-Scientist*. New York: Tudor Publishing Co.
- Einstein, A., Podolsky, B. y Rosen, N. (1935): “Can quantum-mechanical description of physical reality be considered complete?” *Physical Review* vol. 47, 777-780.
- Kargon, R. (1982): *The Rise of Robert Millikan*. Ithaca, NY: Cornell University Press.
- LoSecco, J. (1985): “Status report on the search for proton decay”. *Comments on Nuclear and Particle Physics* vol. 15, 23-34.
- Miller, A. (1981): *Albert Einstein’s Special Theory of Relativity*. Reading, MA: Addison-Wesley Publishing Co.

Analiticidad por la puerta trasera⁴⁰

Gerald J. Massey

1. La pareja dispareja

La filosofía debería haber aprendido una lección de lo que le pasó a su vieja camarada la teología. Cuando se unen fuerzas con un jugador más poderoso, uno puede quedar botado al borde del campo de juego llevándoles agua a los jugadores y alentándolos.

No hace mucho tiempo atrás la filosofía y la ciencia formaban un equipo. En los días de oro la filosofía, avezada y sofisticada, contribuía de manera importante con la tarea de aunar la empresa teórica de la ciencia, en ese entonces relativamente inmadura e inexperimentada. Pero en la medida en que el juego de comprender el mundo pasaba, la ciencia se convirtió en la jugadora estrella. Cada vez más frecuentemente la filosofía se encontró a sí misma sola en el borde del campo o del todo abandonada, una situación patética y bochornosa.

Si el ascenso de la ciencia moderna hizo tambalear a la filosofía, el advenimiento de la ciencia contemporánea la ha dejado totalmente desmoralizada. En el mismísimo amanecer de la revolución Newtoniana, Hume exhortó a sus camaradas investigadores a mandar a la hoguera a cualquier libro que no estuviese repleto de razonamiento

⁴⁰ “Backdoor Analyticity”. En: T. Horowitz y G. Massey (1991, eds.): *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield, pp. 149-166. Traducido y reimpresso con permiso de su autor.

experimental acerca de cuestiones de hecho, o de razonamiento demostrativo acerca de números o cantidades (Hume 1981, p. 114). Afortunadamente, nadie escuchó la reprimenda sediciosa que decía que había que dedicarse sólo a trabajos de ciencia empírica, o uno no disfrutaría hoy el libro en el que esto se publicó.

Una ansiedad visceral de que su entrañable disciplina no jugaba más ningún papel sustantivo obsesionó a muchos filósofos del siglo XX. Lo *a priori* había sido desde hace tiempo la provincia de la filosofía, pero se encontraba en tiempos difíciles (e.g. como consecuencia del desarrollo de las geometrías no-Euclideanas), o había sido reemplazada por las matemáticas y por la nueva lógica matemática, que prosiguió a la independencia natural propia de la madurez. Los filósofos enfrentaron una tarea alternativa: o renunciar a las pretensiones cognitivas de su empresa o encontrar alguna ruta conceptual hacia lo *a priori* que todavía no hubiera sido expropiada por la ciencia. Pronto se convencieron de que en el análisis del significado habían descubierto una ruta privilegiada a lo *a priori*. El análisis del significado, a través del examen analítico del discurso o mediante la construcción de explicaciones Carnapianas, les daba esperanza y les ofrecía a los filósofos la promesa de respeto que da el trabajo productivo de cualquier naturaleza, incluso el intelectual.

Aunque el análisis del significado podría considerarse como productor de verdades *a priori*, corrió un serio riesgo: la trivialidad. Casi sin excepción los grandes filósofos del pasado habían manifestado su abierto desprecio por perogrulladas como “No puede haber injusticia sin propiedad” (Hume) o “Los triángulos tienen tres ángulos” (Descartes), verdades demasiado representativas del tipo al que se llegaba mediante análisis del significado. ¿Podría este haber generado algo que no fuera de Perogrullo? Y si fuese así: ¿Podría haber producido algo de interés filosófico?

La historia reciente, o al menos una manera de leer lo que fue la historia reciente, sugirió una respuesta afirmativa a ambas preguntas.

¿No era la teoría de Russell de las descripciones definidas un paradigma del análisis del significado? Ciertamente, dicha teoría no era trivial, por simple que pudiese parecer en retrospectiva. Ni fue tampoco una teoría que borró las oscuras entidades Meignonianas de la ontología filosóficamente infértil. Aquí, entonces, había un logro semántico que les dio a los filósofos confianza intelectual, incluso en el incansable imperialismo cognitivo de la ciencia contemporánea.

Detrás del análisis del significado yace la analiticidad. El producto final del análisis del significado tiene que ser verdades analíticas, proposiciones verdaderas de una manera u otra en virtud de su significado, como sea que el significado mismo se entienda o se explique. Y ahí, como se verá, yace la complicación.

2. Dos Dogmas del Empirismo

Sin duda todos nos quejamos alguna vez de las fechas que los profesores de historia nos hacían memorizar: 313 C.E. Edicto de Licinio y Constantino de Milán; 1066 Batalla de Hastings; 1492 Colón descubre el nuevo mundo, etc. Y aun así nada tiene más impacto en nuestro sentido de la historia que las pocas fechas que se consideran memorables. Por ejemplo, 1950 es un año cargado de significado histórico para la mayoría de los filósofos analíticos. ¿Por qué? ¿Porque Carnap redescubrió la pragmática? ¡No! Es porque en este año, como todo el mundo pretende saber, Quine *demonstró* en “Dos Dogmas del Empirismo” que la analiticidad es una noción quimérica (Quine 1964).

Declaro que Quine demostró el falso carácter de la analiticidad de la misma manera que Colón descubrió el nuevo mundo, ya que los resultados reales del navegante genovés tienen sólo una vaga semejanza al gran logro con que la historia lo ha coronado. Lo mismo sucede con Quine. De entre las legiones de filósofos, casi nadie que esté ávido de atribuirle a Quine haber desacreditado la analiticidad puede decirnos cuáles eran las tesis de “Dos Dogmas”, mucho menos cuáles eran los argumentos de Quine para estas (si en realidad ofreció

alguno), y si dichos argumentos tenían alguna fuerza lógica. Sin embargo, existe virtual unanimidad en este punto: en 1950 Quine derrumbó exitosamente la analiticidad.

¿Qué hizo en efecto Quine? Bueno, propuso algunas razones para creer que la analiticidad en un sentido amplio (la propiedad poseída por las oraciones que se hacen verdades lógicas o falsedades lógicas mediante la substitución de sinónimos en las expresiones constituyentes) pertenece a una familia de conceptos cuyos miembros tienden a impresionar a gente como él como confusos e ininteligibles. Quine también propuso razones que sugieren que la distinción analítico-sintética no sería clara sino confusa, incluso si uno pudiese tener éxito en darle sentido a la noción de analiticidad. Y, lo más influyente de todo, pintó un escenario pragmático del significado (la red de oraciones y las metáforas de los campos de fuerza), hablándole a un profundo anhelo de la psique filosófica de habla inglesa. Quine también dijo algo en “Dos Dogmas” acerca de la reducción, pero lo voy a omitir aquí.

Junte la suficiente cantidad de gente para creer que algo es de una manera y Ud. casi lo hará ser de esa manera. Suponga que Quine hubiera de hecho demostrado la incoherencia o la naturaleza falaz de la analiticidad. ¿Cómo habrían reaccionado los filósofos responsables a tal demostración? Primero, habrían ciertamente dejado de lado el concepto y la distinción existente en sus ponencias y publicaciones profesionales, solemnemente citando la prueba de Quine de la ininteligibilidad e implausibilidad. Segundo, habrían adoptado, si hubiese estado disponible, o habrían desarrollado, si no hubiese estado disponible, una concepción o teoría del significado que no presupusiera la malograda noción o que incorporase la estropeada distinción. Y, más piamente que nada, hubieran renunciado a todos los métodos filosóficos que presupusieran el concepto de la analiticidad. En aquellos casos donde los métodos descartados parecían indispensables a la filosofía, hubieran encontrado substitutos, métodos reemplazantes que no dependiesen de la noción de analiticidad pero que, no obstante,

hicieran el mismo trabajo a los que se abocaban los viejos métodos.

Es una cuestión de registro histórico que los filósofos respondieron a “Dos Dogmas” de todas estas formas. (Con más certeza, ellos respondieron a “Dos Dogmas” reforzados por las doctrinas de *Word and Object*, más notablemente mediante la tesis de la indeterminación de la traducción de Quine 1960, pp. 26-79). Los filósofos profesionales que estiman su reputación raramente pasan por alto la oportunidad de repudiar la noción de analiticidad o una distinción clara analítico-sintética por medio de un ritual de genuflexión hacia Quine. Ni tampoco dudan en proclamar su alianza con la doctrina del significado que barre la analiticidad y la consiguiente distinción analítico-sintética, a menudo una similar al seductor modelo de las redes de oraciones esbozado en “Dos Dogmas”, y una no menos impresionante.

Debido a que el análisis del significado en sus variados ropajes presupone la analiticidad, los filósofos abandonaron el análisis del significado como un método filosófico legítimo, y buscaron reemplazantes. Pero los métodos filosóficos no vienen de a docenas por centavos; imaginarse incluso un método de esa índole asegura al filósofo un lugar en el panteón de los grandes de la filosofía. ¿Qué método o métodos podrían suplantar entonces el análisis del significado? Mi respuesta no es tan controvertida: el método de los *experimentos mentales* o, como podrían llamárseles también, la apelación a los argumentos de lo concebible. ¿Presupone el método de los experimentos mentales la analiticidad? Mi respuesta es por demás discutible: presuponen la noción de concebibilidad, que es en el fondo indiscernible de la analiticidad.

3. Concebibilidad

Argumentaré, entonces, que mientras que la analiticidad fue echada por la gran puerta de la filosofía, con gran pompa y ceremonia, fue readmitida sigilosa y serenamente por la puerta trasera. Se fue por la puerta delantera como analiticidad; reingresó por la puerta trasera como concebibilidad.

La concebibilidad no es una forastera en la filosofía, ni una noción que sea una primicia del post-empirismo. El concepto de concebibilidad, y el género de argumentos y métodos a los cuales dio origen, se remontan al albor de la filosofía. Sin embargo, fue Descartes quien trajo la concebibilidad al centro del escenario de una forma notoriamente poco crítica y poco rigurosa. A no dudarlo cada época filosófica tiene conceptos y métodos poco rigurosos, que quedan sin examen crítico, y que le dan a la era un tinte particular. En tanto que exista “progreso” filosófico, esto es, en la medida que exista un sentido colectivo de movimiento al progreso o de desarrollo [*momentum*] cognitivo, los conceptos y métodos permanecen sin examen crítico, pues ¿qué filósofo sería tan ruin de cuestionar a un benefactor tan generoso y resuelto? El método carente de sentido crítico del período que se denomina *Filosofía Moderna* es el de los experimentos mentales. El concepto que permanece sin examen es el de la concebibilidad.

Me permitiré explicar ahora cómo propongo argumentar a favor de la tesis de que en el nivel basal la analiticidad y la concebibilidad son la misma noción, y a favor del corolario metodológico de que el análisis del significado y la experimentación mental son al final la misma cosa. Mi estrategia será parecida a la adoptada por David Hume para probar que el sentimiento del altruismo o la fraternidad es el mismo que el sentimiento moral, a saber:

Las mismas dotes de la mente, en toda circunstancia, son aceptables para el sentir de lo moral y del altruismo; la misma disposición se pone hacia un sentimiento y hacia el otro; y el mismo cambio en los objetos por la sola aproximación o por las conexiones, insufla el uno y el otro. Por todas las reglas de la filosofía, entonces, se debe concluir que estos sentimientos son originalmente el mismo; puesto que, en cada

situación, incluso la más nimia, son gobernados por las mismas leyes, y son movidos por los mismos objetos (Hume 1983, p. 52).

Imitando a Hume, entonces, trataré de mostrar que la analiticidad y la concebibilidad son gobernados por las mismas leyes, movidos por los mismos objetos, y que son en el nivel basal la misma noción.

Las conexiones de la analiticidad con lo *a priori* y lo necesario son lo suficientemente conocidas como para eximirme de comentarlas. El enigma de la necesidad de las tesis evidentemente sintéticas de la geometría euclidiana felizmente mostraron ser un pseudo-problema: el descubrimiento de las geometrías no euclidianas hizo patente que la necesidad era sólo aparente, y no real. Pero aquellos filósofos post-kantianos que identificaron lo *a priori* con lo analítico tuvieron que lidiar con el estatus de la aritmética. Era relevante para ellos mostrar que la aritmética es analítica, en tanto *a priori*, o era *a posteriori*, en tanto sintética. La mayoría de ellos optaron por el primer cuerno de la disyunción, eventualmente sosteniendo la reducción de la aritmética a la lógica pura, cuyas leyes se consideran analíticas por cortesía o, tal vez mejor, por lo que Descartes llamaría “eminencia”. Incluso en “Dos Dogmas”, como se recordará, Quine no desafía la noción de analiticidad en un sentido estrecho, i.e., en cuanto verdad lógica, cuando él las emprende en contra de la ininteligibilidad del sentido amplio de analiticidad, que en sí presupone la verdad lógica.

Dos puntos me preocupan aquí. El primero es la relación de la analiticidad con la verdad lógica (y con la falsedad lógica). La misma relación, sostengo, ocurre entre la concebibilidad y la posibilidad lógica. El tópico de que lo que es lógicamente imposible es inconcebible, o por contraposición, que cualquier cosa que es concebible es lógicamente posible, resuena a través de la filosofía moderna como el ritmo constante de un tambor. Hume, el pensador cuya teoría de la concebibilidad veo como representativa de las doctrinas de muchos filósofos modernos y contemporáneos, trata la cuestión de la siguiente forma:

Esto es una máxima consolidada en la metafísica. *Que cualquier cosa que la mente conciba de manera clara incluye la idea de la existencia posible*, o en otras palabras, *que nada que uno imagine es absolutamente imposible*. Uno puede formar la idea de una montaña de oro, y de aquí concluir que tal montaña podría en efecto existir. No se puede concebir la idea de una montaña sin un valle, y por lo tanto considérese esto imposible (Hume 1960, p. 32).

Unas páginas más adelante, Hume continúa esta línea de argumentación con un pasaje claro e inequívoco:

Es vano buscar una contradicción en una cosa que es concebida claramente por la mente. Si implicase una contradicción, es imposible que pudiese haber sido concebida (Hume 1960, p. 43).

Tal como lo lógicamente verdadero o falso es analítico y lo analítico es lógicamente verdadero o falso, así también lo concebible es lógicamente posible y lo inconcebible lógicamente imposible.

Hubo, por supuesto, gente que no estuvo de acuerdo. Entre estos destaca Descartes, cuya multifacética teoría de la concebibilidad evade cualquier clasificación fácil. Para Descartes la concebibilidad, i.e., la concepción clara y distinta, implica la posibilidad, pues Dios puede crear cualquier cosa que uno conciba exactamente de la manera en que se la conciba, un principio que este filósofo necesita para proponer una distinción real entre mente y cuerpo en la Sexta Meditación. Pero la inconcebibilidad no implica imposibilidad, ya que los límites de la mente no pueden ser confundidos con los límites de lo ilimitado e infinito de los poderes de Dios. Escribe a Mersenne que: “En general se puede sostener que Dios puede hacer cualquier cosa que uno puede comprender (concebir) pero no que no pueda hacer aquello que no

podamos comprender”.⁴¹ Descartes está listo para contemplar a Dios haciendo la suma de 1 y 2 otra cosa que 3, o que Dios creara una montaña sin valle o incluso que Dios produjera un vacío, todo lo cual él estima inconcebible. Descartes nos advierte que, de lo inconcebible, de lo que *nuestras* mentes no pueden concebir, afirmemos piamente sólo que un ángel no lo puede crear, no que Dios no pueda hacerlo.⁴²

Dejando de lado las pintorescas ideas de Descartes, y sean lo que sean las leyes de la lógica, el papel regulador que juegan hacia la concebibilidad (las concepciones de uno no pueden violar las leyes de la lógica) es el mismo que el papel constitutivo que estas leyes desempeñan hacia la analiticidad (lo que es lógicamente verdadero o falso es analítico). O puesto al revés, el rol regulador de la analiticidad en la lógica (las leyes lógicas deben estar dentro de los límites de significado semántico) es igual que el rol constitutivo de la lógica hacia la concebibilidad (lo lógicamente inconsistente es inconcebible). Esto, entonces, es el primer paso a mi demostración cuasi-Humana de la identidad original de dos nociones aparentemente diversas.

Procedo ahora al segundo punto tocado más arriba, a saber, el acceso de lo *a priori* a la necesidad y a lo posible dado tanto por la analiticidad como por la concebibilidad. Si nuevamente uno considera que la teoría de Hume es ejemplar, se ve que la concebibilidad implica posibilidad y que la inconcebibilidad implica imposibilidad. Ahora bien, si algo califica como una cuestión *a priori*, la concebibilidad seguro que lo hace; esto es un punto sobre el cual incluso hay consenso entre los denominados filósofos modernos. Entonces, tal como la analiticidad, la concebibilidad posibilita acceso directo *a priori* a la modalidad, pero a lo posible y lo imposible en vez de a lo necesario y lo contingente.

Parece entonces que la concebibilidad proporciona, si no la piedra

⁴¹ Véase en Descartes (1981) la Carta a Mersenne del 15 de Abril de 1630.

⁴² Véase en Descartes (1981) la Carta a Beckman del 17 de Octubre de 1630. También su Carta a Arnauld del 29 de Julio de 1648.

filosofal, al menos una poderosa herramienta filosófica. Mediante la mera concepción de algo el filósofo puede mostrar su posibilidad, o mediante la consideración de la inconcebibilidad de algo el filósofo puede demostrar su imposibilidad. Aquí, entonces, hay un método de sillón [*armchair method*] que llega directamente al corazón metafísico de las cosas, un auténtico camino real hacia la sabiduría.

Pero, ¿es tan real ese camino? ¿Es realmente un juego de niños concebir situaciones, o detectar su inconcebibilidad? Parecería que sí, según Hume:

Nada, a primera vista, podría parecer más ilimitado que el pensamiento de un hombre, que no sólo escapa a todo poder humano y de la autoridad, y que incluso no está restringido por los límites de la naturaleza y de la realidad. Concebir monstruos y formar figuras y formas amorfas no cuesta a la imaginación más problema que concebir los objetos más naturales y familiares. Y mientras que el cuerpo está confinado a un planeta, sobre el cual reptamos lentamente y con dificultades, el pensamiento lo puede teletransportar a uno en instantes a las regiones más distantes del universo; o incluso más allá del universo, al caos total, donde se supone que aquél yace en la más absoluta confusión. Lo que no fue jamás visto u oído podría de todas formas concebirse; ni tampoco está más allá del poder del pensamiento, excepto lo que involucra la más absoluta contradicción (Hume 1981, p. 11).

Nótese por lo demás que uno puede hacer que el pasaje de Hume se convierta en una observación acerca del lenguaje y del significado, sin pérdida de elocuencia literaria o relevancia filosófica del siguiente modo:

Nada, a primera vista, podría parecer más ilimitado que el lenguaje, el cual no sólo escapa a todo poder humano y de la autoridad, y que incluso no está restringido por los límites de la naturaleza y de la realidad. Hablar de monstruos, y unir figuras y formas incongruentes, no cuesta al hablante más que decir cosas acerca de los objetos más naturales y familiares. Y mientras que el cuerpo está confinado a un planeta, sobre el cual se arrastra lentamente y con dificultades, el lenguaje lo puede tele-transportar a uno en instantes a las regiones más distantes del universo; o incluso más allá, al caos total, donde se supone que aquél yace en la más absoluta confusión. Lo que no fue jamás visto u oído podría de todas formas concebirse; ni tampoco está más allá del poder del lenguaje, excepto lo que involucra la más absoluta contradicción.

Pero como Hume mismo dirá en el siguiente pasaje, estas apariencias son engañosas. Lejos de ser libre e ilimitado, nuestro poder de concebir está bastante restringido:

Pero aunque el pensamiento parece poseer esta libertad ilimitada, encontraremos que, luego de un examen más prolijo, que está limitado en estrechos confines, y que todo este poder creativo de la mente no es otra cosa sino el de componer, transportar, argumentar y disminuir los materiales que nos proporcionan la experiencia y los sentidos. Cuando pensamos en una montaña de oro, sólo unimos dos ideas consistentes, *oro* y *montaña*, con las cuales estábamos previamente familiarizados. Podemos

concebir un caballo virtuoso, puesto que de nuestro propio sentimiento, podemos concebir la virtud; y esto lo podríamos unir a la figura o a la forma de un caballo, que nos es familiar. En síntesis, todos los materiales del pensamiento se derivan o de nuestro sentimiento interno o externo; la mezcla de la composición de éstos pertenece sólo a la mente y a la voluntad. O, para expresarme en lenguaje filosófico, todas nuestras ideas o nuestras percepciones tenues son copias de nuestras impresiones más vivaces (Hume 1981, p. 11).

Si hubiera espacio aquí, sería provechoso concentrarse en la observación de Hume, casualmente expresada en el siguiente pasaje de que las ideas complejas se forman mediante la composición de la unión de ideas “consistentes”, tales como *oro* y *montaña*. Esta observación involucra el reconocimiento tácito de Hume del rol regulador que juegan las leyes de la lógica en la concepción. Sería igualmente provechoso, pero del mismo modo poco económico, mostrar que la teoría de Descartes de la concepción es bastante similar a la de Hume. Descartes deja a la mente y la voluntad sólo la mezcla y composición de las materias primas del pensamiento, a saber, las ideas presentadas innatamente a la mente o producidas por ésta con ocasión de ciertas sensaciones corporales.

Pero nos hemos apartado del punto. Los pasajes citados no apuntan a la facilidad de concebir sino más bien a los estrechos límites en los cuales se debe desenvolver. Tal vez dentro de estos límites nada es más

fácil que concebir. Tal vez la agilidad conceptual de nuestras mentes habría impresionado incluso al poeta que escribió esta memorable copla:

Pienso que es ingenioso de la tortuga
Ser de tal laya para ser tan fértil (Nash 1931)

Uno de los primeros y más famosos experimentos mentales de la primera *Investigación* de Hume parece confirmar *la tesis de la concepción fácil*, a saber, el célebre caso del tono faltante del azul. ¿Se puede concebir que alguien se proponga a sí mismo la idea de este tono, alguien que haya disfrutado de larga y directa familiaridad con todos los tonos de todos los colores, excepto con este? Esta es una pregunta escasamente fútil [*idle*]. Una respuesta afirmativa parece subvertir la afirmación universal hecha una página antes de que “toda idea... es una copia de una impresión similar”, un supuesto reforzado por el displicente desafío de que “Aquellos que afirmasen, que esta posición no es universalmente verdadera y sin excepciones, tienen sólo un fácil método de refutar lo anterior; al producir esa idea que, en su opinión, no se deriva de esta fuente” (Hume 1981, pp. 11-12).

Tal vez la respuesta definitiva de Hume a la pregunta puede explicarse solo por su abyecto consentimiento a la tesis de la concepción fácil.⁴³ Hume aceptó a raja tabla lo que se espera ser la respuesta de otros y que sin duda fue su propia respuesta al experimento mental, a saber, que uno ciertamente podría concebir que una persona imaginase la idea del tono faltante del azul. Él, incluso, toma estas respuestas anticipadas “como una prueba de que las ideas simples no siempre son, en toda situación, derivadas de las impresiones correspondientes; aunque esta instancia es tan singular, es escasamente digna de observación, y no merece que, por ésta solamente, se altere nuestra

⁴³ Una mejor explicación del enigmático ejemplo de Hume puede darse al citar su fascinante teoría de las reglas generales. Postulo una explicación tal en mi “Hume’s Theory of General Rules”, una ponencia dada para la *International Hume Society* ofrecida en Marburg, República Federal Alemana, en agosto de 1988.

máxima general” (Hume 1981, p. 13). ¡Qué respuesta más enclenque para un desafío tan apasionante!

Tomar el caso del tono del azul faltante como un paradigma de experimentación mental es abrazar la tesis de la concepción fácil. Tal vez ninguna otra tesis es aceptada más ampliamente en la filosofía contemporánea. No obstante, su aceptación es usualmente implícita, pero se la detecta rápidamente. Para socavarla uno solo necesita recordar cómo los filósofos se aportillan las teorías entre sí.

El arte filosófico es el arte de generar contraejemplos. Las verdades filosóficas, si existe alguna, son afirmaciones que resisten cualquier contraejemplo propuesto. Si uno postula una tesis general, por ejemplo, que la intensión determina la extensión, uno inmediatamente comienza a pensar en cómo falsarla. ¿Cómo? Al concebir una situación imaginaria en la cual la tesis se cae, una situación como el ejemplo de Putnam de la Tierra Gemela en el cual el sentido de “agua” supuestamente no logra determinar su referencia (Putnam 1982, pp. 22-48). ¿Cómo sabe *uno* que ha tenido éxito? ¿Cómo. sabe *Ud.*, por lo demás, que ha tenido éxito? Bueno, si Ud. se siente satisfecho con su concepción, si le parece que se ha dado maña en concebir una situación tal, Ud. simplemente ha hecho su tarea. Y si a mí me parece que Ud. ha tenido éxito, tal vez porque aprovechando su tutela parece que ahora soy también capaz de concebirlo, me rindo. O abandono mi tesis o, como Hume, desecho su contraejemplo como muy esotérico para hacer daño. Hay, por supuesto, una tercera alternativa, una que sorprendentemente es raramente elegida, a saber, desechar su concepción como falsa y apócrifa. Volveré a esta alternativa más adelante.

¿Le suena familiar? Si no, le ruego a Ud. que recuerde la dialéctica de la filosofía, tal como fue practicada durante la edad de oro [*heyday*] del análisis del significado, cuando la analiticidad era todavía un concepto respetable. (Al proponer esta analogía espero que Ud. efectúe los cambios *mutatis mutandis*, algo que tal vez debería considerar como algo obvio).

La reflexión acerca de la manera en que ellos realizan sus experimentos mentales me lleva a pensar que la filosofía contemporánea ha decidido seguir el ejemplo de Hume en el caso del tono faltante del azul. Si es así, han elegido imitar lo peor de Hume. Tal vez ningún otro filósofo ha llevado a cabo sus experimentos mentales con el grado de precaución y sofisticación que Hume le confirió a los suyos, o al menos a algunos. Pero si queremos ver cómo es para un filósofo llevar a cabo un experimento mental de forma responsable, debemos ir al *Tratado*, en especial a la parte II (Sobre las Ideas de Espacio y Tiempo) del libro I (Sobre el Entendimiento), donde Hume trata con el problema de la divisibilidad infinita del espacio y la cuestión de si podría existir un vacío.

Respecto de estos problemas Hume llega a las respuestas contrarias enunciadas por Descartes. Habiendo identificado el cuerpo con la extensión y la extensión con el espacio tridimensional, Descartes anticipadamente concluye que un vacío es inconcebible. Si Ud. logra concebir que toda la materia se extrae de una pieza, encontrará que sus paredes se tocan, pues no hay nada que las separe. Pero si Ud. encuentra que las paredes no se tocan, ha fallado en concebir la remoción de la materia (corporalidad) dentro de la pieza. Sin embargo, Descartes insiste, sería impío pensar que Dios todopoderoso no podría crear una pieza tal simplemente porque la idea de un vacío le choca a *uno* como contradictoria.⁴⁴

Hume nota que se sigue de su teoría del espacio o de la extensión como la manera u orden en el cual los puntos visibles tangibles se distribuyen que “no podemos formar la idea de un vacío, o de espacio, donde no hay nada visible o tangible,” pero se percata de tres objeciones poderosas a su conclusión (Hume 1960, p. 53). La primera objeción, y la que quiero discutir aquí, sostiene que “la mismísima disputa es decisiva respecto de la idea, y de que es imposible que los hombres pudiesen razonar por tanto tiempo acerca del vacío, y

⁴⁴ Véase en Descartes (1981) la Carta a Mersenne del 9 de enero de 1639. Véase también su Carta a Arnauld del 29 de julio de 1648.

refutarla o defenderla, sin tener una noción de lo que han refutado o defendido” (Hume 1960, p. 54). Lo que hace tan atractiva a esta objeción es el principio de Hume de que la concepción involucra la posibilidad. Si uno tiene la idea o concepción de un vacío, entonces la disputa se resolvería decisivamente a favor de su existencia o al menos de su posibilidad.

De ahí en adelante, Hume comienza una intrincada cadena de razonamiento, una mediada a cada paso por un principio u otro de concebibilidad establecido implícitamente. Más aún, él no duda en llevar a cabo experimentos mentales anidados [*nested thought experiments*], i.e., experimentos mentales dentro de experimentos mentales, e.g. uno en el cual se le solicita al lector concebir puntos de luz separados por una oscuridad total, como si se viesan simultáneamente muchas estrellas en una noche oscura. Este experimento mental lleva a Hume a postular un segundo tipo de distancia, una distancia invisible e intangible, que está ligada a una distancia visible y tangible de modos contrafácticamente especificados. Eventualmente, la cadena de razonamiento y los experimentos mentales insertos en dicha cadena [*embedded*] conducen a la “paradoja...de que si uno está satisfecho en dar a la distancia invisible e intangible (separando las estrellas en una noche oscura), o dicho de otro modo, a la capacidad de convertirse en una distancia visible y tangible, el nombre de un vacío, la extensión y la materia son lo mismo, y aun así, hay un vacío. Si Ud. no le da ese nombre, el movimiento es posible en un *plenum*, sin ningún impulso *in infinitum*, sin retornar en un círculo, y sin penetración” (Hume 1960, p. 64).

A diferencia de este manejo del caso del tono faltante del azul, la detallada demostración de Hume de la doctrina propuesta en “Sobre las ideas del espacio y tiempo”, no hace apelación a juicios sacados de un sombrero sobre la concebibilidad, ni mucho menos consiente de manera pusilánime con ellos. Hume posee una teoría bien armada de la concebibilidad, y de principio a fin la cita para aplicarla en todos los casos con que trata. A menos que haya pasado por el colador de su

teoría, no se apela a ninguna idea de la concebibilidad, ni se confía en ningún experimento mental.

Por supuesto, Hume no es el único filósofo que ha presentado principios que se apliquen en la práctica filosófica de la concepción. Otro ejemplo, uno que merece ser más conocido, es el de Helmholtz. Él tomó en serio los postulados neokantianos que, aunque tan consistentes como la geometría euclideana, las geometrías no-euclidianas son estrictamente inconcebibles, i.e., que los seres humanos no pueden concebir espacios que exhiben características no-euclidianas. Al reconocer que uno no puede tratar con esta cuestión sin una teoría articulada de la concepción espacial, Helmholtz se propuso desarrollar una. Sus investigaciones lo llevaron a formular el siguiente criterio de la concepción espacial: concebir el espacio correspondiente a una geometría G es describir la serie de sensaciones que serían experimentadas por criaturas habitando un espacio cuya geometría es descrita por G , criaturas cuya fisiología es similar a la nuestra y está gobernada por las mismas leyes. Dotado con este criterio de la concebibilidad, Helmholtz fue capaz de demostrar que los espacios no-euclidianos no eran menos concebibles que los espacios euclidianos familiares, *pese* a los neokantianos (Helmholtz 1962).

La existencia de un vacío y la concebibilidad de espacios no-euclidianos eran problemas científicos serios, entonces tal vez no es sorprendente que las soluciones a estos incorporasen principios explícitos y doctrinas de la concebibilidad cuidadosamente ramificadas. Parece que la propia disponibilidad de uno a sacar la tesis de la concebibilidad desde un sombrero es inversamente proporcional a la importancia de los problemas a tratar. Si esto es así, podríamos concluir, de la aceptación informal de los experimentos mentales sin base teórica, que la moneda de cambio de gran parte de la filosofía contemporánea consiste en una pura banalidad.

El lado bueno es que se puede esperar el desarrollo de una teoría

general de la concebibilidad para seguirle el tranco a los esfuerzos de lidiar con los serios problemas conceptuales enfrentados u ocasionados por la ciencia en sus diversas ramas. La teoría de conjuntos es un ejemplo prototípico. Tempranamente, los matemáticos pensaron que podrían basar la teoría de conjuntos en concepciones intuitivas o sacadas del sombrero de las propiedades o condiciones que determinan los conjuntos. *Inter alia*, la teoría de Russell mostró lo equivocado que estaban. En respuesta a esto y a otras antinomias, los matemáticos y lógicos gradualmente articularon diversas teorías de la concebibilidad echas a medida para los conjuntos, un proceso de formación de teorías que todavía permanece inconcluso.

La teoría de la computabilidad representa otro ejemplo. Suponga que alguien propone una tesis sacada del sombrero de que él o ella pueden concebir un aparato algorítmico que distingue las formulas bien formadas de primer orden universalmente válidas del resto. ¿Se rendirían los matemáticos y lógicos, renunciando a la tesis de Church o incluso al teorema de Church? La sola sugerencia es ridícula. Primero, el teorema de Church ha sido probado por lo que no hay forma de negarlo. Segundo, hay estándares de concebibilidad conocidos que deben satisfacerse por cualquier afirmación que se tome como aplicable a la tesis de Church. Aquí no se aceptan tesis de la concebibilidad sacadas del sombrero.⁴⁵

Permítaseme volver a mi tesis de que en el fondo la analiticidad y la concebibilidad son la misma noción. Si mi tesis es correcta, o al menos, no demasiado errónea, se debería esperar encontrar que en la edad de oro [*heyday*] del análisis del significado, las teorías articuladas

⁴⁵ La tesis de Church y el teorema de Church son a menudo confundidos, o van de la mano. El teorema de Church propone que, dado el número Gödel de las fórmulas de la lógica de primer orden, la función teorema (la función que da 1 cuando se aplica al número Gödel de un teorema de la lógica de primer orden, y que da 2 cuando se aplica al número Gödel de un no teorema) no es Turing computable. La tesis de Church propone que toda función intuitivamente algorítmica es Turing computable. (Para la computabilidad, por supuesto, uno podría sustituir cualquiera de las muchas nociones que se saben equivalentes para ésta).

del significado —teorías parciales sin duda— fueron citadas para aplicarse a problemas filosóficos precisamente en la medida que pudieron tratar con problemas interesantes para la ciencia. Aquí hay una tesis de que el registro histórico puede resolver. Los invito a echarle un vistazo.⁴⁶

Referencias

- Descartes, R. (1981): *Descartes: Philosophical Letters*, ed. y trad. por Antony Kenny. Minneapolis: University of Minnesota Press.
- Hume, D. (1981): *An Enquiry concerning Human Understanding*. Indianápolis: Hackett Publishing Co.
- _____ (1983): *Enquiry concerning the Principles of Morals*. Indianápolis: Hackett Publishing and Co.
- _____ (1960): *A Treatise of Human Nature*. (Shelby-Bigge, L.A. ed.) Oxford: Clarendon Press.
- Nash, O. (1931): *Hard Lines*. New York: Simon and Schuster.
- Putnam, H. (1982): “A problem about reference”. En: *Reason, Truth and History* Cambridge: Cambridge University Press.
- Quine, W.V.O. (1960): *Word and Object*. New York: John Wiley & Sons.
- _____ (1964): “Two dogmas of empiricism”. Reimpreso en: *From a Logical Point of View*. Cambridge, MA: Harvard University Press.
- von Helmholtz, H. (1962): “On the origin and significance of geometrical axioms”. En: *Popular Scientific Essays*. New York: Dover Publications.

⁴⁶ Este ensayo fue originalmente presentado en el Taller sobre el rol de los experimentos mentales en la ciencia y la filosofía, que fue organizado entre el 18 y el 20 de Abril de 1986, en la U. de Pittsburg. Estoy agradecido a la Dra. Barbara D. Massey, y a los miembros de mi seminario de graduados sobre concebibilidad de la U. de Pittsburg por sus valiosos comentarios y sugerencias.

PARTE II

¿Son los experimentos mentales y las filosofías de sillón
herramientas confiables?
Críticos y defensores

Introducción

Tal como se abordó en la primera parte, los experimentos mentales son moneda corriente en ciencia y filosofía, pese a las limitaciones que poseen, algunas de las cuales ya fueron expuestas. Con todo, ni científicos ni filósofos se frenan usándolos. En el caso de la filosofía, son particularmente populares, pues están ligados al análisis conceptual, el origen de las filosofías de sillón, cuyo método es *a priori*. En disciplinas como la epistemología, la ética y la filosofía de la mente se recurre frecuentemente a experimentos mentales para llevar a cabo análisis “desde el sillón”. Esto ha provocado una reciente ola de críticas por parte de algunos filósofos, quienes piensan que recurrir a métodos *a priori*, que soslayan del todo la investigación empírica, es un error fundamental que debe ser erradicado. Otros, también escépticos, han elaborado ácidos cuestionamientos a un tipo específico de experimento mental que explora la propia mente, denominados “bombas de intuiciones”. Estos serían pintorescas narrativas con detalles sofisticados y poco relevantes, cuyo objetivo central sería mover las perillas de la imaginación en el lector, y así forzarlo a llegar a intuiciones sobre la mente, no solo carentes de fundamento real, sino reñidas con lo informado por algunas disciplinas científicas.

Tanto las críticas a los experimentos mentales como a las filosofías de sillón han tenido importantes repercusiones en dos disciplinas de la filosofía contemporánea: la epistemología y la filosofía de la mente. La primera ha elaborado un tradicional tipo de análisis conceptual para esclarecer qué condiciones necesarias y suficientes deben satisfacerse cuando un agente conoce que *p*, es decir, qué debe

satisfacerse para que exista conocimiento genuino. Se han formulado una serie de respuestas, desde Platón en adelante, tales como las del fundacionalismo, del coherentismo, de la responsabilidad epistémica, del contextualismo, entre varias otras, todas las cuales enfatizarían cuestiones normativas en contra del escepticismo radical, es decir, qué *debe* satisfacerse desde el punto de vista conceptual cuando existe conocimiento. Una escuela influyente es la epistemología naturalizada de Quine (1969) y Kornblith (1994). Ellos ven la epistemología como un continuo con la ciencia, pues aquella debe informarnos *cómo* conocemos, no como *deberíamos* conocer. No es, en consecuencia, el concepto de conocimiento lo que debe analizarse, sino que se debe recurrir a la psicología y a la ciencia cognitiva para examinar en qué forma conocemos realmente. Por supuesto, la epistemología naturalizada, si bien no tiene explícitamente una posición contraria a los experimentos mentales, está contra los métodos de sillón, pues llevan a ignorar el trabajo de la ciencia. Pero, dicha escuela ha encontrado resistencia por parte de los proclives a los métodos de sillón, y en especial por los defensores de los experimentos mentales. Justamente, la relación entre ambos será examinada en la segunda parte del libro.

A diferencia de la epistemología contemporánea, la cual trata de resolver el puzle de la imposibilidad del conocimiento, la filosofía de la mente se ha concentrado, entre otros problemas, en tratar con el problema mente-cuerpo, heredado desde Descartes. Las teorías dualistas, según las cuales, mente y materia son ontológicamente diversas, y las materialistas, de acuerdo con las que la mente no es sino otro objeto material, han atizado la discusión sobre la naturaleza de lo mental. Qué son, finalmente, los estados mentales como creencias (Juan cree que Obama es el presidente de USA), deseos (Juan no desea votar por el *Tea Party*), sentimientos (Juan siente ansiedad), emociones (Juan siente ira), entre otros, es lo que intentan responder las teorías en filosofía de la mente. Particular importancia ha tenido para proponer las mismas el análisis introspectivo de dichos estados

mentales, así como la postulación de experimentos mentales basados en la introspección. Ambos métodos, típicos de la aproximación de sillón a lo mental, han sido criticados por los filósofos que ven en aquellos el peligro de ignorar el trabajo de la ciencia, informado empíricamente. De hecho, el concepto de “bomba de intuición” acuñado por Dennett y Hofstadter en 1980 (Dennett y Hofstadter 1981) a propósito de la Habitación China de Searle (1980) tiene una inspiración muy similar al enfoque naturalizado de Quine y Kornblith. Los dos primeros, al igual que estos últimos, sostienen que la ciencia, no la filosofía, finalmente resolverá los problemas filosóficos de la intencionalidad, de las emociones, de la conciencia, etc.

Luego de los 80's se ha debatido con pasión si el fenómeno de la conciencia puede reducirse explicativamente a alguna teoría que explore la mente desde lo no subjetivo, o desde lo que Searle (2004) llama, desde el punto de vista de la tercera persona. Tal como el análisis de sillón en epistemología, los experimentos mentales de Searle (1980 y 1990), Jackson (1986) y Nagel (1974 y 1986) presentan argumentos que, pese a su carácter controvertido y a basarse en la confiabilidad de la introspección, muestran en qué sentido la conciencia no puede reducirse metafísicamente de manera fisicista, a la manera de Agua= H_{20} , tal como usualmente pretenden los materialistas radicales. En efecto, estar en la pieza china, ponerse en el lugar de *Mary* la científica o imaginar ser un murciélago resulta imposible si no se examinan estas situaciones desde el laboratorio de la mente, el cual permite ejecutar estos experimentos.

De manera análoga a las críticas que suscita la introspección en los adversarios de la filosofía de sillón, frecuentemente se ha atacado que los experimentos mentales como los arriba mencionados no son confiables, justamente por bombear intuiciones indebidamente justificadas, o relacionadas con conceptos *folk*, muy ligados al sentido común. En vista de tales críticas, la pregunta que cabe es la siguiente: ¿Son capaces dichos experimentos mentales de producir creencias verdaderas, tal como lo hacen los experimentos reales, efectivos, pese

a los escépticos? Aunque la crítica de Dennett y Hofstadter (1981) sobre las bombas de intuiciones es tan ácida como relevante, da la impresión de que, al igual que con lo que acontece con Kornblith, se pretende que la filosofía abandone del todo los experimentos mentales y las filosofías de sillón. Pero, ¿será eso necesario?

Mi intención en esta segunda parte es que el lector juzgue por sí mismo las ventajas y desventajas de abandonar los experimentos mentales que incluyen bombas de intuiciones y las filosofías de sillón, con sus respectivos métodos. Al menos, *prima facie*, parece que abandonar la experimentación mental tipo bomba de intuición y la filosofía de sillón arriesga dejar de lado problemas cruciales a los que se ha abocado el análisis contemporáneo. Y, en este sentido, estimo que el fundamento de la crítica de los epistemólogos naturalizados, esto es, que no debe soslayarse el trabajo empírico de las disciplinas científicas puede contestarse de manera análoga: no existe tampoco fundamento adecuado para abandonar las filosofías de sillón y ponerse el “delantal de los científicos”. Hacerlo, a mi juicio, podría perjudicar el desarrollo de la filosofía, y también de la ciencia, ya que los propios científicos recurren a experimentos mentales y, por tanto, a métodos *a priori*, especialmente cuando hay impedimentos técnicos o financieros de llevar a cabo experimentos reales. Con el objeto de que el lector forme una opinión y juicio fundado, se han seleccionado defensas y críticas a los experimentos mentales, tanto en el ámbito de la epistemología como de la filosofía de la mente.

El primer artículo, de Timothy Williamson, justamente examina los métodos de sillón en filosofía, concentrándose en si son o no confiables. Por supuesto, toca los experimentos mentales como un típico método de sillón. Aunque ha habido importantes críticas a los experimentos mentales y las filosofías de sillón, tradicionalmente tales métodos han sido motores fundamentales en la actividad filosófica. ¿O de su estancamiento y falta de progreso, dirían los críticos? Las relaciones de ideas, de conceptos, usualmente se han considerado como el objeto típico de análisis conceptual *a priori*, pero desafortunadamente tal

empresa parece haber significado que, desde el sillón, se soslayara el trabajo empírico de la ciencia, en ámbitos donde ha habido progreso, como por ejemplo los estudios de la psicología de la percepción y de la memoria. La ciencia, asimismo, ha sido mucho más eficiente que la metafísica en dar una cosmovisión del mundo, con apoyo de la experiencia. ¿Puede la filosofía sobrevivir frente a una ciencia exitosa, o como sostiene Massey (en esta compilación), está condenada a alentar desde el borde del campo de juego, restringiendo su actividad al puro esclarecimiento conceptual? Con relación a esta posibilidad, Williamson no insiste sobre la confiabilidad del razonamiento *a priori*; en vez de eso, argumenta que la filosofía es de todas formas pariente cercana de la ciencia, pues los filósofos, incluso si adoptan métodos de sillón, especialmente en la metafísica, hacen contribuciones conceptuales y teóricas que van más allá de lo fenoménico: la noción de mundo posible y la argumentación contrafáctica, por ejemplo, son cruciales para testear si las apariencias descritas por las disciplinas científicas son en efecto verídicas.

Sin embargo, el grueso de la defensa de los métodos de sillón de Williamson se basa en mostrar las falencias de lo que denomina el *empirismo extremo*, posición según la cual se deben abandonar todos los métodos de sillón, filosóficos y científicos por igual, en una suerte de empresa naturalista inspirada por Quine (1969). Pese a esta, la propia ciencia parece adoptar métodos de sillón, por ejemplo, en la matemática, en la lógica, o con la realización de experimentos mentales. En el otro polo está, según Williamson, el racionalismo extremo que justifica los métodos de sillón y, al igual que el empirismo radical, presenta dificultades. La más importante es apelar a una facultad racional especial, de la cual no se tiene una buena explicación desde el punto de vista evolutivo. Frente a los dos polos, Williamson insiste en la postura adoptada en su libro *The Philosophy of Philosophy* (Williamson 2007), en el cual realiza un rescate de los métodos de sillón, sin recurrir a una facultad especial de la razón. En aras de la argumentación, presenta un caso de ficción:

la leyenda de Diógenes, quien desplumó un pollo para “refutar” que el ser humano es un “bípedo implume”, la supuesta definición platónica. En este sentido, ¿tendría que observarse a Diógenes pelar el pollo para persuadirnos de que el ser humano no es un bípedo implume? El análisis de Williamson se concentra en mostrar que es *suficiente* pero no *necesario* imaginar un caso en que Diógenes pelase un pollo para efectos de la refutación, la cual incluso podría apoyarse en análisis contrafáctico de condicionales como este: “Si Diógenes pelara el pollo, se refutaría que el humano es un bípedo implume”. No es tampoco necesario apelar a una facultad racional especial. Como Williamson remarca, el método de sillón *offline*, de comparar un juicio categórico y una suposición, no requiere de confrontar dos juicios categóricos; esto constituiría la vía *online* de refutación, en que la imaginación se conecta con la percepción. En particular, analizar la posibilidad lógica de que “todo bípedo implume sea humano” es suficiente para refutar este juicio, y abre un ámbito de análisis específicamente filosófico: la posibilidad metafísica. Justamente, en la última parte se enfatiza que el examen de la posibilidad es producto de una capacidad anticipatoria de situaciones y circunstancias posibles frente a las que hay que tomar decisiones, la que ha sido incentivada por la evolución. Y esto beneficia el examen de argumentos morales, con respecto a los cuales no es solo innecesario, sino totalmente inapropiado realizar experimentos reales. Finalmente, se hacen varios descargos al escepticismo del juicio que favorecen los filósofos experimentales, quienes aceptan las prácticas epistemológicas ordinarias cuando se trata de la ciencia, pero las rechazan como inadecuadas en el caso de la filosofía.

En vista de la cerrada *defensa de Williamson del sillón filosófico*, el segundo artículo, de Hilary Kornblith, toma las banderas para una aproximación empírica en filosofía, haciendo importantes aclaraciones respecto de la misión asignada a la epistemología por Quine en “*Epistemology Naturalized*” (1969). Aunque ha habido posturas defensivas respecto de la misión quineana de disolver el límite entre ciencia y filosofía, estas no se habrían atrincherado en el razonamiento

a *priori*, sino en que hay métodos de sillón comunes a ciencia y filosofía, lo cual es justamente la postura de Williamson, ciertamente tomada en consideración por Kornblith a lo largo de su texto. En este sentido, Williamson y otros ven la integridad de la especulación desde el sillón la que es ahora considerada como la genuina cuestión. A diferencia de ellos, la filosofía experimental critica los métodos de sillón por abocarse a análisis puramente conceptual. Como una manera de mostrar la inconveniencia de este método, los filósofos experimentales realizan encuestas acerca del uso *folk* de ciertos conceptos epistemológicos. Kornblith, con respecto a este punto, destaca que la epistemología científica que tienen en mente Quine, y también Goldman (1979), no está relacionada con el tipo de investigación llevada a cabo por los filósofos experimentales. Por el contrario, de lo que se trata es de cómo las cuestiones epistemológicas deben abordarse en función del aporte de la psicología, de cómo funciona la mente humana, que es la aproximación original de filósofos como Descartes, Locke, Hume y Kant, quienes no habrían entendido cómo puede caracterizarse qué es el conocimiento prescindiendo de las facultades psicológicas que permiten éste. Pero no es al estilo de los filósofos tradicionales que se debe explorar el conocimiento, mediante la introspección y la reconstrucción racional de nuestras creencias, sino mediante el aporte directo de teorías psicológicas que expliquen de qué forma se adquieren creencias irreflexivamente, mediante procesos mentales operando en el nivel sub-personal, lo que usualmente ignoramos. Lo mismo respecto de la creencia puede decirse del trabajo de la memoria y de la receptividad al testimonio: tampoco operan de acuerdo con el monitoreo introspectivo y reconstructivo de un agente, mucho menos, entonces, según las explicaciones de quienes favorecen los métodos de sillón. Estos tampoco pueden esclarecer la adquisición de conceptos; según Kornblith, solo el trabajo experimental lo hará. Pero, ¿qué tiene todo eso que ver con la epistemología? Para enfocarse en la importancia del trabajo experimental para la epistemología, Kornblith considera la tradicional discusión entre internistas y externistas. Un exponente típico de la primera visión es Bonjour. Para este la propiedad

de estar justificado es una propiedad internamente reconocible, pero las creencias que tienen esta propiedad son proclives a ser verdaderas. El problema, insiste Kornblith, radica en que la introspección, el análisis desde el sillón, dudosamente arroja luz alguna sobre por qué estamos justificados en creer que p. En este sentido, el monitoreo interno es tan poco confiable como la reconstrucción racional de adquisición y justificación de creencias, y lo mismo aplicaría para la responsabilidad epistémica, es decir, con relación a qué debe hacer un sujeto epistémico para conocer que p. Algo análogo sucede para todos los consejos epistémicos que usualmente se hacen: en vez de estudiar *cómo* realmente conocemos, con el apoyo de estudios empíricos, se concentran en un análisis de sillón en que se estudia el concepto de conocimiento, *no cómo este se adquiere*. Para Kornblith, resulta obvio que la vía de sillón debe abandonarse, pues sus métodos deben reemplazarse por métodos experimentales auténticamente confiables, que no apelen a intuiciones o conceptos examinados a priori. Términos centrales del discurso epistemológico son la percepción, la memoria, el testimonio y el razonamiento, todos los cuales parecen identificar varios fenómenos del mundo real, a saber, cómo conocemos de hecho. Por esta razón debe abandonarse el sillón, no hacerlo implica que el filósofo dé la espalda a los resultados experimentales. Como declara Kornblith, el conocimiento mismo, la justificación misma, el razonamiento mismo, la memoria misma, etc. deben explicarse, no *representarse* mentalmente mediante conceptos que den lugar a una discusión filosófica *a priori*. No debemos aclarar nuestros malentendidos conceptuales, debemos estudiar el conocimiento en sí mismo. Pese a estos juicios, Kornblith intenta finalizar con un tenor conciliatorio: la filosofía, al igual que la ciencia, se nutre del trabajo colectivo, el cual es dividido en pos de la eficiencia. Luego, del mismo modo en que se realiza física especulativa, pero contrastada con resultados experimentales, se debe realizar epistemología sin dar la espalda a los resultados experimentales. ¿Qué rol le cabe a los experimentos mentales, entonces? Más aún, ¿son estos a veces no confiables? ¿Cuándo lo son?

El artículo de Daniel Dennett, también inédito, explora un ámbito dentro de la filosofía de la mente, directamente relacionado con la confiabilidad de algunos experimentos mentales denominados por él y Hofstadter (1981) como “bombas de intuiciones”. Ambos consideran que estas, *grosso modo*, son argumentos fascinantes, engañosos y seductores, los cuales se transforman en aduladores de masas. Tal término es acuñado, precisamente, en vista de la Pieza China de Searle (1980 y 1990), una situación imaginaria o narrativa, en que se describe qué sucedería si alguien que habla solo inglés, manipulase símbolos chinos con base en su forma y en un libro de reglas. Supuestamente, tal experimento mental mostraría que computar, mediante sintaxis, ni es suficiente ni constitutivo de semántica, base del entendimiento lingüístico. Por supuesto, Dennett y Hofstadter creen que el experimento mental es falaz y engañoso, y para concluir esto, recurren a *manipular todas sus perillas*. Cuando esto se hace de forma adecuada, resulta claro que la narrativa empuja al lector a conclusiones que no están del todo aseguradas. Es decir, las bombas de intuiciones, aseguran Dennett y Hofstadter, son un particular método de sillón, tan poco confiable como artificioso (Dennett 1991), en tanto basan todo su poder en imaginaciones dependientes de contextos y de narraciones exóticas. Frecuentemente, con las bombas de intuiciones se “infiera un resultado”, un juicio intuitivo emotivo, a partir de un caso no solo imaginado, sino profundamente sobre-simplificado. Pero, ¿son tan dañinas las bombas de intuiciones? Aunque son usualmente vistas como herramientas pedagógicas, como métodos heurísticos, en filosofía ocasionalmente conducen a búsquedas estériles. Un ejemplo paradigmático, sería el genio maligno cartesiano, precursor del *cogito ergo sum*, “argumento” reconocidamente falaz y conducente al dualismo cartesiano. Sin embargo, Dennett defiende que las bombas de intuiciones también pueden servir a un propósito positivo: a diferencia de la ciencia y sus métodos empíricos, la filosofía puede recurrir a bombas de intuiciones para romper malos hábitos de pensamiento, o como sentenció Wittgenstein, para transformar la filosofía en “una batalla en contra del embrujo de nuestra inteligencia

por medio del lenguaje” (1953, sección 109). Un ámbito donde estas confusiones son abundantes es la filosofía de la mente y, en particular, en el problema de la experiencia consciente, con la discusión acerca de los denominados *qualia*. De hecho, Dennett también se refiere a otra bomba de intuición, *Mary* de Jackson (1986), la cual mediante premisas que invitan a la confusión y la ambigüedad intenta establecer que no todos los hechos son físicos, tal como el fisicismo propone. Aunque más abajo introduzco el artículo de Crane, que versa justamente sobre este problema, la pregunta que cabe formular entonces es la siguiente: dada toda la crítica de Dennett y Hofstadter, ¿es claro el significado y criterio de “bomba de intuición”? Paradójicamente, mi intuición es pese a su esfuerzo, no lo es.

Mi postura acerca de las bombas de intuiciones puede encontrarse en el quinto artículo de este libro. Tal ensayo se centra en cómo dichas bombas han recibido distinto tratamiento por parte de Dennett, desde que acuñó el término junto a Hofstadter, en 1980, hasta sus últimas reflexiones, en 2013. Según la postura original de ellos, las bombas de intuiciones son narrativas que dudosamente cuentan como argumentos. Por el contrario, serían simples *aduladores de masas*, artefactos epistémicos cuyo objetivo es la persuasión mediante el bombeo de intuiciones. Estas son, a su vez, movidas por la imaginación y la especulación, y no mediante la argumentación rigurosa. Sin embargo, la variación de este enfoque negativo de Dennett es el centro del ensayo que presento en esta compilación. En efecto, ya en 1984, en *Elbow Room* presenta comentarios más morigerados acerca de las bombas de intuiciones, afirmando que estas no son del todo negativas. Y no lo son porque son una especie de aguijón teórico que invita al lector a considerar un problema filosófico de manera general, sin perderse en los detalles, en los árboles del bosque. Pese a que pueden ser como direcciones erradas para encontrar un lugar, llevan a aventuras filosóficas fascinantes, lo cual muestra el cariz pedagógico, ligado a la discusión racional y crítica, propia del filosofar. Las bombas de intuiciones son, así, poderosos instrumentos pedagógicos

que, según él, deben ser analizadas caso a caso, “perilla a perilla”. Ello, ciertamente, permitiría la depuración de algunos problemas filosóficos, exorcizando el embrujo de la palabra, tal como, por ejemplo, intenta Wittgenstein (1953). Dennett 1988, a propósito de los *qualia*, y Dennett 1991, en relación con el problema de la conciencia sigue las mismas directrices. En este sentido, según el nuevo enfoque de Dennett, las bombas de intuiciones son herramientas teóricas exploratorias poderosas, y recalca, no hay muchas herramientas en la caja de herramientas del filósofo. Son, dichas herramientas, prótesis de la imaginación, la cual se encamina a la aventura de resolver un problema, o en el mejor de los casos, de exorcizarlo a lo Wittgenstein. Por ello considero que el segundo enfoque de las bombas de intuiciones, posterior a 1984, ciertamente plasmado y del todo desarrollado en 2013, es bastante positivo. De hecho, pone las bombas de intuiciones en relación con actividades filosóficas básicas: etiquetar, dar ejemplos, proponer analogías y metáforas y, sobre todo, armar el andamiaje de la ubicación de un problema. Es por este motivo que, incluso, afirma que “*ama las bombas de intuiciones*” (Dennett 2013, p. 11, énfasis mío), pese a lo que algunos filósofos han llegado a creer. Esta nueva postura con relación a las bombas de intuiciones, ciertamente más refrescante, se relaciona, a mi juicio, con el papel que juega el ensayo, pero por sobre todo el *error*, en la filosofía. A veces, y ello ocurriría seguido en esta disciplina, es necesario cometer errores, y como énfasis, hacerse consciente de los mismos. Tal cuestión ciertamente reivindica a las bombas de intuiciones porque, en síntesis, permiten aprender a pensar, a veces de manera gruesa, otras de manera más rigurosa. O bien a discutir, racional y críticamente, problemas característicos de la filosofía.

Finalmente, el artículo de Tim Crane muestra un tipo de dependencia contextual, a propósito del problema de la conciencia y la supuesta bomba de intuición que implicaría *Mary*, a saber: ciertos contextos de experiencia subjetiva, propios de la posición de un sujeto en el mundo, implican un tipo de conocimiento no objetivo. Este, dependiente

de contextos, es lo que Crane denomina *conocimiento subjetivo*, a propósito del problema metafísico que entrañaría la conciencia (Levine 1993). El argumento de Jackson, apoyado por el experimento mental de *Mary*, mostraría este último tipo de conocimiento. Junto con hacer una defensa de dicho conocimiento, ciertamente un aliciente para continuar con la experimentación mental como método de sillón, Crane realiza un examen de varias malas interpretaciones del argumento de Jackson (1982). Entre estas destacan que *implicaría* la existencia de los *qualia*, que es inválido debido a que involucra de manera equívoca dos nociones de conocimiento, proposicional y por familiaridad, y que la segunda premisa, que versa sobre cómo *Mary* aprende algo nuevo, es falsa. De hecho, el contra-experimento mental de Dennett (2005) precisamente se basa en la supuesta falsedad, o al menos el carácter dudoso de dicha segunda premisa, con base en que ella *debería* saber *todo* lo que hay que saber sobre el color; luego, ¿cómo podría “aprender” algo que de antemano no sabía? Crane defiende punto por punto el argumento, con el objeto de mostrar que la lección y moraleja que pueden aprenderse debe todavía ser captada: el experimento mental muestra, más que la falsedad del *fisicismo*, que algunos hechos son *subjetivos*, y que se tiene conocimiento, ciertamente *proposicional*, tal como cuando se emplean *indexicales* de la forma: “¡Estoy aquí!” Si esta frase se emplea en el contexto apropiado, argumenta Crane, como cuando uno señala un punto de un mapa en un contexto determinado, ciertamente se aprende algo nuevo, algo que no se sabía antes. En este respecto, este filósofo enfatiza que el argumento del conocimiento es un argumento sólido respecto de la conclusión de que existen hechos *subjetivos*: hechos acerca del carácter *subjetivo* de la experiencia de un sujeto. Curiosa e interesantemente, también sostiene que tal conclusión no amenaza una versión plausible del *fisicismo*, porque este no implica que no haya hechos *subjetivos*, ni tampoco que *todos los hechos son objetivos*. Al contrario, suponer que el argumento de Jackson socava al *fisicismo* involucra malinterpretar este, como una posición *epistemológica* y no *metafísica*. Justamente, concluye el artículo sugiriendo que ni la

ciencia física, ni ninguna otra ciencia objetiva, puede establecer todos los hechos, pero esto no puede ser la base de una crítica al fisicismo. Al contrario, su postura en relación con este punto revelada por el epígrafe de su trabajo, la notable intuición de Russell: un vidente ve cosas que un ciego no puede saber y, aún así, este puede saber toda la física. Esto indica que el ciego, aunque puede saber teóricamente toda la física y luego, el conocimiento que otros hombres tienen, no tendrá una parte de la física. Y esta intuición, tal como las intuiciones que ganamos mediante experimentos mentales, no solo muestra que no hay, pese a Dennett y Hofstadter, una buena razón para abandonarlos. Por el contrario, vale la pena conocer las bombas de intuiciones; en efecto, ellas son herramientas de exploración de la naturaleza de lo mental, y ciertamente un objeto de estudio en las filosofías de sillón.

Referencias

- Dennett, D. (1984): *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: Bradford Books.
- _____ (1988): “Quining qualia”. M. y E. Bisiach (eds.) *Consciousness in Contemporary Science*. Oxford: OUP, pp. 43-77.
- _____ (1991): *Consciousness Explained*. New York: Little Brown.
- _____ (2005): *Sweet Dreams*. Cambridge, MA: MIT Press.
- _____ (2013): *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Co.
- Dennett, D. (1991): *Consciousness Explained*. New York: Little Brown.
- _____ (2005): *Sweet Dreams*. Cambridge, MA: MIT Press.
- Dennett, D. y Hofstadter, D. (1981): *The Mind's I: Fantasies and Reflections on Mind and Soul*. New York: Basic Books.
- Goldman, A. (1979): “What is justified belief?” En: H. Kornblith (ed.) *Naturalizing Epistemology*. Cambridge, MA: MIT Press, pp. 104-130.
- Jackson, F. (1982): “Epiphenomenal qualia”. *Philosophical Quarterly* 32, 127-136.

- _____ (1986): "What Mary Didn't Know". *The Journal of Philosophy* 83 (5), 291-5.
- Kornblith, H. (1994): "Introduction: What is naturalistic epistemology?" En: *Naturalizing Epistemology*. Cambridge, MA: MIT Press, pp. 1-14.
- Levine, J. (1993): "On Leaving out What It's Like". En: *M. Davies and G. Humphreys (eds.) Consciousness*. Oxford: Blackwell. Reimpreso en: N. Block, O. Flanagan y G. Güzeldere (eds.) *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, pp. 543-556.
- Nagel, T. (1974): "What Is It Like to Be a Bat?" *Philosophical Review* 83, 435-50.
- _____ (1986): *The View from Nowhere*. Oxford: OUP.
- Quine, W.V.O. (1969): "Epistemology naturalized". En: *Ontological Relativity and Other Essays*. New York: Columbia University Press, pp. 69-90.
- Searle, J. (1980): "Minds, Brains and Programs". *Behavioral and Brain Sciences* 3, 417-24.
- _____ (1990): "Is the Brain's Mind a Computer Program?" *Scientific American*, January 1990, 20-25.
- _____ (2004): *Mind: A Brief Introduction*. Oxford: OUP.
- Williamson, T. (2007): *The Philosophy of Philosophy*. Oxford: Blackwell.
- Wittgenstein, L. (1953): *Philosophical Investigations*. Oxford: Blackwell.

Conocimiento de Sillón y Naturaleza de la Filosofía⁴⁷

Timothy Williamson

1. El surgimiento de la ciencia moderna ha tornado cada vez más apremiante un desafío epistemológico a la filosofía. Los métodos estándar de la filosofía son de sillón. Involucran el pensamiento sin observación o experimento. Pero, ¿cuán confiados podemos estar de nuestras teorías filosóficas si rehusamos pararnos del sillón para chequear si acaso el mundo que está más allá de éste se conforma a lo que [dichas teorías] dicen? ¿Son los departamentos de filosofía sobrevivientes anacrónicos de un estadio pre-científico?

La rama de la filosofía más obviamente vulnerable a este desafío es la metafísica, pues parece erigirse como una rival a la ciencia natural básica, tratando, mediante métodos *a priori*, de responder las mismas preguntas acerca de la naturaleza esencial de cualquier realidad que subyazca a las apariencias. La crítica a la metafísica, o al menos ese tipo de metafísica, ha sido de hecho la versión canónica del desafío desde Hume y Kant, pasando por los Positivistas Lógicos, hasta los críticos contemporáneos de la metafísica reciente. Sin embargo, todas las ramas de la filosofía están, en principio y usualmente en la práctica, abiertas a tales desafíos. ¿Pueden los filósofos morales simplemente ignorar los descubrimientos antropológicos acerca de las sociedades con sistemas morales muy diferentes a los nuestros? ¿Pueden los epistemólogos simplemente ignorar los resultados experimentales

⁴⁷ Conferencia en el Instituto de Filosofía, PUC, Santiago de Chile, 14 de mayo de 2009. Traducción y publicación con permiso del autor.

acerca de la psicología de la percepción y de la memoria? ¿Pueden los filósofos del lenguaje simplemente ignorar los resultados de la lingüística?

La respuesta positiva predominante al reto ha sido el intento de demarcar un tipo especial de pregunta para el cual los métodos de sillón de la filosofía son apropiados, dividiendo, mediante ello, la filosofía de sillón en filosofía legítima, que contesta las preguntas de este tipo, e ilegítima, que intenta responder a preguntas de otros tipos. La idea guía es que las preguntas filosóficas legítimas no suscitan conjeturas acerca de cómo es el mundo más allá del sillón, de manera que no hay necesidad de levantarse del sillón para chequear si acaso es realmente así. Por ejemplo, Hume distinguió entre *relaciones de ideas* y *cuestiones de hecho*. De acuerdo con su visión, los métodos de sillón son apropiados para encontrar relaciones de ideas, pero no para investigar cuestiones de hecho. Aunque lo sintético *a priori* de Kant va más allá que las relaciones de ideas de Hume (lo analítico *a priori*), todavía se ocupa de lo que introducimos a la experiencia del mundo y no de cómo es el mundo independiente de nuestra experiencia de éste. Tanto Hume como Kant confinan los métodos legítimos de sillón a la investigación sobre lo que la mente ha traído a este último, la estructura de nuestras ideas o de nuestra experiencia. En el siglo veinte, el Giro Lingüístico implicó una estrategia similar, centrada en el lenguaje en vez de la mente: los métodos de sillón son legítimos cuando consisten en una reflexión sobre el lenguaje del cual somos maestros en el sillón, ya sea el lenguaje ordinario que traemos a éste [el sillón] (como para Wittgenstein y Austin) o el lenguaje artificial que ideamos cuando ya estamos en el sillón (como para Frege y Carnap). La idea de legitimar una metodología de sillón confinando la filosofía a un tipo especial de preguntas está extendida no sólo en la tradición de la filosofía analítica, sino también en el racimo de tradiciones conocida como “filosofía continental”. Por ejemplo, los fenomenólogos se concentran en cómo las cosas se aparecen desde la perspectiva del sujeto, independientemente de si son realmente de esa

manera (Austin una vez describió su filosofía del lenguaje ordinario como “fenomenología lingüística”) (Austin 1961). Ni tampoco el Giro Lingüístico está confinado a la filosofía analítica; una forma de este puede discernirse en el trabajo de filósofos postmodernistas como Derrida (1974).

Aún hoy día, muchos filósofos explicarían que ellos tratan con “preguntas conceptuales” antes que con “preguntas empíricas”. Exactamente qué significan por “conceptuales” y “empíricas” es a menudo poco claro. Lo que usualmente resulta claro es que ellos intentan que la distinción se correlacione con una distinción entre preguntas para las cuales una metodología de sillón es apropiada y preguntas para las cuales dicha metodología es inapropiada. Pero por “pregunta conceptual” ellos simplemente no *significan* “pregunta para la cual una metodología de sillón es apropiada”; más bien, decir que una pregunta es conceptual supone *explicar por qué* una metodología de sillón es apropiada para ella. Se supone que una pregunta conceptual es acerca de conceptos (sea lo que sea que son estos exactamente) de algún modo que posibilita que ésta sea contestada mediante la reflexión sobre esos conceptos (y quizá sobre otros).

En décadas recientes, la filosofía se ha movido en una dirección que hace que cualquier distinción entre sus preguntas “conceptuales” y las preguntas “empíricas” de la ciencia natural sea cada vez más difícil de mantener. Quine ha insistido que la filosofía es continua con la ciencia natural. Más recientemente, ha habido un resurgimiento de la metafísica pre-kantiana que formula preguntas acerca de la naturaleza esencial de cualquier realidad que subyazca a las apariencias, sin ninguna restricción a la realidad “tal como la experimentamos”. Tales metafísicos están más cerca en espíritu a Leibniz que a Kant, y no solamente porque invoquen mundos posibles (David Lewis es un ejemplo destacable). Sus preguntas no son “conceptuales” en ningún sentido que las distinga en cuanto a su naturaleza de las que formulan los científicos naturales. Ellos se ven a sí mismos como contribuyendo con algunas partes muy generales de una teoría total del

mundo a la que los científicos hacen contribuciones más específicas. De hecho, en la práctica la filosofía de la física es continua con la física altamente teórica; la filosofía de la psicología es continua con la psicología altamente teórica, etc. Los filósofos no están simplemente “clarificando” los conceptos de los científicos; están usándolos para construir teorías especulativas que intentan ser integradas con las mejores teorías del dominio que los mismos científicos han producido. Ellos no están describiendo como el dominio se nos aparece a nosotros o incluso a los científicos; están preguntando si esas apariencias son verídicas.

De este modo, en muchas de las áreas más vitales de la filosofía contemporánea, los filósofos están rehusándose a ser confinados a preguntas de algún tipo peculiarmente “conceptual” o “fenomenológico” que explicara el carácter apropiado de los métodos de sillón. Donde los resultados de la observación y del experimento son relevantes, ellos los toman en cuenta. No obstante, en muchos casos continúan usando una metodología de sillón, por ejemplo en la metafísica más contemporánea. Puesto que se han ubicado más allá del alcance de la naturaleza “conceptual” o “fenomenológica” de sus preguntas al justificar una metodología tal, para ellos el reto epistemológico asume una forma particularmente urgente.

El problema para la apelación a la naturaleza “conceptual” o “fenomenológica” de las preguntas filosóficas al justificar la metodología de sillón no es solamente que muchas de las preguntas a las cuales los filósofos aplican la metodología no tienen esa naturaleza. Es también poco claro por qué una filosofía de sillón debiera ser adecuada incluso para preguntas cuya naturaleza parece ser “conceptual” o “fenomenológica”. Por ejemplo, preguntas acerca de lo que lógicamente entraña aquellos que a menudo se consideraban como paradigmas de preguntas “conceptuales”. La esperanza podría ser que nuestra captación de los conceptos relevantes o nuestra comprensión de las palabras que las expresan nos permite reconocer si acaso el entrañe en cuestión se sostiene y, mediante ello, responder

a la pregunta desde el sillón. Obviamente, si nosotros no captáramos los conceptos o no entendiéramos las palabras, estaríamos en una mala posición para responder la pregunta, pero eso no muestra nada especial acerca de este tipo de pregunta, puesto que estamos en mala posición para responder cualquier pregunta que no entendamos. El problema es hasta dónde nuestra competencia lingüística o conceptual por sí sola nos lleva hacia la respuesta de preguntas de este tipo especial. La dificultad es que uno puede entender una palabra de su lenguaje nativo y, en virtud de ello, captar el concepto que ésta expresa, a través de tener la habilidad para usarla efectivamente en la comunicación, aun cuando uno tenga creencias completamente erróneas acerca de sus propiedades lógicas, si uno se ajusta apropiadamente al error de otros para compartir esas creencias. Muchos filósofos usan esos ejemplos. Reconocer las implicaciones [*entailments*] requiere más habilidad lógica que reconocer las palabras. Llamar a la pregunta “conceptual” no explica por qué la destreza lógica es mejor ejercida desde el sillón. Similarmente, una pregunta acerca de cómo se le aparecen a uno las cosas puede ser ‘fenomenológica’, pero no se sigue que ésta pueda ser mejor respondida desde el sillón. Porque puede argüirse que ni siquiera las apariencias son exactamente lo que parecen ser; las cosas pueden aparecerse a uno de algún modo aun cuando uno no esté en condiciones de saber que se aparecen de ese modo. Quizás un psicólogo que esté escaneando nuestro cerebro pueda en principio incluso ser capaz de descubrir que no estamos en el estado mental consciente que nosotros asumimos que estamos. De este modo, incluso en este posible campo de aplicación, la apelación a la naturaleza “conceptual” o “fenomenológica” de las preguntas filosóficas es mucho menor que lo que uno pudiera haber esperado para justificar una metodología de sillón.

Gran parte de la filosofía del siglo XX fue conformada por teorías profundamente erróneas de la naturaleza de la filosofía. Pero no debiéramos concluir que dicha filosofía no contenía algo de valor. Por ejemplo, en el curso del Giro Lingüístico aprendimos mucho acerca

de la estructura de grano fino de los argumentos. Eso a su vez ha mejorado nuestra habilidad para distinguir entre razonamiento válido e inválido, porque la validez de los sutiles argumentos filosóficos es a menudo sensible a pequeñas diferencias en la sintaxis y semántica de las premisas y de la conclusión. Nos hemos vuelto significativamente más confiables para discriminar el razonamiento válido del inválido, aun cuando el asunto del argumento no tenga nada que ver con la mente o el lenguaje. La confianza en la forma de superficie a menudo no es suficientemente buena. Los metafísicos, quienes insisten que están interesados en la estructura del mundo, no en la estructura de nuestro discurso habla acerca del mundo, y, por lo tanto, rehúsan a examinar esta última, son los más vulnerables a confundir las dos. Similarmente, los científicos, quienes se rehúsan a examinar sus instrumentos científicos sobre la base de que ellos están interesados en aquello que el uso de sus instrumentos les permite estudiar, no en los instrumentos mismos, son los que están más vulnerables a ser engañados por esos instrumentos. La mejor filosofía contemporánea es extremadamente cuidadosa acerca de la estructura lingüística y lógica de sus argumentos. No obstante, esos valiosos resultados del Giro Lingüístico no responden el reto epistemológico central a la filosofía de sillón. En algunos respectos, ellos lo hacen más apremiante, puesto que algo de nuestra presente sofisticación acerca de la estructura lingüística viene de la lingüística “empírica”.

Si buscamos una respuesta al desafío epistemológico que se ajuste a la filosofía analítica tal como se ha desarrollado en los últimos cuarenta años no necesitamos una que justifique una metodología *puramente* de sillón, puesto que dicha filosofía típicamente acepta los resultados de la ciencia “empírica” como relevantes en principio. Sin embargo, incluso el examen más superficial deja claro que la metodología usada es una de sillón. ¿Cómo puede una metodología tal ser apropiada si los filósofos ya no aceptan ningún límite para preguntas de un tipo “conceptual” o “empírico” especial?

De acuerdo con los empiristas extremos, el reto no puede

enfrentarse. La filosofía no puede obtener conocimiento más allá del sillón mediante el uso de métodos de sillón. Debe comenzar haciendo observaciones y experimentos, o al menos confiar en los que hacen otros. Pero el empirismo extremo es poco convincente. El mismo es un trozo tosco de epistemología de sillón. La observación y el experimento por sí solos no nos dicen que los métodos de sillón nunca son buenos —por ejemplo, en lógica. De hecho todas las ciencias naturales hacen un uso esencial de los métodos de sillón. La matemática es el ejemplo más obvio, pero está lejos de ser el único. Los métodos de sillón se usan al moverse desde los resultados estadísticamente analizados de las observaciones y experimentos a los juicios epistemológicos informales respecto de cuál de varias teorías está mejor sustentada por los datos. Por razones que previamente se han insinuado, describir las inferencias de sillón meramente como “conceptuales” o “analíticas”, no es de mucho valor para explicar su epistemología. En cualquier caso, los empiristas realmente extremos niegan que haya un dominio de verdades “conceptuales” o “analíticas” para las cuales sea apropiada una metodología de sillón. Ellos carecen de una buena explicación de por qué los métodos de sillón de la ciencia son apropiados, pero no los de la filosofía. De hecho, carecen de una buena explicación de lo que son los métodos de sillón, para si acaso los métodos de la ciencia *incluyen* a los de la filosofía.

En el otro polo, los racionalistas extremos responden al reto epistemológico postulando una facultad especial de la intuición racional que de algún modo u otro proporciona desde el sillón conocimiento del mundo que está más allá. Pero el racionalismo extremo es psicológicamente implausible. No tiene una buena historia acerca de por qué o cómo debiera haber evolucionado en la especie humana la supuesta facultad de la intuición racional. Tampoco hay afirmaciones convincentes respecto de una fenomenología de la intuición racional. Las afirmaciones filosóficas algunas veces nos parecen verdaderas o incluso necesarias, pero ¿qué es aquello que está más allá de nuestro estar conscientemente inclinado a juzgar que

[estas afirmaciones] son verdaderas o necesarias, del mismo modo que estamos inclinados a hacer todo tipo de otros juicios?

La filosofía contemporánea de sillón necesita una buena respuesta al reto epistemológico, una que evite los peligros opuestos del empirismo extremo y el racionalismo extremo sin confinar a la filosofía a un tipo especial de pregunta, “conceptual”, “fenomenológica” o del tipo que sea. Esa es la preocupación central de mi libro *The Philosophy of Philosophy* (Williamson 2007). Allí se hace algo del trabajo duro requerido para sustanciar las aserciones negativas deliberadamente impresionistas que he estado haciendo acerca de esas aproximaciones negativas. La crítica desarrollada en dicho libro puede aplicarse a los escritos metafilosóficos de muchos diferentes filósofos, aunque los detalles variarán extensamente entre caso y caso. Pero también desarrollo los inicios de una concepción positiva de la filosofía, la cual es mucho menos excepcional que lo que a los filósofos a menudo les gusta aspirar, tanto en sus preguntas como en los modos de responderlas. Sus métodos de sillón son aplicaciones de los métodos que usamos en forma mucho más extendida tanto en la ciencia como en la vida diaria. No necesitamos ni tenemos una facultad de la intuición racional. En lo que sigue usaré algunos ejemplos simples para ilustrar esta aproximación positiva.

2. Es típico que se afirme en la filosofía que una variedad de métodos son relevantes en principio, y a menudo en la práctica, para su evaluación. En particular, tanto las metodologías de sillón como las que no lo son pueden usarse. Uno podría preguntarse cómo la misma proposición puede pertenecer tanto a la categoría para la cual una metodología de sillón es apropiada como a la categoría para la cual una metodología que no es de sillón es apropiada. Sin embargo, las consideraciones lógicas muestran que si ambas categorías no son vacías, debiéramos esperar que se traslapen.

Supóngase que p es una proposición que puede establecerse (pero solamente con alguna dificultad) mediante la metodología de sillón,

mientras que q es una proposición diferente que puede establecerse (pero solamente con alguna dificultad) por medio de la observación y el experimento, y donde no hay ninguna conexión entre las dos proposiciones. Ahora, considérese la proposición disyuntiva p -o- q , la cual es verdadera si y solamente si o bien p es verdadera o bien q es verdadera. En un escenario posible establecemos p por medios de sillón e ignoramos q ; puesto que p -o- q es una consecuencia lógica de p , podemos deducirla de p , a partir de esto, p -o- q mediante metodologías de sillón. En otro posible escenario establecemos q por medio de observación y experimento e ignoramos p ; puesto que p -o- q es una consecuencia lógica de q , podemos deducirla de q y, a partir de esto, establecer p -o- q mediante observación y experimento. De este modo, la misma proposición disyuntiva puede establecerse tanto por medios de sillón como por medio de observación y experimento.

Obviamente, es improbable que los filósofos encuentren mucho interés intrínseco en la pregunta si acaso la aparentemente aleatoria disyunción p -o- q es verdadera. Sin embargo, el argumento muestra que es un error confiar en una dicotomía exclusiva entre proposiciones que pueden establecerse por medios de sillón y aquellas que pueden establecerse por medios que no son de sillón. Veremos que el mismo fenómeno ocurre también en casos no disyuntivos. Comenzaré con un ejemplo ficticio elemental por dos razones. En primer lugar, su estructura simple, esquemática acentúa los puntos epistemológicos más perspicuamente. En segundo lugar, si tomamos ejemplos de controversias importantes en la historia de la filosofía, probablemente estaremos lidiando con casos en los cuales la aplicación de los métodos estándar es problemática y, de este modo, arriesgamos perder de vista el hecho de que aquellos métodos también tienen muchas aplicaciones que no son problemáticas.

Según reza la leyenda, después de que Platón hubo definido al hombre como un bípedo implume, Diógenes de Sinope desplumó un pollo y dijo, “¡He aquí el hombre de Platón!”; Platón entonces revisó la definición como “bípedo implume con uñas anchas”

(Diógenes Laercio, *Vidas de los Filósofos*, 6.40). La anécdota es presumiblemente apócrifa, pero podemos usarla como una ficción conveniente para pensar en varias posibilidades. Continuaré usando los nombres ‘Platón’ y ‘Diógenes’, pero sin la menor pretensión de rigurosidad histórica.

Supongamos que la definición de Platón venía a ser lo mismo que esta proposición:

- (1) Ser un bípido implume es necesario y suficiente para ser humano.

Podemos asumir que (1) no requiere de la palabra ‘humano’ para ser sinónima con la frase “bípido implume” o del concepto *humano* para ser idéntica con el concepto bípido implume. Más bien, podemos suponer que (1) es una afirmación precipitada acerca de la propiedad de ser un humano y la propiedad de ser un bípido implume, no una afirmación acerca de las palabras o conceptos: nada podría ser un bípido implume sin ser un humano, y nada podría ser un humano sin ser un bípido implume —las propiedades coinciden individuo por individuo, no solamente especie por especie. Así, una consecuencia lógica de (1) es (2):

Todos y solamente los bípedos implumes son humanos.

La conversa no se sostiene: (1) no es una consecuencia lógica de (2), porque (2), a diferencia de (1), es consistente con que sea una coincidencia contingente que todos y solamente los bípedos implumes son humanos.

Mirando al pollo desplumado de Diógenes podemos decir:

Este es un bípido implume; no es humano.

Tal como se entiende la conjunción (3), es contingentemente verdadera; en particular un pollo cuenta como implume cuando se despluma aunque naturalmente tenga plumas. Sabemos (3) mediante percepción y no mediante reflexión de sillón. Claramente, (3) es

lógicamente inconsistente con (2). Puesto que (2) es una consecuencia lógica de (1), (3) es también lógicamente consistente con (1). De este modo, tenemos una refutación empírica de la tesis proto-filosófica (1) mediante la observación (3).

Obviamente, no importa mucho si vemos a Diógenes desplumar el pollo o solamente escuchamos acerca de ello. Suponga que un testigo confiable nos afirma que Diógenes ha desplumado un pollo corriente. Mediante su testimonio nos enteramos de que hay un pollo desplumado de dos patas. Conscientemente juzgamos que hay un bípedo desplumado y que no es humano. Llegamos a saber (4):

(4) Hay un bípedo desplumado que no es humano.

Claramente (4) es lógicamente inconsistente respecto de (2) y, entonces, con (1). De este modo, tenemos aún una refutación “empírica” de la tesis proto-filosófica (1), mediante testimonio en vez de observación directa.

¿Es realmente necesario que Diógenes desplume de verdad al pollo? Si lo vemos enfrente nuestro parado con el pollo en sus manos, evidentemente capaz y con la intención de desplumarlo, ¿no es eso ya suficiente para refutar (1), incluso si no se pone a desplumarlo? Sabemos que puede desplumar al pollo. Podemos ver que tiene dos patas. Sabemos conscientemente que si me pusiese a desplumarlo, sería un bípedo desplumado y no un humano. Al juntar estas dos informaciones, deducimos y por consiguiente concluimos (5):

(5) Puede haber un bípedo desplumado que no es humano.

No tenemos más una refutación de (2), debido a que (5) es lógicamente consistente con la verdad de (2), como una mera coincidencia contingente. Pero aún tenemos una refutación de la tesis proto-filosófica (1), dado que su fuerza modal hace que (1) sea lógicamente inconsistente con (5). Más aún, la refutación aún tiene un aspecto “empírico”, porque el caso presente para (5) depende de características observacionales de la situación en virtud de la cual

Diógenes tiene tanto la capacidad como la oportunidad de desplumar al pollo. Nuevamente, no importaría mucho si supiésemos acerca de estas características mediante testimonio en vez de mediante observación directa.

La comparación entre los dos casos testimoniales es particularmente cercana. En el primer caso, se nos dice que Diógenes ha desplumado al pollo. Categóricamente juzgamos que lo ha hecho. Nos movemos al juicio categórico de que existe un bípedo desplumado que no es humano, de modo que hay una cosa como esa, la cual es lo que (4) afirma. En el segundo caso, se nos dice que Diógenes *puede* desplumar el pollo. Categóricamente juzgamos que puede hacerlo. Hipotéticamente suponemos que lo ha hecho, y nos movemos a la suposición del juicio de que hay un bípedo desplumado que no es humano, de modo que existe tal cosa; entonces nos salimos de la suposición para concluir categóricamente que puede haber un bípedo desplumado que no es humano, que es lo que (5) asevera. Así, movimientos paralelos del pensamiento ocurren en ambos casos, entre los mismos contenidos, de que Diógenes ha desplumado el pollo y que hay un bípedo desplumado que no es humano. La diferencia es que en el primer caso, hay un movimiento desde juicio categórico a juicio categórico, mientras que en el segundo hay un movimiento desde una suposición a un juicio de acuerdo con una suposición. En jerga contemporánea, la primera movida tiene lugar “online” mientras que la segunda tiene lugar “offline”.

Podemos más aún reducir el elemento empírico en la refutación de (1) al reflexionar que no importa mucho si Diógenes realmente tiene la habilidad y oportunidad de desplumar al pollo. El mero hecho de que alguien pudiera haberlo desplumado es suficiente para refutar (1). Por supuesto, mientras menos específico el tipo de posibilidad en cuestión, más amplia es la gama de posibilidades que podrían haberse tomado en consideración en la movida desde la suposición que Diógenes ha desplumado el pollo al juicio hecho según esa suposición, que hay un bípedo desplumado que no es humano.

Evidentemente, tenemos un espectro de casos, desde la refutación observacional original de (1) a una refutación de (1) solo por medios de sillón. Podemos describirlos como ordenados desde lo altamente “empírico” a lo altamente “no empírico”. Aun así, las similitudes son tan impactantes como las diferencias. Una característica que varía de una manera sorprendentemente insignificante a través del espectro es el grado de persuasión [*cogency*] epistémica de las refutaciones. Si desplumamos nosotros el pollo o meramente imaginamos uno desplumado no significa una diferencia sustancial en la confianza que razonablemente podemos tener respecto de la falsedad de (1). La posibilidad en principio de percibir incorrectamente a Diógenes o al pollo tiene poco que ver de manera práctica con hacer que la refutación observacional original sea menos convincente. A la inversa, puro depender de un experimento mental basado en la imaginación no es relevante para que la refutación resulte menos convincente.

Sería absurdo de parte del empirista extremo insistir que Diógenes deba en la realidad desplumar al pollo para que (1) sea refutada. Pero no sería mucho más plausible para el racionalista extremo invocar la intuición racional al explicar la refutación de (1). Incluso cuando Diógenes se abstiene de desplumar realmente al pollo, sabemos el hecho contingente de que fácilmente podría haberlo desplumado, mediante observar y considerar de manera imaginaria la situación. No requerimos de una facultad especial de intuición racional para hacer eso. Aunque podemos gradualmente volcar nuestra atención hacia formas de imaginación más inespecíficas, de modo que menos y menos características de la situación real son relevantes, hasta que nos concentramos en la mera posibilidad metafísica, en ningún momento se necesita invocar una facultad especial. Del mismo modo, en ningún momento se requiere postular un nuevo rol para conexiones “conceptuales”; es simplemente que el rol de la observación disminuye.

El mismo punto subyacente puede hacerse en términos de condicionales contrafácticos en vez de adscripciones de posibilidades de grados de especificidad que varían. Por ejemplo, si Diógenes

no despluma el pollo, podríamos aún saber esta negación de un condicional contrafáctico:

- (6) No es el caso que si Diógenes hubiera tratado de desplumar el pollo, hubiera fracasado.

Por supuesto, (6) establece un hecho contingente. En algunas circunstancias posibles pero contrafácticas, si Diógenes hubiera tratado de desplumar el pollo, él *habría* fracasado. Sin embargo, mediante concebir vía imaginación qué habría pasado o podría haber pasado si Diógenes hubiera tratado de desplumar el pollo, a la luz del conocimiento de uno de los hechos contingentes de la situación real, uno podría llegar a saber (6). Pero (6) implica que es metafísicamente posible para Diógenes tratar de desplumar el pollo y no fallar, luego, es que es metafísicamente posible para él tener éxito en desplumar el pollo; al considerar esa posibilidad podemos llegar a conocer que es metafísicamente posible que exista un bípedo desplumado que no es humano, y por consiguiente que refute (1).

Es innecesario utilizar un hecho altamente contingente como (6). Podríamos usar (7) en reemplazo:

- (7) No es el caso que si Diógenes hubiera desplumado el pollo, una contradicción se hubiera dado.

Debido a que (7) implica que es metafísicamente posible para Diógenes desplumar el pollo y que la contradicción no hubiese sido el caso, es metafísicamente posible que Diógenes desplume el pollo; al considerar la posibilidad nuevamente llegamos a saber que es metafísicamente posible que exista un bípedo desplumado que no es un humano, y por consiguiente refutar (1). Ciertamente, (7) es de manera controvertida equivalente a la proposición de que es metafísicamente posible para Diógenes desplumar el pollo, lo cual es debatiblemente, una verdad metafísicamente necesaria (mediante el principio comúnmente aceptado de que si algo es posible metafísicamente, es metafísicamente necesario que sea metafísicamente posible).

Representa una diferencia notoriamente poco significativa al grado de persuasión de la refutación si va a través de (6) o de (7). Confiar en la observación de la situación de Diógenes para (6) pero no para (7) ni fortalece la refutación (al añadirle soporte “empírico”) ni la debilita considerablemente (al añadirle riesgo “empírico”). Una vez más, no requerimos de una facultad de intuición racional para saber (6), y para saber (7) no se necesita de una facultad extra con respecto a saber (6), por lo que no se necesita de una facultad racional para saber (7). De forma similar, las conexiones “conceptuales” no juegan papel en la refutación de (1) vía (7) que no hayan jugado en su refutación vía (6); es solo que la observación desempeña un papel en la primera refutación que no juega en la última.

En nuestra propia situación real, por supuesto, sabemos de millones de pollos desplumados y podemos imaginar a millones más, por lo cual la refutación de (1) está sobredeterminada, pero esto es un punto irrelevante. Incluso, si los pollos desplumados hubieran sido conocidos previamente, el empirista extremo aún se equivoca al insistir en que Diógenes debe completar el desplumado del pollo para refutar (1). Acciones infinitas son posibles, pero demasiado irrelevantes para haber sido llevadas a cabo. En muchos casos podemos saber que son posibles sin haber tenido que agarrar a alguien para llevar a cabo tales acciones. Por ejemplo, sabemos que es posible dar un discurso ante la Asamblea General de las Naciones Unidas mientras uno sostiene una piña, incluso si (supongo ciertamente) que nadie ha llevado a cabo tal acción, o lo hará en el futuro. De la misma manera, si reinterpretemos “bípedo desplumado” como significando algo así como “miembro de una especie cuyos miembros normalmente tienen solo dos patas y ninguna pluma”, y así que el pollo desplumado no cuenta como un “bípedo desplumado” en este nuevo sentido, aún podemos saber que podrían haber una especie de no humanos bípedos desplumados, por ejemplo, si los pollos evolucionaran para transformarse naturalmente y no tener plumas.

Sugiero el siguiente escenario. Al tomar decisiones, a menudo

tenemos que considerar nuestras opciones, y los peligros y oportunidades que enfrentaremos. Así, debemos embarcarnos en pensamiento modal acerca de una serie de variedad de posibilidades, la mayor parte de las cuales son contrafácticas. Tenemos que estimar cuál de estas podría realizarse con facilidad, y qué sucedería si se lo hiciese. Nuestra imaginación ha evolucionado al menos en parte debido a que nos ayuda a hacer tales estimaciones con algún grado de confiabilidad. Lo hace al capacitarnos para emplear habilidades en la aplicación de conceptos “offline” de acuerdo con situaciones hipotéticas, habilidades originalmente desarrolladas en la aplicación “online” de los mismos conceptos, dado cierto “input” perceptual. Ciertamente, nosotros típicamente usamos la imaginación en conjunción con algunos “inputs” perceptuales, pero como un caso límite de lo que podemos hacer sin esos mismísimos inputs —tal como nosotros aplicamos lógica deductiva a premisas que dependen de la observación, pero como un caso límite de que podemos aplicar lógica deductiva a un conjunto vacío de premisas, y por consiguiente derivar verdades lógicas. Tal como sucede con nuestras capacidades cognitivas, la confiabilidad de tales técnicas de imaginación se extienden sobre los casos prácticos significativos para los cuales primariamente evolucionaron, porque aquellos casos no forman una clase natural: no hay una manera factible de ser confiable acerca de estos sin ser confiable acerca de un rango de casos mucho más amplio. De esta forma, podemos usar nuestras técnicas de imaginación para pensamiento modal confiablemente, tal como en nuestras otras capacidades cognitivas, de propósitos teóricos para los cuales no evolucionaron. En particular, podemos usar dichas capacidades para investigar filosóficamente formas de posibilidad, menos específicas, que aquellas de las cuales usualmente nos preocupamos. La posibilidad metafísica es en su especificidad, la que es filosóficamente más importante de todas.

Por supuesto, la imaginación es solo uno de nuestros recursos de método de sillón. La lógica y la matemática son otros recursos

sobresalientes, usados en las ciencias; los críticos de los métodos de sillón a menudo soslayan la cuestión de cómo la lógica y la matemática se deben exceptuar de sus objeciones. Pero es la imaginación la que es un recurso más distintivamente filosófico.

The Philosophy of Philosophy desarrolla una explicación tal de forma más profunda, y se concentra en analizar lo que sucede en los experimentos mentales de los filósofos, uno de los métodos de sillón más destacables en la filosofía de hoy. Pero aún estamos en una etapa temprana de la exploración de este tópico.

3. Aunque queda mucho por explicar respecto de cómo la imaginación tiene éxito al jugar un rol confiable en la cognición, las consideraciones precedentes sugieren que desempeña un rol legítimo en la filosofía. Más aún, reemplazar su aplicación con otras que correspondan a métodos que no sean de sillón bien podría no aumentar la confiabilidad.

Considérese, por ejemplo, un argumento estándar en contra del utilitarismo. Uno construye un escenario imaginario en el cual castigar al inocente maximizaría la utilidad, juzga que tal escenario en el que se castiga al inocente aún sería incorrecto, y concluye que el utilitarismo es falso porque implica una predicción falsa acerca de qué debería hacerse en esa situación posible. ¿Podríamos construir un mejor argumento en contra del utilitarismo reemplazando el experimento mental moral por un experimento moral de la vida real, llevando a cabo la situación descrita por el escenario y castigando a una persona real, y así determinar que habría sido incorrecto hacerlo? Tal experimento moral real sería anti-ético sin ser filosóficamente más efectivo. Aunque reemplaza un método de sillón por uno que no es de sillón, pocos considerarían que el reemplazo como significativamente mejorando el grado de persuasión del argumento anti-utilitarista. Podría incluso hacer lo contrario, ya que las consideraciones de intereses personales [*self-interest*] o la presión de la opinión pública podrían tener un efecto distorsionador más fuerte en el juicio moral

en el caso “online” que en el caso “offline”. La distinción entre los métodos de sillón y los que no son de sillón parece ajena [*orthogonal*] a las más apremiantes dudas que uno podría tener acerca de *ambos* argumentos. Puesto que los utilitaristas de línea-dura responderán que el juicio de que es incorrecto castigar a una persona inocente, sea de modo “offline” u “online”, meramente manifiesta nuestra teoría falsa de la moral de carácter “folk”. El enfoque apropiado de esa duda estriba en la confiabilidad, o su ausencia, de nuestro juicio para aplicar nuestros conceptos morales, no en si lo hacemos desde el sillón, o desde fuera de este.

Más en general, sugiero, las objeciones al razonamiento de sillón en la filosofía son articulaciones típicamente inadecuadas de la desazón de los filósofos de aplicar en prácticas ordinarias conceptos “folk”, “online” y “offline”. Tales dudas peligran en convertirse en una nueva forma de escepticismo, un “escepticismo del juicio”. Podrían llevar a sus defensores a posiciones epistémicas que se auto refuten [*self defeating*]. Los escépticos del juicio sostienen que no tenemos garantía cuando dependemos de posiciones comunes en la aplicación de conceptos, porque carecen de apoyo adecuado independiente. Al elaborar los juicios negativos, parecen apoyarse en una práctica común de aplicación de conceptos epistemológicos como *no garantizado*. ¿Podría todo eso constituir una *reductio ad absurdum* de las prácticas de aplicación de conceptos epistémicos? El problema es que los escépticos de los juicios están en lo correcto cuando son renuentes a renunciar a las prácticas ordinarias de aplicación de conceptos epistemológicos en la ciencia natural. Por ejemplo, son renuentes a renunciar al juicio de que la teoría de la evolución está mejor fundamentada por la evidencia que el creacionismo, incluso si al hacer el juicio se apoyan en una práctica ordinaria de aplicar conceptos epistemológicos tales como *mejor apoyado por la evidencia*. Los escépticos del juicio requieren de alguna forma que su escepticismo resulte manejable dentro de límites.

Si la motivación para el escepticismo del juicio es simplemente

la carencia de independencia de apoyo para nuestras prácticas ordinarias, entonces es probable que se generalice en exceso en la forma ya descrita, porque cualquier supuesto apoyo independiente para nuestras prácticas corrientes de aplicación de conceptos tendrá ella misma que confiar en nuestra prácticas ordinarias de la aplicación de conceptos tales como *apoyo independiente*. Así, es mejor que la motivación para el escepticismo del juicio no sea que hay carencia de apoyo independiente para nuestras prácticas conceptuales ordinarias.

Recientemente, los defensores de la escuela llamada “filosofía experimental” tales como Stephen Stich han dado una motivación más positiva para una forma limitada de escepticismo del juicio. Argumentan que los datos experimentales socavan la confianza de los filósofos en ciertas prácticas conceptuales corrientes al mostrar que los juicios en cuestión son objeto de efectos distorsionadores [*framing effects*], por lo cual varían con factores que usualmente se piensa son irrelevantes a la verdad del juicio, tal como la inocuidad del ambiente en el cual se hace, el orden en el cual las preguntas se formulan, o el origen étnico de los acusados. No hay duda de que tales efectos distorsionadores a veces ocurren, y que pueden ser bastante perjudiciales a la práctica.

La pregunta es cuánto se extiende el perjuicio, y en qué direcciones. Los filósofos experimentales tienden a asumir sin argumento que lo que socava es la confianza en las prácticas ordinarias conceptuales con conceptos claves *en la filosofía de sillón*. Cuando se señala cuán insignificante puede resultar la diferencia que puede haber entre un experimento mental con un concepto ordinario y un experimento correspondiente de la vida real con el mismo concepto, reaccionan extendiendo su escepticismo a aplicaciones tanto “online” como “offline”. Pero, entonces se incrementa fuertemente la presión para que expliquen por qué su escepticismo no debería aplicarse a sus propias aplicaciones “online” de los conceptos epistemológicos ordinarios, al argumentar a favor de su escepticismo. Peor, no han proporcionado evidencia de que los juicios hechos *en filosofía* constituyen una clase

psicológica, y la evidencia de la última sección sugiere que no lo hacen. Entonces, es puramente tendencioso para ellos extender su escepticismo a la confianza en prácticas ordinarias de aplicar conceptos epistemológicos en filosofía, o al menos en el tipo de filosofía que ellos no gustan, pero no a la confianza en las prácticas ordinarias de aplicación de conceptos epistemológicos —tal vez los mismísimos conceptos epistemológicos— en la ciencia, tal como en el juicio de que la teoría de la evolución está mejor apoyada por evidencia que el creacionismo. Después de todo, la voluntad a hacer un juicio varía fuertemente con factores independientes de su verdad, tales como el origen étnico o las creencias religiosas del acusado. Hasta ahora los filósofos experimentales no han ejecutado los experimentos que les autorizan a tratar la distinción entre filosofía y no filosofía como inclusive relevante al problema. Simplemente se han concentrado en resultados en la filosofía, sin aplicar el control necesario mediante el cotejo de resultados similares fuera de la filosofía. Es como si hubieran testeado los efectos distorsionadores por prejuicios en la mujer, los hubieran encontrado, y hubieran decidido que el problema socava la confianza de los juicios de la mujer, sin haber testeado los efectos distorsionadores en el hombre.

Los efectos distorsionadores se dan en la filosofía, y son una fuente de error. Pero podrían ocurrir en *todas* las disciplinas académicas, y producir error en todas estas. Necesitamos saber mucho más sobre dónde son peores, cómo evitarlos, cómo se eluden mediante la estructura normal de una disciplina académica, mediante el entrenamiento, la retroalimentación, la evaluación de doble ciego, el alcance de consenso a través de debate público que involucra a muchos individuos en un periodo extenso de tiempo, y así sucesivamente. Los resultados experimentales de seguro arrojarán luz sobre estos problemas. Pero la investigación empírica no ha ido muy lejos aún, y está actualmente siendo distorsionada por un ánimo preconcebido en contra de la filosofía de sillón con la cual se desarrolla— incluso por aquellos que conceden que la distinción entre juicios de sillón

y de no sillón no es lo que importa. Una meta más productiva para la investigación empírica es mejorar nuestras maneras actuales de mitigar los efectos distorsionadores de juicios tanto dentro y fuera del sillón, en vez de desarrollar un programa de inspiración inadecuada para eliminar la filosofía de sillón. La filosofía puede contribuir mejor a las ciencias al refinar sus propios métodos distintivos intelectuales, no al abandonarlos e imitar los de otras disciplinas.

Referencias

- Austin, J. (1961): “The meaning of a word”. En: J. Urmson y G. Warnock (eds.) *J.L. Austin Philosophical Papers*. Oxford: OUP, pp. 55-74.
- Diógenes Laercio (s. III d.c.): *The Lives and Opinions of Eminent Philosophers*. *Forgotten Books*.
- Derrida, J. (1974): *Of Grammatology*. Baltimore, Maryland: The Johns Hopkins University Press.
- Williamson, T. (2007): *The Philosophy of Philosophy*. Oxford: Blackwell.

¿Hay lugar en epistemología para la filosofía de sillón?⁴⁸

Hilary Kornblith

Quine, desde luego, argumentó en favor del rechazo de la noción misma de lo *a priori*⁴⁹, y en 1969, en su “Epistemology Naturalized”, argumentó a favor de una aproximación general derechamente empírica respecto de las cuestiones epistemológicas (Quine 1969). Sin embargo, lo que precisamente tal programa podría ser fue un asunto muy disputado, tanto entre aquellos que anhelaban rechazarlo como entre los que adscribían con entusiasmo a él. En las décadas de 1980 y 1990, en gran medida debido a la influencia de Alvin Goldman (1978, 1986 y 1992), la posibilidad de una aproximación cabalmente empírica a la epistemología se puso una vez más al frente de la discusión, y la cuestión de si acaso la especulación epistemológica podría ser abordada *a priori* de modo provechoso, y qué, si acaso en algo, podrían contribuir las ciencias empíricas a la epistemología, fueron cuestiones bastante debatidas.⁵⁰ Aquellos que querían defender aproximaciones tradicionales a las cuestiones epistemológicas acostumbraban reconocer la relevancia de los resultados empíricos para ciertos tópicos acerca del conocimiento, pero luego pasaban enseguida a minimizar la importancia de este hecho. Richard Feldman, por ejemplo, es un caso típico en este respecto. Él comenta:

⁴⁸ Traducción y publicación con permiso del autor.

⁴⁹ Véase especialmente “Truth by convention” (publicado originalmente en 1935) y “Carnap and logical truth” (publicado originalmente en 1954), ambos en Quine (1951 y 1961).

⁵⁰ Véase también los artículos reunidos en Kornblith (1994), y mi propia postura en Kornblith (1993).

Como es obvio, el trabajo empírico es relevante para “el estudio del conocimiento humano”. Pero, esto logra mostrar su relevancia para la epistemología solo si esta es tan amplia como el estudio del conocimiento humano. El estudio completo del conocimiento humano incluiría, probablemente, estudios históricos de lo que la gente solía saber; estudios en neurociencias concernientes a los modos en que el cerebro procesa la información, estudios sociológicos acerca de los modos en que el conocimiento es transmitido en las sociedades, etc. Mientras algunos filósofos pueden que piensen que ellos, desde sus sillones, tienen algo que decir sobre muchos de estos tópicos, ninguna persona sensata podría pensar que tales investigaciones puedan resultar exitosas sin el *input* científico. Así, dada tan amplia interpretación de lo que cuenta como epistemología, es difícil imaginar algún desacuerdo con la posición relativa a la verdad del naturalismo metodológico. (Feldman 1999, p. 171).

Como Feldman a continuación sugiere, sin embargo, la epistemología no necesita ser entendida como tan cercana a una empresa tan amplia, y, como insistía, uno puede no necesitar del *input* experimental una vez que los asuntos epistemológicos han sido concebidos de un modo más acotado y plausible.

Aquí, la respuesta de Feldman, la que no era en absoluto atípica⁵¹, involucra dos puntos que merecen ser distinguidos. En primer lugar, está la cuestión acerca de la amplitud de la epistemología propiamente tal. Pero, en segundo lugar, Feldman ya no habla acerca de si acaso la especulación epistemológica es *a priori*. En lugar de eso, él habla de si acaso la epistemología puede ser llevada a cabo provechosamente desde el sillón. Esta es una distinción importante.

⁵¹ Para observaciones de esta misma índole acerca de la importancia del externalismo, y del modo en el cual la información empírica es, de este modo, traída a desempeñar un rol protagónico en epistemología, véase BonJour (2001).

Así, por ejemplo, en el último libro de Timothy Williamson (2007), *The Philosophy of Philosophy*, se propone defender un enfoque de sillón para las cuestiones filosóficas. Sin embargo, como él es enfático en señalar, esto no significa que la información empírica no sea relevante. En efecto, Williamson rechaza la distinción a priori/a posteriori, aunque por razones significativamente diferentes a las de Quine.⁵² Más aún, Williamson se esmera en rechazar lo que denomina “excepcionalismo filosófico”, esto es, la posición que sostiene que los métodos apropiados en filosofía son diferentes en tipo a otras clases de investigación. No obstante, Williamson, como Feldman, quiere defender la integridad de la filosofía como una disciplina de sillón. Cuando teorizamos desde el sillón, nuestras creencias empíricas jugarán propia e inevitablemente algún rol no-trivial en las conclusiones a las que lleguemos. Sin embargo, hay una diferencia real entre la especulación de sillón y el trabajo netamente empírico que uno ve, por ejemplo, en la psicología experimental o en la física experimental. Williamson, como Feldman, está interesado en defender la legitimidad de los métodos de sillón en filosofía en general, y especialmente en epistemología. De este modo, la cuestión acerca del estatus de la especulación epistemológica parece haber cambiado. Muchos defensores de los métodos tradicionales en epistemología ya no buscan defender a estos métodos tradicionales como métodos a priori; en lugar de ello, es la integridad de la especulación desde el sillón la que es ahora vista como la genuina cuestión.

Unatendenciamás en la filosofía reciente debiese de ser mencionada, a modo de introducción. Una de las principales alternativas a los métodos de sillón, en los últimos años, es el reciente y emergente campo de la así llamada “filosofía experimental”. Digo “así llamada” no porque crea que el rótulo no sea apropiado, ni mucho menos con el fin de poner, de alguna manera, en entredicho este enfoque, sino más bien para indicar que este rótulo es un nombre y no una descripción. La filosofía experimental no es meramente una aproximación a las

⁵² Véase Williamson (2007, capítulos 3 y 4).

cuestiones filosóficas que intenta hacer uso y apoyarse en el trabajo experimental de la clase que sea. Más bien, este enfoque, defendido en trabajos hechos por Stephen Stich, Jonathan Weinberg, Joshua Knobe, Shaun Nichols, y por muchos otros, busca arrojar luz en cuestiones filosóficas, usualmente, mediante cuestionarios que exploran nuestras concepciones *folk* relativas a múltiples nociones filosóficas claves.⁵³ Esto es, sin lugar a dudas, una aproximación experimental a cuestiones filosóficas, pero no es en modo alguno el tipo de trabajo empírico que fue sugerido por Quine o Goldman como para formar la base para una epistemología científica propiamente tal. Gran parte de la discusión reciente en torno a los métodos de sillón se ha centrado, a diferencia de los períodos anteriores en los cuales este asunto fue ampliamente discutido, en la pregunta de si acaso es probable que el enfoque de los filósofos experimentales arroje luz sobre cuestiones de real importancia filosófica.

Permítaseme poner mis cartas sobre la mesa. Argumentaré aquí que es improbable que los métodos de sillón arrojen verdadera luz sobre cuestiones centrales en epistemología. Sin el *input* directo y sustantivo del trabajo experimental, argumentaré, es posible que permanezcamos ignorantes de, o mal entendamos, la naturaleza misma de los problemas epistemológicos y su eventual solución. Aquí, la verdadera fuente de iluminación, como lo veo, no es probable que venga del trabajo de los filósofos experimentales, sino más bien del trabajo más tradicional en ciencias cognitivas, trabajo no dirigido a nuestras concepciones *folk* acerca de nada. Mi disconformidad relativa al trabajo de los filósofos experimentales no será, desde luego, una objeción a la idea de que el trabajo empírico puede que arroje luz sobre los problemas filosóficos, sino más bien que gran parte (aunque no todo) del trabajo que los filósofos experimentales han hecho y quieren hacer está mal encaminado, es más, está mal encaminado de manera bastante similar, creo, en que lo está gran parte de la especulación de sillón en filosofía. De este modo, argumentaré en defensa de una

⁵³ Véase, por ejemplo, Weinberg, Nichols y Stich (2001).

aproximación a la epistemología informada a cabalidad de modo empírico, pero una que es fundamentalmente diferente a la intentada por los filósofos experimentales.

I

Mientras las cuestiones relativas a la relevancia de la psicología para la epistemología han sido ampliamente discutidas en la literatura reciente, estas no eran en absoluto objeto de discusión entre los filósofos de los siglos XVII y XVIII. Descartes, Locke, Hume y Kant, por ejemplo, probablemente no habrían entendido la pregunta de si acaso ellos estaban comprometidos con la especulación filosófica o la especulación psicológica. La disciplina de la psicología no emergió como una disciplina independiente sino hasta finales del siglo XIX, pero, más allá de eso, las posiciones epistemológicas de todos estos filósofos estaban claramente influenciadas por sus posiciones en psicología. Para ellos, una epistemología que no estaba informada por una teoría de la mente habría sido inconcebible.

Desde luego, la concepción de la mente que estos filósofos tenían no estaba informada por los resultados experimentales en psicología, ya que en ese entonces no había tal cosa como la psicología experimental. Más bien, todos estos filósofos sostenían que la introspección y la reflexión juntas proporcionaban la clave para la comprensión de la mente, y de esta forma, incluso si las creencias a las que se llega mediante introspección no están justificadas de manera *a priori*,⁵⁴ el uso de la introspección es claramente un método de investigación de sillón. Si podemos comprender totalmente desde el sillón las características de la mente que se requieren para una epistemología totalmente informada, entonces los métodos de sillón pueden bien ser todo lo que se requiere una vez que uno se vuelca más directamente hacia las cuestiones epistemológicas mismas.

⁵⁴ Al menos en el caso típico. Hay un desacuerdo real acerca de si ciertos juicios a los que se ha llegado mediante introspección están justificados *a priori*. Para discusión una particularmente útil de este asunto véase Kitcher (1980).

Estos filósofos estaban en lo correcto, creo, al pensar que una comprensión del modo en que la mente funciona es un prerequisite para la construcción de una epistemología razonable, pero estaban equivocados, desde luego, al pensar que somos capaces de comprender las características relevantes de la mente mediante la investigación de sillón. No es tan solo que el trabajo experimental en psicología se ha mostrado imprescindible para tapar vacíos menores en la visión de la mente que la introspección nos proporciona, o que el trabajo experimental se necesita para corregir algunas pequeñas imperfecciones alrededor de los cantos, por decirlo así, del cuadro ampliamente preciso que de otro modo proporcionan los métodos de sillón. Al contrario, ha resultado que en cada materia, punto por punto, no podemos siquiera comenzar a comprender cómo funciona la mente sin un exhaustivo trabajo experimental, y la imagen de la mente que nos proporcionan los métodos de sillón es, de principio a fin, fundamentalmente defectuosa. Consideremos ahora una serie de asuntos que se mostrarán especialmente importantes cuando pasamos de la naturaleza de la mente a la epistemología.

Descartes sostuvo, célebremente, que la introspección nos proporciona una imagen de nuestros estados mentales actuales que es infalible y completa; en una palabra, la mente es completamente transparente a sí misma. Tal teoría ya no es más defendida en estos días, sin embargo, la extraordinariamente rica imagen de nuestros estados y procesos mentales que la introspección nos da es una parte importante de la concepción de sillón de la mente. Considérese, por ejemplo, el modo en el cual los problemas epistemológicos son a menudo introducidos. Chisholm sugiere, en una discusión que es en múltiples aspectos bastante típica, que deberíamos involucrarnos en el cuestionamiento socrático. Practique la introspección, y detectará un gran número de sus creencias. Escoja una de éstas y pregúntese a sí mismo: “¿Por qué sostengo esta creencia particular?”. Al menos en el caso típico, notará que la creencia-objetivo inicial es una que sostiene por varias razones, esto es, la sostiene todavía sobre la base de otras

creencias. Y entonces nos podremos preguntar acerca de cada una de éstas por qué la sostiene. Puede que continuemos con este proceso, argumenta Chisholm, hasta que alcancemos creencias que brinden a la investigación algún tipo de parada apropiada.⁵⁵

Dejemos de lado la defensa de Chisholm del fundacionalismo. Cuando nos involucramos en el tipo de cuestionamiento socrático que él recomienda, la introspección no sólo nos proporciona una idea acerca de nuestras creencias, sino que nos proporciona una idea acerca del origen de nuestras creencias. De este modo, cuando considero una creencia mía *A*, y me pregunto por qué es que la sostengo, la introspección parece revelarme por qué es que actualmente creo *A*: lo creo, por decirlo así, porque creo *B* y *C*. Por rica y poderosa que pueda ser la visión introspectiva del origen de nuestras creencias, podríamos preguntarnos cuán certeras son las impresiones que comunica la introspección. Este es un problema que los psicólogos sociales han investigado con algún detalle.

Gran parte de lo que la introspección parece revelar, cuando nos involucramos en este tipo de ejercicios, asemeja no ser mucho más que confabulación.⁵⁶ Lo que en la introspección parece ser la aprehensión directa de relaciones causales entre nuestros estados mentales es, en el fondo, el resultado de un proceso de reconstrucción racional: estamos de hecho involucrados en un proceso subconsciente acerca de cuál debe haber sido el origen de nuestras creencias. Aún más, este proceso de reconstrucción no es, en una amplia gama de casos, estremecedoramente certero. Una buena parte de nuestras creencias está formada de maneras que no son en modo alguno razonables, sin embargo, parece que al reconstruir lo que eran nuestras razones, nos valemos de una premisa menor crucial: que sea lo que sea que creamos,

⁵⁵ Según mi parecer, la mejor presentación de Chisholm de esta manera de pensar acerca de las cuestiones epistemológicas se puede encontrar en la segunda edición de su *Theory of Knowledge* (Chisholm 1977, cap. 2).

⁵⁶ Véase, por ejemplo, Wilson (2002).

probablemente llegamos a creerlo sobre la base de buenas razones⁵⁷. Así, no debe de parecer ninguna sorpresa que, cuando nos volvemos reflexivos y escudriñamos las razones por las cuales sostenemos variadas creencias, en general resulta que encontramos que nuestras razones son lo suficientemente adecuadas. Lo que parece, desde el punto de vista de la introspección, como una responsable revisión extra de nuestro irreflexivo proceso de adquisición de la creencia de primer orden, resulta, en efecto, no ser mucho más que pura autocomplacencia. Pero, esto no es algo que podamos decir desde el sillón. Hay, en efecto, una poderosa ilusión que surge precisamente del uso de métodos de investigación de sillón. A menos que nos levantemos y salgamos del sillón, somos proclives a mal entender los orígenes de nuestras creencias.

Gran parte de nuestra adquisición de la creencia es irreflexiva. Por ejemplo, cuando entro en mi oficina, inevitablemente llego a creer que hay un escritorio enfrente de mí. No me detengo a reflexionar acerca de si acaso tal creencia, dada la evidencia que tengo, está justificada. Y en este respecto, esta creencia mía particular no es inusual; la mayoría de mis creencias está formada de modo irreflexivo. Es más, esto no es una peculiaridad mía; la vasta mayoría de las creencias están formadas irreflexivamente. No obstante, sí nos detenemos, a veces, a reflexionar y preguntarnos qué es lo que deberíamos creer. Cuando actuamos así, tenemos la clara impresión de la agencia epistémica. Formar creencias, al menos en estas ocasiones reflexivas, parece ser algo que hacemos más que algo que meramente nos ocurre.⁵⁸

¿Hasta qué punto esta impresión de agencia debe ser tomada en sentido literal? Con el fin de responder a esta pregunta, necesitamos salir del sillón y atender al trabajo experimental en psicología social. En general, nuestros juicios acerca de nuestra propia agencia no

⁵⁷ He argumentado con cierto detalle en favor de esto en el capítulo 4 de Kornblith (2002).

⁵⁸ Una serie de autores han adjudicado gran importancia a este hecho. Véase, por ejemplo, Korsgaard (1996, capítulo 3) y Moran (2001).

son tremendamente confiables. Las personas emocionalmente sanas tienden a tener una impresión exagerada de su propia eficacia en múltiples ámbitos de la vida.⁵⁹ Además, la idea de que hay una asimetría profunda con respecto a la agencia cuando se trata de la diferencia entre la adquisición de la creencia reflexiva e irreflexiva es una tal que no resiste el examen cuidadoso. Mientras la adquisición de la creencia irreflexiva está, por supuesto, mediada por un montón de procesos mentales que operan en el nivel sub-personal de maneras de las que solemos ser ignorantes,⁶⁰ lo mismo es verdadero de modo bastante claro de la adquisición de la creencia reflexiva. No estamos más al tanto, mediante introspección, de la completa gama de procesos mentales que operan en la adquisición de la creencia reflexiva que lo estamos en el caso de la irreflexiva. Por último, el hecho de que nuestros juicios introspectivos acerca del origen de nuestras creencias resulten en gran medida ser una cuestión de racionalización después del hecho más que monitoreo auténticamente simultáneo muestra, a lo menos, que los dictámenes de la introspección exageran enormemente hasta qué punto y en qué grado jugamos un rol activo en la formación de nuestras creencias cuando dejamos de reflexionar.⁶¹ Una comprensión precisa del grado y alcance en que somos agentes genuinos en la adquisición de la creencia reflexiva no puede ser alcanzada desde el sillón.

El funcionamiento de algunos de los procesos más fundamentales subyacentes a la cognición no son solo invisibles para la introspección, sino que estos funcionan de modos que son sorprendentemente diferentes de la concepción de sillón que tenemos de los mismos. Así, por ejemplo, la operación de la memoria deja bastante poco a través de la huella introspectiva. Cuando la memoria está operando con gran eficiencia, puede que estemos al tanto de estar formando una creencia relativa a la memoria, pero nada más que la creencia misma

⁵⁹ Véase, por ejemplo, Taylor (1989).

⁶⁰ Sosa plantea este punto en un intento de apoyar la idea de que “la reflexión ayuda a la agencia”. Véase en Greco (2004, p. 292).

⁶¹ He discutido estos asuntos con mayor detalle en “The myth of epistemic agency”, manuscrito.

está disponible a la introspección. La visión de sillón de la memoria es sencillamente que obtenemos creencias inicialmente por otros medios —por ejemplo, mediante la percepción— y luego estas creencias son almacenadas en la memoria; posteriormente, en las ocasiones convenientes, aquellas son extraídas del almacenamiento y hechas disponibles para la introspección. Pero este sencillísimo cuadro del funcionamiento de la memoria es por completo impreciso. El proceso inicial de codificación, el proceso de almacenamiento, y el proceso de recuperación involucran todos bastante construcción y reconstrucción. La memoria humana genuina no es en absoluto como el proceso de tomar una foto e ingresarla a un archivo, sólo para sacarla del archivo en ocasiones subsiguientes cuando puede que se necesite.⁶²

Prácticamente lo mismo podría ser dicho acerca de nuestra receptividad al testimonio. La introspección revela poco, si es que algo, acerca del proceso cognitivo que tiene lugar cuando un hablante afirma que *p* y nosotros, inmediatamente después, llegamos a creer que *p*. La especulación de sillón acerca del proceso ha llevado a una serie de explicaciones, algunas de las cuales sugieren que aquí, en el caso típico, de hecho ocurre bastante poco; la transmisión de la información no involucra más que la “preservación del contenido”.⁶³ Según otras teorías,⁶⁴ hay una cierta cantidad de monitoreo que se lleva a cabo para chequear, por ejemplo, la sinceridad y capacidad del hablante para saber de lo que él o ella habla. No obstante, la literatura empírica sobre tal monitoreo revela aspectos de la situación que los métodos de sillón no podrían siquiera comenzar a detectar.⁶⁵ En efecto,

⁶² Christensen y yo hemos discutido esto (Christensen y Kornblith 1997).

⁶³ La frase proviene de Burge (1993). Debería aclarar que no es la visión de Burge que, en el caso típico, muy poco sucede más allá de la preservación del contenido.

⁶⁴ Véase, por ejemplo, Fricker (1994) y Lackey y Sosa (2006). Como es característico de este cuerpo de literatura, los artículos de la última antología no se involucran con la literatura empírica relevante sobre el tema.

⁶⁵ Aquí me apoyo en Michaelian, “In defence of gullibility: The epistemology of testimony and the psychology of deception detection” *Synthese*, próxima aparición. También me ha ayudado la lectura de un manuscrito sobre este tema de Shieber.

monitoreamos testificadores con respecto a la sinceridad sobre la base de una variedad de marcas observables, y hay, de hecho, una serie de marcas conductuales que son indicadores confiables de insinceridad; desafortunadamente, tendemos a confiar en marcas que no son las que indican de modo confiable si el testificador es auténticamente sincero. Las situaciones psicológicas y epistemológicas son aquí altamente complejas,⁶⁶ y ellas son complejas de una manera que la introspección y los métodos de sillón no pueden revelar.

El contenido y formato de nuestros conceptos juega un rol importante en una variedad de asuntos epistemológicos, pero el carácter de nuestros conceptos no puede ser determinado y resuelto desde el sillón.⁶⁷ Según la visión clásica de los conceptos, ellos están representados por un conjunto de condiciones necesarias y suficientes de modo muy similar al formato de los análisis filosóficos tradicionales. Así, por ejemplo, el concepto de un pájaro podría ser representado como un animal con un pico, alas y dos patas. Siguiendo el trabajo de Eleanor Rosch,⁶⁸ se ha sugerido que el formato en el cual los conceptos son representados, no obstante, podría ser bastante diferente. En lugar de un conjunto de condiciones necesarias y suficientes en el formato estándar, un concepto podría representarse mediante un caso prototípico, en conjunto con una colección de dimensiones de similaridad tales que cualquier objeto lo suficientemente similar al caso prototípico caiga bajo el concepto.⁶⁹ Y hay también otras explicaciones.⁷⁰ En efecto, no es en absoluto claro que nuestros

⁶⁶ Nuevamente, véanse los artículos de Michaelian y Shieber.

⁶⁷ Para una discusión de estos asuntos, véase Goldman (2007) y Kornblith (2007).

⁶⁸ Véase especialmente "Principles of categorization" de Rosch y Lloyd (1978).

⁶⁹ Desde luego, es un error afirmar que este tipo de formato para la representación de conceptos no proporciona condiciones necesarias y suficientes para caer bajo el concepto. Cualquier explicación de lo que los conceptos son inevitablemente proporcionará condiciones necesarias y suficientes. Lo que en este punto está en juego es si acaso aquellas condiciones necesarias y suficientes son presentadas en el tipo de formato característico de los análisis filosóficos tradicionales.

⁷⁰ Para una colección útil de artículos sobre estos asuntos véase Margolis y Laurence (1999).

conceptos estén representados en un solo formato. No sólo diferentes conceptos podrían representarse en diferentes formatos; uno y el mismo concepto podría representarse bien en más de un formato.⁷¹ Como sea que los conceptos sean representados, es perfectamente claro que no podemos determinar el formato de su representación mediante la introspección. Sólo el trabajo experimental puede ser de utilidad al tratar con este asunto.

Cuestiones relativas a la estructura de gran escala de la mente han sido el centro de importante trabajo filosófico desde los tiempos de Platón hasta la actualidad. Aquí, como en otras partes, la especulación de sillón ha sido reemplazada por trabajo que está informado empíricamente. Asuntos relativos a la modularidad, e inquietudes similares relativas a la inferencia específica de dominio, sólo pueden ser provechosamente tratados mediante trabajo experimental. A estas alturas, cómo parezcan las cosas desde el punto de vista de la primera persona o cómo parezcan desde el sillón no es más relevante para estas investigaciones que lo es para una exploración de los mecanismos involucrados en la digestión.

Lo mismo es verdadero acerca del rol de la emoción en la cognición. La introspección nos da una rica impresión de nuestros estados emocionales y sus efectos. Las visiones de sillón acerca de estos asuntos que están basadas sobre este cuadro introspectivo pueden ser de interés para los antropólogos cognitivos o sociólogos, pero aquellas son sencillamente irrelevantes para una investigación acerca de los efectos reales de la emoción en la cognición. Si queremos comprender cómo funciona realmente la mente, nuestra concepción introspectiva de las cosas no solo no es la última palabra sobre estas materias; ya no es ni siquiera considerada como la primera. Una investigación apropiada de estos fenómenos no comienza con nuestra visión introspectiva de los mismos, o nuestra visión de sillón, para pasar a continuación a hacer correcciones donde resulte necesario. La

⁷¹ Véase Horvath (1998), por ejemplo.

visión de sillón sobre estas materias es sencillamente irrelevante para una investigación seria acerca del funcionamiento de la mente.

II

¿Tiene esto algo que ver con la epistemología? La comprensión de cómo funciona la mente, podríamos dar por sentado, no se puede alcanzar desde el sillón. Al mismo tiempo, la epistemología no es una simple investigación acerca de cómo es que opera la mente. Incluso si concedemos, entonces, como por supuesto debemos, que la psicología no puede ser llevada a cabo de manera provechosa desde el sillón, ¿qué relación tiene esto con la metodología de la epistemología?

Comenzaré a contestar estas preguntas mediante ejemplos. Argumentaré que una serie de asuntos centrales para la epistemología no pueden ser tratados de modo provechoso sin la clase de orientación y restricción que ofrece el trabajo experimental. Esto, desde luego, dejará abierta la pregunta de si acaso hay otros asuntos dentro de la epistemología que podrían ser tratados de modo provechoso sin la ayuda de trabajo experimental. Algo también tendré que decir acerca de esta cuestión.

Considérese, entonces, el debate en epistemología entre internistas y externistas.⁷² Internistas y externistas discrepan con respecto a los determinantes de la justificación: las características que vuelven a una creencia justificada, insisten los internistas, deben ser, en algún sentido importante, internas al agente; los externistas niegan esto. Así, por ejemplo, los internistas del acceso insisten en que las características que vuelven a una creencia justificada deben ser accesibles para la introspección o reflexión. Algunos internistas, como Laurence Bonjour, argumentan que las creencias que están justificadas en el sentido internista son más proclives a ser verdaderas que las creencias que no consiguen satisfacer el estándar internista (Bonjour 1985, p. 8). Una vez que uno tiene sobre la mesa una explicación internista

⁷² Para una colección de algunos de los artículos centrales sobre este asunto, véase Kornblith (2001).

de la justificación, por decirlo así, el requisito de BonJour produce la aseveración empírica de que: las creencias que satisfacen estos estándares son más proclives a ser verdaderas que las creencias que no lo hacen. Dado que hay bastante investigación empírica sobre lo que ocurre cuando hacemos introspección y reflexionamos, este trabajo experimental se hace relevante para examinar la afirmación de BonJour. Y tal como tenemos una visión de sillón acerca de lo que ocurre cuando nos involucramos con la introspección y la reflexión, esta visión de sillón, como he señalado más arriba, no es en absoluto precisa. Así, con el fin de examinar si acaso la afirmación de BonJour es realmente verdadera, y no solo si parece verdadera desde el sillón, necesitamos echar un vistazo serio a la investigación empírica. La visión desde el sillón nos da un cuadro por completo erróneo.

Ahora bien, BonJour es, la verdad, un internista inusual en tanto que quiere identificar la propiedad de estar justificado con una propiedad internamente reconocible, mientras, al mismo tiempo, insistir en que las creencias que tienen esta propiedad son proclives a ser verdaderas. No todos los internistas quieren hacer esta segunda afirmación, una que equivale a insistir en que las creencias internamente justificadas son producidas de modo confiable. Pero, la investigación empírica sobre los mecanismos involucrados en la introspección y reflexión no es, por esta razón, del todo irrelevante para las afirmaciones de estos otros internistas. Si, como he argumentado, la literatura empírica muestra que los procedimientos mismos que los internistas consideran como una clase de revisión de los procesos naturales y genuinos [*unselfconscious*] de adquisición de creencia no llevan realmente a cabo de modo efectivo ninguna revisión tal, entonces la motivación para identificar la justificación con tal monitoreo interno está, por lo menos, severamente comprometida. El internismo obtiene una aparente plausibilidad a partir del hecho de que una visión de sillón de la introspección y la reflexión genera la ilusión de que tales procedimientos de revisión son auténticamente efectivos.

Lo mismo es verdadero, por supuesto, de buena parte de la

discusión en torno a la responsabilidad epistémica. Se hacen una serie de afirmaciones acerca de qué es lo que un agente responsable debe hacer a fin de que se vuelva probable que sus creencias sean verdaderas, pero la mayoría de la literatura sobre responsabilidad epistémica no es más que especulación de sillón. La literatura empírica sobre los efectos reales de los mismísimos procedimientos de revisión que los teóricos de sillón insisten en que se constituyen como conducta epistémicamente responsable puede ser bastante saludable.⁷³ Desde luego, siempre está abierto para estos teóricos insistir en que ellos no están identificando la adquisición de creencia epistémicamente responsable con la adquisición de la creencia confiable, y, en efecto, que ellos no están comprometidos con la visión de que la creencia responsablemente adquirida tiende a ser confiable. Todo esto es verdad, pero no muestra en modo alguno que los resultados empíricos son irrelevantes para la evaluación de afirmaciones relativas a la responsabilidad epistémica. A lo menos, los resultados empíricos ponen una cierta presión sobre estos teóricos, ya que los procedimientos mismos en que ellos insisten son definitorios de la conducta epistémicamente responsable resultan hacer, bajo el examen empírico, poco más que aumentar la confianza de agentes ya ignorantes [*benighted*]. Dependiendo de los detalles de la visión particular acerca de la responsabilidad epistémica, los resultados empíricos no necesitan constituir ningún tipo de refutación concluyente de aquellos, sin embargo, sería sin duda un error concebir tal información como irrelevante para una completa evaluación de estas afirmaciones. Los resultados son profundamente sorprendentes, y lo son porque contradicen nuestra visión de sillón sobre estas materias. Pero, fue justamente la visión de sillón la que en un primer momento hizo aparecer como plausible las afirmaciones acerca de la responsabilidad epistémica.

Hay una larga tradición en epistemología con respecto a ofrecer

⁷³ Por primera vez traté este asunto en Kornblith (1989). Además del capítulo cuatro de Kornblith (2002), véase también Bishop (2000) y Bishop y Trout (2005).

consejos epistémicos, esto es, sugerencias acerca de cómo es que uno debe proceder para alcanzar las creencias de uno si es que la meta es obtener creencias verdaderas. Hablar de “reglas para la dirección de la mente” fue otrora común, y, en años recientes, los epistemólogos han discutido con algún detalle hasta qué punto una explicación de la creencia justificada debe de dar algún tipo de orientación para el agente epistémico respectivo. Al igual que con la discusión de la responsabilidad epistémica, es bastante claro que la literatura empírica sobre los efectos de la introspección, reflexión y varias estrategias inferenciales es directamente relevante aquí. Y, nuevamente, precisamente porque este cuerpo de literatura ha producido con regularidad resultados sorprendentes —resultados que contradicen directamente el modo en que las cosas parecen desde el sillón— resulta claro que estos problemas no pueden ser tratados de modo provechoso con maneras que no están informadas por el trabajo experimental.

Muchos filósofos creen que el método apropiado en epistemología, y más generalmente en filosofía, es una cuestión de análisis conceptual, y el famoso contraejemplo de Ed Gettier (1963) a la sugerencia de que el conocimiento podría ser identificado con la creencia verdadera justificada es visto como un modelo de cómo es que tal análisis conceptual podría ser llevado a cabo. Según este planteamiento, se estima que el objeto del análisis filosófico es un concepto, y los métodos de sillón se consideran como los medios para elucidar el contenido de nuestros conceptos.⁷⁴ De este modo, hay un vasto cuerpo de literatura filosófica que intenta dar una explicación del contenido de varios conceptos filosóficamente interesantes. Hay, además, un formato estándar en el que estos análisis proceden, a saber, proporcionando condiciones necesarias y suficientes de alguna clase. Hay, además, un vasto cuerpo de literatura en psicología que investiga el formato y contenido de nuestros conceptos, y esta literatura ha producido un cuerpo de interesantes y sorprendentes resultados. El formato en el

⁷⁴ Para un defensa de este enfoque, véase, por ejemplo, Jackson (1998).

cual nuestros conceptos son mentalmente representados parece ser bastante diferente del formato de los análisis filosóficos tradicionales, y el contenido de nuestros conceptos es también notablemente diferente del modo en que aparece desde el sillón. ¿Cómo debemos de considerar la relación entre estas dos investigaciones? ¿Cómo, si es que lo hace, se relaciona la investigación experimental de nuestros conceptos con la investigación de sillón?

Alvin Goldman ha argumentado que la investigación de sillón es un tipo de versión proto-científica del trabajo experimental en que se ocupan los psicólogos (Goldman 2007). Él argumenta que, de hecho, los métodos de sillón resultan ser bastante confiables al momento de darnos información acerca del contenido de nuestros conceptos, aunque el trabajo experimental podría ciertamente otorgar una revisión necesaria de nuestros resultados, y, en ocasiones, una fuente de corrección para la visión de sillón. De acuerdo con este enfoque, los métodos de sillón nos dan una explicación bastante precisa del contenido de nuestros conceptos, pero aquellos deben ser complementados por el trabajo empírico. Otros son menos optimistas en relación con el alcance y grado en el que los métodos de sillón se aproximarán aquí a los resultados conseguidos por la experimentación cuidadosa,⁷⁵ y esto ha llevado a los filósofos experimentales a sugerir que la tarea del análisis conceptual requiere ser llevada a cabo, no ya desde el sillón, sino desde el laboratorio psicológico. No obstante, la diferencia aquí entre Goldman y los filósofos experimentales es sólo una cuestión de grado. Ellos están en total acuerdo con respecto a que el objeto del análisis filosófico y el objeto de la investigación experimental son uno y el mismo, y concuerdan, del mismo modo, en que la investigación experimental es el medio más confiable para llevar a cabo esta investigación. En particular, cuando los métodos de sillón y los métodos experimentales arrojan distintos resultados, es en los últimos en los que se ha de confiar.

⁷⁵ Ofrezco razones para dudar que Goldman esté en lo cierto en Kornblith (2007).

Lo que resulta difícil de defender, creo, es cualquier intento de aislar el proyecto filosófico del análisis conceptual de los resultados de las investigaciones empíricas. Después de todo, no es un accidente ortográfico que ambos, filósofos y psicólogos, se refieran al objeto de su investigación como si fuese un concepto. Sin duda, este no es un caso de simple homonimia, como con ‘banco’ y ‘banco’. Tanto el objeto de la investigación filosófica como el objeto de la investigación psicológica se piensa están implicados en la explicación del contenido de nuestros pensamientos, los tipos de inferencia que estamos inclinados a hacer, y en nuestra conducta categorizadora. Dado este amplio traslape [*overlap*] entre los esfuerzos causales y explicativos en que el discurso en torno a los conceptos se supone debe estar abocado, sin duda parece al mundo como si los objetos de estudio en estas dos investigaciones fuesen idénticos. Por lo menos, si incluso en parte se pueden exponer argumentos de que hay diferencias reales en este punto, aún será difícil ver cómo podrían los resultados de la investigación experimental resultar irrelevantes para una comprensión del objeto del análisis filosófico. Una vez más, dado el alcance y grado en el que los resultados experimentales resultan ser sorprendentes —esto es, dado que los resultados experimentales están tan profundamente reñidos con la visión que adoptemos desde el sillón— esto logra adelantar un serio desafío a los métodos de sillón en epistemología.

Hay otra manera en que los resultados empíricos son relevantes para el proyecto del análisis conceptual, según se le ha concebido tradicionalmente. La mayoría de los epistemólogos rechazan el escepticismo, y su rechazo del escepticismo es un compromiso bastante profundo.⁷⁶ Ellos ofrecen análisis de la justificación y del conocimiento, y ellos creen que estas aclaraciones sirven para explicar

⁷⁶ Mis comentarios sobre este punto se aplican igualmente a los contextualistas. Así, aunque los contextualistas quisieran conceder que el escepticismo es verdadero en contextos de alto nivel, también querrán insistir que el escepticismo es falso en contextos cotidianos. La clase de asuntos que aquí trato adelantarían inquietudes con respecto a los contextos cotidianos.

por qué es que estamos justificados y por qué es que conocemos en varios casos paradigmáticos. De este modo, una parte importante de la defensa de estos análisis es que los casos paradigmáticos de justificación y conocimiento resulten concordar con sus análisis. Una explicación de la justificación, o una explicación del conocimiento, que tiene por consecuencia que tenemos poca, si es que algo, creencia justificada o conocimiento, es, por esa misma razón, considerada implausible. No obstante, ahora los hechos psicológicos relativos a la adquisición y retención de la creencia son directamente relevantes para la evaluación de los análisis filosóficos, incluso en la concepción más tradicional del proyecto del análisis. La mayoría de las explicaciones del conocimiento, por ejemplo, incluyen alguna clase de requerimiento basal, y este requerimiento basal es típicamente explicado en términos causales o contrafácticos: una creencia no puede estar basada en otra a menos que la primera sea causada por la segunda, o a menos que no se tendría la primera si no fuese porque se tiene la segunda.⁷⁷ Nuestra comprensión de sillón de las dependencias causales y contrafácticas entre las creencias es, sin embargo, notoriamente deficiente. De este modo, a fin de ver si los casos paradigmáticos de conocimiento satisfacen en efecto las condiciones estipuladas en nuestros análisis, necesitamos salir del sillón y entrar en el laboratorio psicológico.⁷⁸ Lo mismo es verdad de las explicaciones de la justificación. En el caso de la justificación doxástica, un requerimiento basal es una característica tan esencial como lo es en las explicaciones del conocimiento. Pero, incluso las explicaciones de la justificación proposicional

⁷⁷ Véase, por ejemplo, la discusión de Feldman y Conee (2004) de la noción de “bien fundado” [*well-foundedness*], especialmente pp. 92-93.

⁷⁸ No quiero sugerir que los filósofos requieran ellos mismos de llevar a cabo experimentos. En la mayoría de los casos, los filósofos simplemente carecen del entrenamiento relevante para llevar a cabo experimentos con propiedad, y uno no hace mejor al involucrarse en la especulación filosófica mediante la experimentación amateur que por medio de la especulación de sillón. Lo que se requiere, sin embargo, y lo que aquí pretendo afirmar, es que la especulación filosófica debe estar informada por el trabajo experimental, ya sea hecha por filósofos o psicólogos (apropiadamente entrenados).

deben aceptar sus consecuencias para situaciones paradigmáticas. Una explicación de la justificación proposicional que tuviera la consecuencia de que la mayoría de las personas raramente satisface sus condiciones sería, por esa misma razón, en extremo implausible. Pero, si acaso las condiciones son de hecho satisfechas en situaciones del mundo real no es algo que pueda determinarse desde el sillón.

Este punto, creo, tiene implicaciones muy amplias. Considérense algunos de los términos centrales del discurso epistemológico: percepción, memoria, testimonio, razonamiento. Estos términos parecen identificar varios fenómenos del mundo real, y hay bastante trabajo psicológico que avanza en la comprensión de qué, justamente, se tratan estos fenómenos. Junto con este trabajo filosófico, vemos filósofos intentando ofrecer análisis de estas nociones epistemológicas fundamentales, y estos análisis son a menudo presentados en completo aislamiento de y, en efecto, con total ignorancia de la literatura psicológica. Los epistemólogos están interesados en estos temas, sin embargo, porque están interesados en la naturaleza y alcance del conocimiento humano. Un análisis filosófico de estas nociones que no logre concordar con los fenómenos del mundo real, por esa misma razón, fracasa en su objetivo. Tal como en el caso del conocimiento y la justificación, hay un problema serio con cualquier explicación de la percepción, la memoria, del testimonio o del razonamiento que tenga la consecuencia de que no haya mucho que en efecto esté ocurriendo. Sin embargo, la concordancia entre los análisis filosóficos y los fenómenos del mundo real no pueden ser determinados desde el sillón. Esto no es tan sólo una preocupación en principio. Algunos de estos fenómenos —particularmente la memoria, como he mencionado más arriba— resultan ser bastante diferentes de lo que parecen desde el sillón. Los análisis de sillón de tales fenómenos, de este modo, corren el serio riesgo de simplemente fracasar en captar los mismísimos fenómenos que intentan analizar. Aquí, el análogo del escepticismo es el eliminativismo, la afirmación de que realmente resulta que no hay percepción, memoria, testimonio o razonamiento; y el eliminativismo

es casi tan plausible como el escepticismo. Cuando los fenómenos de la cognición resultan ser significativamente diferentes de lo que pensamos que eran —esto es, cuando nuestra visión de sillón de estos fenómenos está profundamente equivocada— no podemos permitir que nuestra epistemología sea construida sobre nuestros errores de sillón. Sin embargo, tal como en algún otro lugar en ciencia, la psicología periódicamente revela que nuestra visión de sillón de las cosas no es meramente incompleta o en detalle errada; nuestra visión de sillón de los fenómenos psicológicos está, por momentos, extremadamente lejos del blanco. Sería un error muy grave dedicarse a la especulación epistemológica de una manera tal que simplemente ignore este hecho. Pero, esto significa que no podemos descansar contentos y estar satisfechos con una epistemología llevada a cabo desde el sillón.

Si se acepta este argumento, como creo que se debería, entonces cuenta en contra de considerar nuestros conceptos como el objeto del análisis epistemológico.⁷⁹ No es nuestro concepto del conocimiento, o de la justificación, o del razonamiento, o de la memoria el objeto del análisis filosófico; es el conocimiento mismo, la justificación misma, el razonamiento mismo, la memoria misma, etc. Tal como podemos representar mentalmente cosas tales como el oro y el agua de un modo tal que no logremos hacer justicia de las sustancias mismas, podríamos representar mentalmente los objetos de la investigación epistemológica de maneras tales que introduzcan malentendidos sustantivos. Estamos interesados en ciertos fenómenos reales —el conocimiento humano, por ejemplo— y no hay mucho sentido en elucidar, con el detalle que sea, cuáles son nuestras concepciones personales o *folk* de estos fenómenos. Por esta razón, no simpatizo más con el trabajo de los filósofos experimentales que con el de aquellos

⁷⁹ He presentado este argumento en detalle en Kornblith (2002). Williamson (2007) también defiende esta posición. Él y yo discrepamos con respecto al resultado de esta tesis para el uso de los métodos de sillón en filosofía.

que practican el análisis conceptual de una forma más tradicional. Es verdad, como han argumentado los filósofos experimentales, que si uno quiere dar con lo que son nuestros conceptos, entonces se debe proceder experimentalmente. Los métodos tradicionales de análisis conceptual equivalen a poco más que psicología social mal hecha. Pero, mientras los filósofos experimentales responden a este problema tratando de hacer buena psicología social, me parece que han tratado aquí con el problema equivocado. Hacer un mejor trabajo en dar con nuestros conceptos no constituye progreso alguno, pienso, ya que el objeto propio del análisis filosófico simplemente no es nuestros conceptos. Estos son los vehículos mediante los cuales representamos el objeto del análisis conceptual, pero esto es totalmente diferente de considerar a los conceptos mismos como el objeto de estudio.

De este modo, hay en epistemología muchísimos proyectos centrales que serían dramáticamente afectados por los resultados experimentales de las ciencias cognitivas. Nuestra visión de sillón de los fenómenos psicológicos está a menudo profundamente errada, y estos errores son luego incorporados en nuestros análisis filosóficos si no confiamos en nada más que en los métodos de sillón. El tema de la epistemología está tan estrechamente vinculado a estos fenómenos psicológicos, y nuestra visión de sillón de estos es tan gravemente sesgada, que nuestras teorías epistemológicas inevitablemente sufrirán si son aisladas de los resultados experimentales. Sin embargo, no es tan sólo un puñado de proyectos epistemológicos reconocidamente importantes los que necesitan estar informados por el trabajo de laboratorio. Los fenómenos mismos que los epistemólogos buscan iluminar son unos que son objeto de estudio experimental, y este trabajo experimental ha llevado a menudo a resultados sorprendentes. Ignoramos este trabajo bajo nuestro propio riesgo. Cuando las teorías epistemológicas en epistemología son construidas desde el sillón, corren el serio peligro de divorciarse de los fenómenos mismos que buscan iluminar. La única manera de asegurarnos de que no estamos construyendo elaborados castillos en el aire, desconectados de los fenómenos del mundo real que

motivaron nuestro trabajo desde el comienzo, es basar nuestro trabajo en la mejor comprensión experimental disponible de esos fenómenos. Cualquier epistemología informada así empíricamente dejará el sillón inevitablemente muy atrás.

III

Aunque Timothy Williamson y yo concordamos mucho con respecto al método filosófico,⁸⁰ hay también bastante con respecto a lo cual discrepamos. Williamson es lejos más optimista respecto de las posibilidades y proyecciones para la filosofía en general, y para la epistemología en particular, en tanto llevadas a cabo desde el sillón.⁸¹ Ahora bien, él comenta que “la legitimidad en principio de la filosofía experimental no vuelve ilegítima en principio a la filosofía de sillón”.⁸² Esto es bastante acertado, desde luego, y nadie ha argumentado que no lo sea. Mi argumento aquí, como en otros lugares,⁸³ no es que los métodos de sillón podrían, en principio, llevar a errores y omisiones, sino más bien que, en la práctica real, los métodos de sillón han llevado regularmente a errores bastante sustanciales en epistemología. Los rasgos de la cognición que son más directamente relevantes para la epistemología son unos que mal interpretamos gravemente cuando permanecemos en el sillón. Es por esta razón que me causan inquietud los modelos formales diseñados para cuadrar con nuestra concepción de sillón sobre estas materias. No es que piense que haya algo malo con el desarrollo de modelos formales —nada más lejos de eso. Como Williamson correctamente señala (2007, p. 128), la modelación formal y los métodos experimentales no tienen por qué ser vistos como rivales; en la práctica ordinaria de la ciencia, ellos son, por lo común, mutuamente beneficiosos. El *input* experimental proporciona datos

⁸⁰ Véase Kornblith (2009) y el ensayo de Williamson (2009).

⁸¹ De esta manera, Williamson comenta: “espero que los métodos de sillón jueguen legítimamente un rol más dominante en la filosofía futura de lo que [Kornblith] espera —desde luego, tal diferencia en el énfasis puede resultar en amplias diferencias en la práctica” (Williamson 2009, p. 126).

⁸² *Ibid.*, p. 126.

⁸³ Por ejemplo, Kornblith (2006 y 2002).

útiles para informar nuestros modelos formales; la modelación formal otorga información útil acerca de dónde es mejor testear nuestras teorías disponibles. Lo que objeto, entonces, no es la modelación formal *per se*, sino más bien que los modelos formales en epistemología no tienen el beneficio de esta interacción sustantiva con los datos experimentales. Cuando el *input* para nuestros modelos formales proviene del sillón más que del laboratorio, y cuando el *output* de nuestros modelos formales no se usa para informar pruebas experimentales, sino, en lugar de eso, meramente para provocar más interacciones con la visión obtenida desde el sillón, es probable que las teorías epistemológicas resultantes estén gravemente equivocadas. Incluso, cuando este método no origina errores, es un método sobre el cual no debiésemos depositar nuestra confianza. La modelación formal que no está informada ni testada por resultados experimentales solo contribuye a la mala epistemología.

Nadie podría oponerse razonablemente al uso de los métodos de sillón como un componente de la metodología filosófica. Después de todo, sí necesitamos, a veces, dar un paso atrás con respecto a los datos que se han recopilado, y las teorías que hemos construido sobre esa base, y pensar sobre sus implicaciones. Desde el sillón, podemos construir diferentes modelos formales y examinar sus consecuencias. Pero, reconocer esto es decir nada más que la filosofía, como cualquier actividad intelectual, involucra considerar las cosas detenidamente. Lo que está en disputa entre aquellos que son optimistas con respecto a las posibilidades y proyecciones de la metodología de sillón y aquellos que son más pesimistas no es si acaso dejar de pensar es una buena idea. Es, en lugar de eso, la pregunta sobre qué es lo necesitamos traer al sillón para ser pensado. Aquellos que favorecen una metodología completamente de sillón creen que ningún *input* experimental especial es realmente necesario para, al menos, gran parte de la especulación filosófica. Aquellos de nosotros que somos más pesimistas con respecto a esta clase de metodología creemos que los datos experimentales se requieren para mantenernos en contacto con los fenómenos que tratamos de comprender.

IV

He argumentado que sería un error continuar con la especulación epistemológica completamente desde el sillón. Permítaseme terminar, sin embargo, con una nota conciliatoria. La filosofía es una empresa colaborativa. Tal como hay una división del trabajo en física, con algunos físicos involucrados activamente en trabajo de laboratorio, y otros, más teóricamente inclinados, involucrados en la especulación de sillón, hay obviamente espacio para tal división del trabajo en filosofía, incluso según mi planteamiento. Sin embargo, esto no quiere decir que sea probable que cualquiera especulación de sillón añeja resulte constructiva. Los físicos que se involucran en la especulación de sillón no lo hacen en total y supina ignorancia del trabajo de sus colegas experimentales. La física de sillón llevada a cabo con ignorancia de los resultados experimentales carecería por completo de algún valor. Cuando hablamos del fenómeno real de la especulación de sillón en física, entonces, lo que tenemos en mente no es trabajo que es completamente independiente de los resultados experimentales. En vez de eso, hablar de especulación de sillón en física equivale al simple reconocimiento de los diversos intereses y talentos de aquellos que contribuyen a la teorización física, junto con el reconocimiento de que el progreso en física no es típicamente atribuible a individuos actuando perfectamente aislados, sino que es, en lugar de eso, el producto de una comunidad de investigadores. Lo mismo es verdad, desde mi perspectiva, de la epistemología. Los teóricos individuales pueden llevar a cabo su trabajo desde el sillón de modo constructivo, siempre y cuando sean parte de una comunidad que los mantenga en contacto con los fenómenos epistemológicos reales mediante trabajo experimental. Mal interpretamos el rol de la especulación de sillón, según mi parecer, si pensamos que es posible progresar en epistemología sin que el trabajo experimental juegue este importante rol.⁸⁴

⁸⁴ Estoy en deuda con Timothy Williamson por las conversaciones sobre este tema en múltiples ocasiones.

Referencias

- Bishop, M. (2000): “In praise of epistemic irresponsibility: How lazy and ignorant can you be?” *Synthese* 122(1), 179-208.
- Bishop, M. y Trout, J. D. (2005): *Epistemology and the Psychology of Human Judgment: Oxford: Oxford University Press.*
- Bonjour, L. (1985): *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.
- _____ (2001): “The indispensability of internalism”. *Philosophical Topics* 29 (1/2), 47-65.
- Burge, T. (1993): “Content preservation”. *Philosophical Review* 102, 4, 457-488.
- Chisholm, R. (1977): *Theory of Knowledge*. Englewood Cliffs, N.J.: Prentice-Hall.
- Christensen D. y Kornblith, H. (1997): “Testimony, memory and the limits of the a priori”. *Philosophical Studies* 86, 1, 1-20.
- Feldman, R. (1999): “Methodological naturalism in epistemology”. En: J. Greco y E. Sosa (eds.) *The Blackwell Guide to Epistemology*. Oxford: Blackwell.
- Feldman, R. y Conee, E. (2004), “Evidentialism”. En: *Evidentialism: Essays in Epistemology*. Oxford: OUP, pp. 83-107.
- French P. A. y Wettstein, H.K. (2007): *Midwest Studies in Philosophy: Philosophy and the Empirical*, 31, 1.
- Fricker, E. (1994): “Against gullibility”. En: B. K. Matilal y A. Chakrabarti (eds.) *Knowing from Words*. Dordrecht: Kluwer Academic Publishers, pp. 125-161.
- Gettier, E. (1963): “Is knowledge justified true belief?” *Analysis* 23, 6, 121-3.
- Goldman, A. (1986): *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

- _____ (1992): *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.
- _____ (2007): "Philosophical intuitions: Their target, their Source, and their epistemic status". *Grazer Philosophische Studien* Vol. 74, 1, 1-26.
- Greco, J. (1978): "Epistemics: The regulative theory of cognition". *The Journal of Philosophy* 75, 10, 509-523.
- _____ (2004): *Ernest Sosa and his Critics*. Oxford: Blackwell.
- Horvath, J. (manuscrito) "Conceptual analysis and naturalized epistemology".
- Jackson, F. (1998): *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: OUP.
- Kitcher, P. (1980): "A priori knowledge". *Philosophical Review* 89, 1, 3-23.
- Knobe, J. y Nichols, S. (2008): *Experimental Philosophy*. Oxford: OUP.
- Kornblith, H. (manuscrito) "The myth of epistemic agency".
- _____ (1989): "Introspection and misdirection". *Australasian Journal of Philosophy* 67, 4, 410-422.
- _____ (1993): *Inductive Inference and its Natural Ground*. Cambridge, MA: MIT Press.
- _____ (1994): *Naturalizing Epistemology*. Cambridge, MA: MIT Press.
- _____ (1998): "The role of intuition in philosophical inquiry". En: M. DePaul y W. Ramsey (eds.) *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham, MR: Rowman and Littlefield Publishers, pp. 129-141.
- _____ (2001): *Epistemology: Internalism and Externalism*. Oxford: Blackwell.
- _____ (2002): *Knowledge and its Place in Nature*. Oxford: OUP.

- _____ (2006): “Appeals to intuition and the ambitions of epistemology”. En: S. Hetherington (ed.) *Epistemology Futures*. Oxford: OUP, pp. 10-25.
- _____ (2007): “Naturalism and Intuitions”. *Grazer Philosophische Studien* 74, 1, 27-49.
- _____ (2009): “Timothy Williamson’s *The Philosophy of Philosophy*”. *Analysis*, 69, 1, 109-16.
- Korsgaard, C. (1996): *Source of Normativity*. Cambridge: Cambridge University Press.
- Lackey, J. y Sosa, E. (2006): *The Epistemology of Testimony*. Oxford: OUP.
- Margolis, E. y Laurence, S. (1999): *Concepts: Core Readings*. Cambridge, MA: MIT Press.
- Moran, R. (2001): *Authority and Estrangement: An Essay on Self-Knowledge*. NJ: Princeton University Press.
- Quine, W. V. O. (1951): *The Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press.
- _____ (1961): *From a Logical Point of View*. Cambridge, MA: Harvard University Press.
- _____ (1969): *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Rosch, E. y B. Lloyd (1978): *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum.
- Taylor, S. (1989): *Positive Illusions: Creative Self-Deception and the Healthy Mind*. London: Basic Books.
- Weinberg, J. Nichols, S. y Stich, S. (2001): “Normativity and Epistemic Intuitions”. *Philosophical Topics* 29, 429-460.
- Williamson, T. (2007): *The Philosophy of Philosophy*. Oxford: Blackwell.
- _____ (2009): “Replies to Kornblith, Jackson and Moore”. *Analysis* 69, 125-135.
- Wilson, T. (2002): *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Harvard University Press.

Bombas de Intuiciones⁸⁵

Daniel C. Dennett

Acuñé el término “bomba de intuición” en 1980, aunque el mérito también debería atribuirse a Douglas Horfstadter, pues tal bautizo tuvo lugar luego de discusiones con él del experimento mental de John Searle “La Pieza China” (Searle 1980), el que reimprimimos con comentarios en nuestra antología, *The Mind’s I* (Dennett y Hofstadter 1981). Encontramos el experimento mental de Searle fascinante porque, por una parte, era un argumento tan atractivo y engañoso, e incluso así, tan claramente, un adulator de masas y seductor. ¿Cómo y por qué este experimento funcionaba? Junto a Doug lo examinamos desde el punto de vista de la *ingeniería en reversa* [*reverse engineering*], y Doug hizo notar que cuando uno se enfrenta con un artefacto como este, uno debe “girar todas las perillas” [*turn all the knobs*] para saber qué hace para que funcione: Ud. debe mentalmente variar todos los detalles del experimento para saber si las mismas intuiciones son bombeadas, o bombeadas con más fuerza, cuando los detalles son alterados. ¿Es este argumento (si es este un argumento, y no solo una historia cautivadora) sólido bajo alteraciones, o depende de manera crítica de detalles que deberían ser optativos? Doug se centró en la frase “hojas de papel” [*slips of paper*] (¿o era “pedacitos de papel” [*bits of paper*] en el ensayo de Searle, y mostró cómo incentivaba a la gente a subestimar el tamaño y complejidad del software involucrado

⁸⁵ Inédito. Traducido y publicado en esta compilación de artículos con permiso de su autor.

en muchos niveles de magnitud. Su comentario sobre Searle en nuestro libro presentó esta crítica, y condujo a una respuesta feroz de él en las páginas de *New York Review of Books*, debido a que, aunque habíamos reimpresso su artículo correctamente, en su comentario Doug se equivocó y escribió “pedacitos” [*bits*] donde Searle había dicho hojas [*slips*] (¿o fue a la inversa?). De esta manera, Searle argumentó, ¡Doug describía su argumento de forma totalmente errónea! Esto justamente demostró que nuestro punto era correcto, en una forma: si tal pequeño desajuste habilita o inhabilita a un experimento mental, ¡esto es algo que debería captar la atención de todo aquél al que se le están bombeando las intuiciones!

Teníamos mucho más que decir acerca de las bombas de intuiciones en *The Mind's I*, el cual es, en verdad, una colección de bombas de intuiciones. Las reflexiones de Doug acerca de Searle incluyeron un virtuoso análisis de la Pieza China, una bomba de intuición con al menos cinco perillas, las cuales procedimos a girar, de a una, para derivar una familia de experimentos mentales diferentes con conclusiones distintas respectivas. En las reflexiones finales, hago hincapié en lo siguiente:

Cada ajuste de los detalles en una bomba de intuición produce una narrativa ligeramente diferente, con problemas distintos que se difuminan en el contexto y con diferentes moralejas que extraer. En qué versión o versiones uno debería confiar, es un problema a resolver mediante el examen cuidadoso de aquellos, para ver qué características de la narrativa están haciendo el trabajo. Si las sobresimplificaciones son la *f fuente* de las intuiciones en vez de tan solo artefactos para suprimir las complicaciones irrelevantes, deberíamos no confiar en las conclusiones que se nos invita extraer. Estas son un asunto de juicio prudente, por lo que no

es de extrañar que una sospecha generalizada y tan justificada rodeen tales ejercicios de la imaginación y de la especulación”. (p. 460, cursivas en cita original).

Felizmente, el término “bomba de intuición” ha resultado ser de provecho para otros y ha emigrado de su lugar de nacimiento a ambientes más interesantes y fructíferos, y su significado está, hasta donde puedo ver, todavía cambiando. (Como acuñador, no tengo derecho sobre la fijación [*fixation*] por supuesto, aunque podría intentar influir a otros usuarios). ¿Es este [*intuition pump*] sinónimo de “experimento mental” (en su uso común) o se refiere a una subvariedad de experimento mental (lo cual fue mi intención original): aquellos que no dramatizan un argumento lógico simplemente contando una historia? De esta manera es como elucidado el término en *Elbow Room*:

Tales experimentos mentales (a diferencia de los de Galileo o de Einstein por ejemplo) *no* se supone que abrigan argumentos estrictos para probar conclusiones desde premisas. Más bien, su punto es producir una familia de reflexiones imaginativas en el lector que en último término no lleve a una conclusión formal sino que a una “intuición”. Las bombas de intuición están astutamente diseñadas para que la atención del lector se centre en las características “importantes”, y para desviar al lector, y así no se atasque en detalles difíciles de seguir. En principio, no hay nada malo con esto. Ciertamente, uno de las tareas más importantes de la filosofía es encontrar maneras para ayudar a la gente a ver el bosque y no los árboles. Pero se abusa de las bombas de intuición, aunque de manera rara deliberadamente.

Tal vez el abuso más frecuente es inferir un resultado —un juicio intuitivo emotivo— de la tremenda simplicidad del caso imaginado, en vez de desde el contenido actual de los ejemplos descritos tan simple y claramente. (Dennett 1984, p. 12) [N. del T.: todos los destacados corresponden a la cita original]

En otro pasaje, le doy a las bombas de intuición un rol crucial en la filosofía:

Una institución medianamente poderosa de la represión profesional nos ha ocultado la centralidad e influencia de las bombas de intuición en el desarrollo de la filosofía. No es en el área de la voluntad libre que las bombas de intuición han sido la fuerza dominante. Sugiero que la reflexión sobre la historia de la filosofía muestra que las grandes bombas de intuición han sido siempre los mayores alicientes [*movers*]. Piense en la caverna de Platón, Menon enseñándole geometría al niño esclavo, el genio maligno de Descartes, y el estado de la naturaleza de Hobbes. Piense, más recientemente, en el lingüista de Quine (1960) tratando de traducir “Gavagai”, el puzle de Goodman (1983), la posición original de Rawls (1971), y la seductora pregunta de Farrell (1950) acerca de cómo sería ser un murciélago, sin dejar de mencionar el notable ejemplo de Putnam (1975) de la tierra gemela e incluso la aún más notoria pieza china de Searle (1980).

Uno podría decir que estas bombas de intuiciones son melodías que se mantienen a través del

tiempo [*enduring*], con el poder duradero que hace que permanezcan y que serán recordadas por nuestros novatos [*freshmen*], de manera vívida y precisa, años después de que ellos han olvidado los intrincados contrapuntos que rodean al argumento y su análisis. Una buena bomba de intuición es más poderosa que cualquier versión de este. (¿Cuántas variaciones han sido reiteradamente dadas acerca de los temas de Rawls de la posición original o de la tierra gemela de Putnam?).

Las bombas de intuiciones son artefactos pedagógicos poderosos. En general, hay consenso de que el experimento mental de Descartes “*cogito ergo sum*” es lógicamente sospechoso, si no totalmente defectuoso. Literalmente, este ha inspirado a docenas de interpretaciones y defensas; muchos profesores de filosofía desecharían todos estos comentarios aunque nunca soñarían remover la dramática idea de Descartes del programa del curso. Incluso las más grandes bombas de intuición pueden perderlo tanto a uno como pueden enseñarle. Cuando enseñamos Descartes, por ejemplo, típicamente no enseñamos su experimento mental como revelando la verdad —o incluso conduciendo a la verdad— acerca del conocimiento. De hecho, es típico que echemos la culpa a Descartes y a su seductora bomba de intuición por llevar a los filósofos a una búsqueda estéril [*a wild goose chase*]. Por lo menos estamos agradecidos de que él, de la misma manera en que nosotros estaríamos agradecidos de alguien que nos dio

una dirección errónea, pero cuyas indicaciones nos llevaron a una fascinante aventura de la cual aprendimos un montón.

El rol central de las bombas de intuiciones en la filosofía es mostrarnos que la filosofía no es y que razonablemente no aspiraría a ser, ciencia. La filosofía sin las bombas de intuiciones ocasionalmente tiene éxito al purificar y delimitar un área conceptual que es suficiente para que la ciencia se enseñoree, pero en general, estas no son triunfos de las corrientes principales de la filosofía. La filosofía con las bombas de intuiciones no es ciencia en absoluto pero en su forma informal es una valiosa manera —incluso necesaria ocasionalmente— compañera de la ciencia. No debería ser embarazoso para los filósofos darse cuenta que las bombas de intuiciones hacen bastante para el trabajo duradero de la filosofía (para peor o para mejor). Después de todo, una bomba de intuición debería ser la herramienta ideal en el *kit* del filósofo, si tomamos en serio una de las perspectivas más conocidas de *para qué* es la filosofía. Es para ampliar nuestra visión de lo posible, para romper los malos hábitos de pensamiento. Como dijo Wittgenstein, “la filosofía es una batalla en contra del embrujo de nuestra inteligencia por medio del lenguaje” (1953, sección 109). Para tales tareas, la organización estricta del argumento riguroso es pocas veces más que una póliza de seguro, un control sobre la intuición desbordada promovida que ha centrado su atención en una nueva teoría (Dennett 1984, p. 12 y ss.).

Uno de los motivos centrales en *Elbow Room* es demostrar, una y otra vez, que la bibliografía sobre la voluntad libre tiene una gran cantidad de bombas de intuiciones defectuosas o mal empleadas, experimentos mentales cuyo carácter no conclusivo puede mostrarse mediante la implementación de experimentos *de control*, ajustando las perillas y observando de nuevo. En *Freedom Evolves* (2003), resalté que nuestro pensamiento filosófico acerca de la voluntad libre tiene una nueva herramienta. Ahora tenemos una variedad de bombas de intuición mejoradas mediante prótesis: simulaciones computacionales que pueden resguardar nuestra transparencia, mostrando bajo una variedad de condiciones que ciertos resultados son sólidos.

La literatura sobre la conciencia está densamente poblada de bombas de intuiciones de dudoso valor —pero a menudo, como La Pieza China de Searle— de gran popularidad. El caso de Frank Jackson (1986), de *Mary* la científica del color es con seguridad el más famoso y más egregio, defectuoso, he argumentado, porque “¡En verdad lo incentiva a uno a malinterpretar sus premisas!” (Dennett 1991, p. 398). Este veredicto es defendido también en “What RoboMary Knows” (Dennett 2005, p. 122). Pero recuerde: las bombas de intuiciones son herramientas poderosas, y pueden aplicarse de buena manera, si se realizan con cuidado. En “Quining Qualia” (Dennett 1988), intenté minar la conciencia de los filósofos de que sabían de qué estaban hablando cuando lo hacían acerca de los *qualia* mediante “una serie de quince bombas de intuiciones, formuladas en una secuencia con el objeto de exorcizar y de deshacerse de las enojosas intuiciones.

Algunos experimentos mentales argumentos formales mostrados de manera vívida —ciertamente inolvidables en los mejores casos— a través de ejemplos escogidos para aglutinar [*populate*] las premisas. Otros tienen un argumento simple y formal como columna vertebral, pero dependen críticamente del ejercicio informal que “establece” una premisa crucial. Todo el trabajo es efectuado por la manera en que imaginamos los detalles, y aquí es mucho menos obvio lo que es relevante —hasta que hemos realizado un montón de variaciones.

Pero la mayor parte de los experimentos mentales en la filosofía son incluso menos formales, y dependen para su eficacia en algún grado de conformidad de parte de la audiencia en vez de la suspensión propia de la incredulidad que es la respuesta apropiada a una buena cuenta cuentos. Pienso que el valor clave de examinar tales experimentos mentales como bombas de intuición es que llama la atención sobre el artificio involucrado, su diseño, su supuesta funcionalidad, y nos incentiva a ser cautos e inquisitivos sobre cómo nuestras mentes podrían manipularse mediante un dispositivo tal. Las bombas de intuición son parte de la caja de herramientas del filósofo, y si usted quiere ser un trabajador confiable, usted debe entender cómo funcionan sus herramientas.

Referencias

- Dennett, D. (1984): *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: Bradford Books.
- _____ (1988): “Quining qualia”. M. y E. Bisiach (eds.) *Consciousness in Contemporary Science*. Oxford: OUP, pp. 43-77.
- _____ (1991): *Consciousness Explained*. New York: Little Brown.
- _____ (2003): *Freedom Evolves*. New York: Penguin Books.
- _____ (2005): *Sweet Dreams*. Cambridge, MA: MIT Press.
- Dennett, D. y Hofstadter, D. (1981): *The Mind's I: Fantasies and Reflections on Mind and Soul*. New York: Basic Books.
- Farrel, B. (1950): “Experience”. *Mind* 59, 170–198.
- Flanagan, O. (1997): “The robust phenomenology of the stream of consciousness”. En: N. Block, *et. al. (eds.) The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, pp. 89-93.
- Goodman, N. y Putnam, H. (1983): *Fact, Fiction and Forecast*. Cambridge, MA: Harvard University Press.

- Jackson, F. (1986): “What Mary Didn’t Know”. *The Journal of Philosophy* 83 (5), 291-5.
- Putnam, H. (1975): *The meaning of ‘meaning’*. En: *Mind, Language and Reality. Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press.
- Quine, W. V. O. (1960): “Translation and meaning”. En: *Word and Object. Cambridge, MA: MIT Press, pp. 26-73*.
- Rawls, J. (1971): *A Theory of Justice. Cambridge, MA: Harvard University Press*.
- Searle, J. (1980): “Minds, Brains and Programs”. *Behavioral and Brain Sciences* 3, 417-24.
- Wittgenstein, L. (1953): *Philosophical Investigations*. Oxford: Blackwell.

“Bomba de intuición” La evolución de una herramienta epistemológica⁸⁶

Rodrigo A. González Fernández

1. Introducción

La epistemología de los experimentos mentales es cautivante no solo para los desarrollos teóricos de científicos y filósofos. Lo es, por ejemplo, para los físicos y biólogos, pero también para los investigadores de la filosofía del lenguaje, la filosofía de la mente, la ética, y para varias disciplinas más de la filosofía contemporánea. En general, todos los defensores de los experimentos mentales de alguna manera asumen que las operaciones que estos describen son pertinentes, relevantes y efectivas para generar un argumento, y así para llegar a una conclusión. Ello ocurre incluso si los experimentos mentales son positivos, o propedéuticos, o bien negativos, o con el fin de refutar una teoría luego de asumida la verdad de esta en aras de la argumentación.

Pero, el vulgo opina que las excepciones confirman la regla, y esto ciertamente parece acontecer con algunos experimentos mentales, especialmente en el ámbito de discusión de qué es lo mental. En efecto, a propósito de la Habitación China, publicada en 1980,

⁸⁶ Una versión preliminar de este trabajo fue presentada en las Jornadas Chuaqui 2016, Santiago de Chile. Agradezco los comentarios de Andrés Bobenrieth y de varios asistentes a la ponencia. También agradezco la generosidad de Jan Heylen, Felipe Morales y Guido Vallejos, quienes han enriquecido la discusión de este trabajo.

Dennett y Hofstadter introdujeron un término del todo novedoso en la mencionada epistemología: las denominadas bombas de intuiciones. Estas son una clase de experimento mental, que tendrían como *leit motiv* la seducción de las masas. Ello sucedería por el bombeo de intuiciones mediante una narrativa que parece argumentativa, pero que dudosamente lo es. Según el enfoque original de Dennett/Hofstadter, las bombas de intuiciones son, entonces, artilugios para lograr la persuasión de una audiencia. Sus premisas son dudosas y discutibles, y por ello recurren a ciertas imágenes plasmadas mediante detalles narrativos, que resultan tan pintorescos y originales como seductores. Las operaciones descritas por algunas bombas justamente muestran ese carácter pintoresco y seductor, como ocurre en la Habitación China, la Pieza de *Mary*, el murciélago de Nagel y en varios más. En consecuencia, Dennett y Hofstadter introdujeron un término para intentar desacreditar algunos experimentos mentales, denominadas por ellos bombas de intuiciones.

En vista de estas críticas, en el presente ensayo hago una revisión del término “bomba de intuición”, mostrando la evolución que ha tenido el mismo. En efecto, parece existir un importante cambio de posición, al menos en el pensamiento de Dennett, el cual originalmente trata a algunos experimentos mentales como carentes de virtud epistémica.

De hecho, que el término “bomba de intuición” haya sido acuñado por él y Hofstadter una vez aparecido el argumento de la Habitación China de Searle, me llevó a pensar que el problema con estas herramientas epistemológicas no es inherente a ellas, sino más bien derivado del problema mente-cuerpo (González 2013). Pero, esta tesis requiere de algunas precisiones. Dennett ha hecho modificaciones a su tratamiento de las bombas de intuiciones, y ha encontrado para ellas una serie de ventajas para el razonamiento, la imaginación, la argumentación, etc. En consecuencia, en este ensayo muestro la evolución de su tratamiento de las bombas de intuiciones, y énfasis de qué forma han pasado de ser elementos negativos a positivos en la epistemología de los experimentos mentales.

2. La Habitación China y la definición original del concepto “bomba de intuición”

Como dice la canción de un famoso cantante romántico, todo empezó a ir mal en cierto momento. Dicho momento clave ocurrió cuando Searle publicó su famoso y siempre controvertido experimento mental de la Habitación China. Lo fue, estimo, porque Dennett con la inteligencia que lo caracteriza, dudó que fuera un argumento, y más aún, calificó a la narrativa de Searle como un adulator de masas, un artefacto epistémico puramente seductor y persuasivo. En vista de estos calificativos, acuñó junto a Hofstadter el término “bomba de intuición”, un concepto que ciertamente engrosó los términos de la epistemología de los experimentos mentales.

Para Dennett, el argumento serleano no cuenta como un experimento mental más. Tal como enfatiza:

[...] Era un argumento tan atractivo y engañoso, e incluso así, tan claramente, *un adulator de masas y un seductor*. ¿Cómo y por qué este experimento funcionaba? Junto a Doug lo examinamos desde el punto de vista de la *ingeniería en reversa* [*reverse engineering*], y Doug hizo notar que cuando uno se enfrenta con un artefacto como este, uno debe “girar todas las perillas” [*turn all the knobs*] para saber qué hace que funcione: Ud. debe mentalmente variar todos los detalles del experimento para saber si las mismas intuiciones son bombeadas, o bombeadas con más fuerza, cuando los detalles son alterados. ¿Es este argumento (si es este un argumento, y *no solo una historia cautivadora*) sólido bajo alteraciones, o depende de manera crítica de detalles que deberían ser optativos? (Dennett, en esta compilación, p. 117)

Hay, sin embargo, un tenor algo ambiguo en la crítica. Tanto Dennett como Hofstadter consideran que lo característico de la Habitación China es ser un adulator de masas, una narrativa seductora que, pese a ser fascinante, está lleno de detalles de los cuales penden una serie de intuiciones, las que al ser bombeadas, llevan al lector a conclusiones tendenciosas. De ahí entonces, uno podría generalizar así: una característica crucial de las bombas de intuiciones, es ser narrativas que, al depender de detalles persuasivos, son capaces de bombear las intuiciones de una audiencia. Pero, ¿qué tipo de intuiciones? Y, más importante aún, ¿son dichas intuiciones constitutivas de conocimiento genuino? O, en términos más simples aún, ¿son confiables?

Con relación a la primera pregunta, es posible sostener que no son intuiciones como aquellas que usualmente se llama del “sexto sentido femenino”. Eso de que María intuye que Juan le es infiel con Mónica, aunque no haya evidencia alguna. Al contrario, estimo que la cuestión de las intuiciones tiene que ver con que estas tienen como objetivo central disimular la debilidad de algunas premisas para establecer una conclusión. Luego, *suponen* la verdad y la solidez de las mismas, las cuales tienen como objetivo permitir la inferencia de la conclusión, también intuita. Así, uno puede responder, en relación con las preguntas segunda y tercera, que las bombas de intuiciones no establecen conocimiento genuino, ni serían del todo confiables.

Aquí el análisis de un caso será esclarecedor con respecto a lo que quiero decir, y especialmente a lo que quiere decir Dennett con “algunas bombas de intuiciones”. Tómese nuevamente el vapuleado experimento mental de la Habitación China de Searle (1980), el cual sufrió el embate de críticas por doquier. Entre ellas, justamente la de haber planteado de manera laxa, poco rigurosa, una serie de premisas para establecer que la sintaxis de un programa es incapaz, en principio, de proveer semántica.

Teniendo presente esta crítica particular, Searle (1990) da un cariz más analítico a su argumento, al ordenarlo con axiomas y conclusiones.

Para discutirlo brevemente, lo cito a continuación:

Axioma 1: Los programas computacionales son formales (sintácticos)

Axioma 2: Las mentes humanas tienen contenidos (semántica)

Axioma 3: La sintaxis en sí no es constitutiva ni suficiente para la semántica

Conclusión 1: Los programas no son constitutivos ni suficientes para las mentes

Axioma 4: Los cerebros causan mentes

Conclusión 2: Cualquier sistema capaz de causar mentes debe tener los mismos poderes causales equivalentes a los de los cerebros

Conclusión 3: Los programas computacionales no pueden duplicar los poderes causales de los cerebros que producen mentes (p. 21)

A propósito de la axiomatización de la Habitación China, una crítica de los Churchlands, que es muy similar a la de Dennett y Hofstadter, es que Searle no puede saber la verdad del axioma 3, esto es, que la sintaxis en sí no es constitutiva ni suficiente para la semántica (P.M. y P.S. Churchland 1990, p. 34). Y de ahí Searle desprendería la aventurada conclusión, del todo *intuida*, de que las máquinas programadas no pueden tener, en principio, estados mentales con contenido, por ser puramente sintácticas.

De hecho, Dennett critica que la Habitación China sea un ejemplo válido, paradigmático, de que la sintaxis es insuficiente para la semántica. Su crítica de que se mueven perillas en el experimento mental, y de que es una narrativa seductora y tendenciosa, tiene ese

preciso fundamento. En otras palabras, Searle recurriría a una serie de imágenes mentales con el objetivo de mostrar que las máquinas programadas son incapaces en principio de tener estados mentales con contenido. El hecho de que Searle manipule ideogramas en chino en función de un programa para correlacionar símbolos sobre la base de su pura forma tiene precisamente ese sentido, a saber, proporcionar una imagen mental, una narrativa, en que el operador del experimento mental supone que la premisa 3 lleva a la primera conclusión de la Habitación China, esto es, que se refuta la tesis central del cognitivismo. En consecuencia, esa es la principal función de la narrativa de la Habitación China, según Dennett y Hofstadter: el disimulo de la debilidad de la premisa 3, la cual ni Searle, ni ningún operador del experimento mental, podría saber con certeza. En relación con este punto, se bombea una intuición clave: que el cognitivismo es falso porque manipular símbolos es insuficiente para que emerjan estados mentales genuinos.

Los detalles a los que alude Hofstadter serían cruciales para bombear la mencionada intuición. Por ejemplo, si alguien cambiara “manipular símbolos” por “traducir cadenas de símbolos”, se arruinaría el experimento mental, en el sentido de que las intuiciones correctas no serían bombeadas. Esto es ciertamente correcto, porque “manipular” no implica saber qué se está manipulando, y por ello uno logra intuir que alguien que manipula símbolos, o ideogramas en chino, no entiende, y por tanto, no necesariamente habla chino. De esta forma, los detalles sí son relevantes para que emerjan a las intuiciones correctas en una bomba de intuición.

Para Dennett, no obstante, cada detalle del experimento mental cuenta como una perilla que girar, que ciertamente podría llevar a intuiciones diferentes. Enfatiza esta idea del siguiente modo:

Cada ajuste de los detalles en una bomba de intuición produce una narrativa ligeramente diferente, con problemas distintos que se

difuminan en el contexto y con diferentes moralejas que extraer. En qué versión o versiones uno debería confiar, es un problema que resolver mediante el examen cuidadoso de dichos detalles, para ver qué características de la narrativa están haciendo el trabajo. Si las sobresimplificaciones son la *fuerza* de las intuiciones en vez de tan solo artefactos para suprimir las complicaciones irrelevantes, deberíamos no confiar en las conclusiones que se nos invita extraer. Estas son un asunto de juicio prudente, por lo que no es de extrañar que una sospecha generalizada y tan justificada rodee tales ejercicios de la imaginación y de la especulación. (Dennett y Hofstadter 1981, pp. 459-460, énfasis en original).

Nótese que aquí se alude a las palabras imaginación y especulación, dos actividades que pueden carecer de suficiente rigor. Esto no es necesariamente negativo, si el objetivo de un experimento mental tipo bomba de intuición lleva a una conclusión simplificando detalles irrelevantes. Pero, cuando los detalles son relevantes, entonces las intuiciones bombeadas pueden desviar al sobre simplificar. Por otra parte, cuando descartan detalles irrelevantes, las bombas son útiles. Recalca Dennett esta idea así:

En tales experimentos mentales (a diferencia de los de Galileo o de Einstein por ejemplo) *no* se supone que abrigan argumentos estrictos para probar conclusiones desde premisas. Más bien, su punto es producir una familia de reflexiones imaginativas en el lector que en último término no lleve a una conclusión formal sino que a una “intuición”. Las bombas de intuición están astutamente diseñadas para

que la atención del lector se centre en las características “importantes”, y para desviar al lector y que no se atasque en detalles difíciles de seguir. En principio, no hay nada malo con esto. Ciertamente, una de las tareas más importantes de la filosofía es encontrar maneras para ayudar a la gente a ver el bosque y no los árboles (Dennett 1984, p. 12, énfasis en original).

Nótese el cambio de postura de Dennett aquí: en vez de ser aduladores de masas, las bombas de intuiciones no serían del todo negativas y no lo son, pese a que dependen de narrativas imaginativas y a que apelan a intuiciones. No son del todo negativas, reza su nueva posición, porque algunas de esas reflexiones imaginativas ayudan al lector a comprender un problema filosófico de manera general, sin perderse en los detalles, o en los árboles del bosque. Incluso, la simplificación de las bombas puede ayudar a reflexionar y a criticar las aristas de un problema. Por eso Dennett agrega lo siguiente:

Sugiero que la reflexión sobre la historia de la filosofía muestra que las grandes bombas de intuición han sido siempre los mayores agujones teóricos [*movers*]. Piense en la caverna de Platón, Menon enseñándole geometría al niño esclavo, el genio maligno de Descartes, y el estado de la naturaleza de Hobbes. Piense, más recientemente, en el lingüista de Quine (1960) tratando de traducir “Gavagai”, el puzle de Goodman (1983), la posición original de Rawls (1971), y la seductora pregunta de Farrell (1950) acerca de cómo sería ser un murciélago, sin dejar de mencionar el notable ejemplo de Putnam (1975) de la tierra gemela e incluso la aún más notoria habitación china de Searle (1980) (Dennett 1984, p. 18).

Es curioso que aquí la Habitación China ya no es descalificada como un adulator de masas, ni es considerada como un pseudo argumento. Por el contrario, se la considera, al igual que a otros experimentos mentales tipo bomba de intuición, como un agujijón, un verdadero incentivo para reflexionar acerca de ideas relacionadas con el cognitivismo. Por esta razón, estimo que las bombas de intuiciones, sean malas o buenas, son melodías que se preservan en el ideario colectivo de la filosofía, haciendo que los novatos sean introducidos a temas y problemas clásicos de alguna sub disciplina filosófica.

Es por esto también que, según la nueva postura de Dennett, las bombas de intuiciones pueden considerarse poderosos instrumentos pedagógicos. Pero, ve en estos problemas también. En efecto, los compara con una serie de indicaciones para llegar a una dirección, las cuales pueden acertar, pero también pueden desorientar. El *cogito ergo sum*, por ejemplo, lo compara con una serie de direcciones erradas, que han inspirado a un grupo considerable de filósofos a una búsqueda estéril [*wild goose chase*], pero que incluso así, han llevado a una aventura fascinante.

En relación con este interesante punto, considera que el carácter pedagógico de las bombas muestra por qué la filosofía, pese a Quine, no aspira a ser ciencia. Por el contrario, dichas bombas depuran problemas filosóficos y por ello sirven “para ampliar nuestra visión de lo posible, para romper los malos hábitos de pensamiento” (Dennett 1984, pp. 12 y ss.). Aquí Dennett claramente se inspira en Wittgenstein (1953).

Sin embargo, de todas formas parece que la posición original de Dennett tiende a persistir en relación con las bombas de intuiciones. Tal parece que se debe hacer un análisis caso a caso, perilla a perilla, en los experimentos mentales tipo bomba de intuición. No queda claro qué criterio debe cumplir una bomba de intuición para ser considerada un experimento mental de peso. Y ello es especialmente relevante en el campo de la filosofía de la mente, en que Dennett y Hofstadter descalifican una serie de experimentos mentales

tildándolos, precisamente, de ser meras bombas de intuiciones. Lo anterior puede apreciarse en el siguiente párrafo, en que el primero descalifica a algunos de estos experimentos mentales, y, no sorprende, defiende los propios:

La literatura sobre la conciencia está densamente poblada de bombas de intuiciones de dudoso valor —pero a menudo, como La Pieza China de Searle— de gran popularidad. El caso de Frank Jackson (1986), de *Mary* la científica del color es con seguridad el más famoso y más egregio, defectuoso, he argumentado, porque “¡En verdad lo incentiva a uno a malinterpretar sus premisas!” (Dennett 1991, p. 398). Este veredicto es defendido también en “What RoboMary Knows” (Dennett 2005, p. 122). Pero recuerde: las bombas de intuiciones son herramientas poderosas, y pueden aplicarse de buena manera, si se realizan con cuidado. En “Quining Qualia” (Dennett 1988), intenté minar la idea de los filósofos de que sabían de qué estaban hablando cuando lo hacían acerca de los *qualia* mediante “una serie de quince bombas de intuiciones, formuladas en una secuencia con el objeto de exorcizar y deshacerse de las enojosas intuiciones”. (Dennett, en esta compilación, p. 120)

Creo que es posible concluir que no hay claridad en la primera postura de Dennett respecto de cuándo un experimento mental tipo bomba de intuición es positivo, es decir, un aporte a la discusión racional y crítica. La cuestión de que permite depurar problemas filosóficos, exorcizar teorías falsas, tampoco resulta del todo clara, al menos de cuándo ello ocurre válidamente.

En efecto, el hecho de que a las bombas de intuición originales

Dennett postule contra experimentos mentales, o más bombas de intuiciones, lo lleva a uno a pensar que el problema con estas herramientas de pensamiento no tiene que ver con su mecánica, con las operaciones que describen, sino más bien con su ámbito de aplicación (González 2013). Y por ello es que sostengo que las bombas de intuiciones, especialmente en el dominio de lo mental, tienen una suerte de carácter pirrónico, a lo Sexto Empírico, porque los experimentos mentales y contra experimentos mentales conducen a una especie de epojé. Es decir, dado el problema involucrado, estimo que resulta creíble la tesis y la antítesis de igual forma. Tómese en cuenta, por ejemplo, el caso de Mary la neurocientista, que aprendió a ver en blanco y negro, y Robo Mary, la versión de Dennett, en que ella sabe perfectamente qué es el color con base en la descripción física de los colores y en la experiencia de estos.

Incluso así, pienso que él ha variado notoriamente su postura desde 1980 acerca de las bombas de intuiciones.⁸⁷ La próxima sección tiene como propósito dar una visión panorámica de su postura, que parece final.

3. La evolución de “bomba de intuición”: una herramienta exploratoria en el trabajo filosófico

Las bombas de intuiciones son herramientas de la caja de herramientas del filósofo. Esa es la principal tesis que defiende Dennett en relación con este tipo de experimentos mentales. Con relación a esta tesis no podría estar más de acuerdo. La caja de herramientas del filósofo no tiene demasiadas herramientas, por lo que saber y conocer cada una de ellas es útil y necesario.

Sin embargo, también sostiene que las bombas de intuiciones son armas, y que se emplean en la argumentación y en el análisis de problemas. Dado que en nuestra cultura no tenemos gran apego

⁸⁷ Algo ya ha mostrado de este cambio en la comunicación personal que hemos sostenido. Posteriormente a esta, en 2013, ha publicado un libro enteramente dedicado a las bombas de intuiciones.

por las armas, ni tampoco por la 2ª enmienda de la constitución de Estados Unidos de 1791, en que se autoriza el uso de armas a civiles, solo usaré la primera caracterización y lo haré en aras de exponer la última postura de Dennett acerca de las bombas de intuiciones.⁸⁸

Es sorprendente su nueva postura positiva acerca de las bombas de intuiciones, a diferencia de lo mostrado entre 1980 y 1984, con textos como la refutación de la Habitación China, *The Mind's I y Elbow Room*. No obstante, un aspecto que es del todo novedoso acerca de las bombas de intuiciones es que considera a estas como herramientas exploratorias de un problema, que si bien pueden fallar, generalmente encaminan u orientan a la solución de un dilema filosófico. Por eso la frase que abre la introducción de su libro es, si bien de Perogrullo, muy cierta: “El pensamiento es difícil” (Dennett 2013, p. 9). Por ello, las tentaciones de salirse del camino son múltiples, e incluso el abandono de la búsqueda parece siempre una alternativa cómoda.

La filosofía es una actividad que requiere pensar problemas difíciles, la mayoría de resolución muy intrincada. La tentación de abandonar el camino siempre está presente, y es difícil encontrar atajos que realmente lleven a la meta. Las bombas de intuiciones son, según el último Dennett, una especie de prótesis de la imaginación, son potentes focos para centrarnos y concentrarnos en un problema filosófico y, especialmente, en el marco teórico en el que se da. Para tales efectos se necesitan herramientas, como menciono arriba, pero hay que tener claro algo: no se pueden usar mal, ni menos comprar a la vuelta de la esquina. Según Dennett, el filósofo se parece al herrero en que tiene que hacer sus propias herramientas para resolver problemas.

⁸⁸ Pero, aquí haré una confesión, que me parece importante para el análisis de la segunda sección de este ensayo. Debo confesar que el libro *Intuition Pumps and Other Tools for Thinking* me sorprendió de manera doble. Lo hizo, en primer lugar, porque Dennett publicó un libro sobre las bombas de intuiciones luego de haber sostenido comunicación personal sobre el tema. Y lo hizo luego de haberme confiado traducir y publicar un artículo intitolado *Intuition Pumps*, que precisamente cito aquí.

Este pensamiento rememora la historia de Babbage, en los albores de la Inteligencia Artificial, en el siglo XIX. En ese entonces no podían comprarse herramientas a la vuelta de la esquina, ni en la multitienda preferida. Claro que no. Babbage, quien quiso construir La Máquina de las Diferencias para erradicar el error humano de las tablas de cálculo, tuvo que hacer y, más aún, diseñar gran cantidad de herramientas con torneros y una variopinta gama de maestros. No había medidas estándar, ni trabajo hecho en serie para construir dichas herramientas, de modo que Babbage tropezó una y otra vez con problemas técnicos, humanos, sindicales y hasta financieros. La consecuencia fue que, gracias a su tesón y esfuerzo, se pudo completar una porción de La Máquina de las Diferencias, la cual trabaja tal como se previó: sin esfuerzo y sin producir ningún error matemático. De esta manera, Babbage logró crear una serie de herramientas para conseguir, medianamente, su objetivo inmediato. No pudo cumplirse todavía su objetivo de largo alcance: erradicar el error humano de las matemáticas mediante la mecanización del pensamiento.

El filósofo está en la misma posición de Babbage, qué duda cabe. Lo está porque tiene que diseñar y hacer sus herramientas teóricas y conceptuales a partir de la nada, punto que remarca Dennett de la siguiente manera:

Como todos los artesanos, un herrero necesita herramientas, pero –de acuerdo con una vieja observación (ciertamente casi extinta)– los herreros son únicos en que hacen sus propias herramientas. Los carpinteros no hacen sus sierras y martillos, los sastres no hacen sus tijeras y agujas, los plomeros no hacen sus llaves, pero los herreros pueden hacer martillos, yunques, espumillones y tongas de metal. ¿Qué hay de las herramientas de pensamiento? ¿Quién las hace? Y, ¿de qué son hechas? Los filósofos

han confeccionado algunas de las mejores —a partir de nada más que ideas, estructuras útiles de información— René Descartes nos dio las *Coordenadas Cartesianas*, los ejes de X e Y sin los cuales el *cálculo* —una herramienta de pensamiento *par excellence* simultáneamente inventada por Isaac Newton y el filósofo Gottfried Wilhelm Leibniz— habría sido imposible. (Dennett 2013, p. 10).

Hay herramientas de pensamiento simples. Por ejemplo, Dennett nombra algunas conocidas tales como etiquetar, dar ejemplos, proponer analogías y metáforas y, finalmente, construir el andamiaje de ubicación de un problema. Los experimentos mentales son, justamente, herramientas típicas de la filosofía. En general, son medios efectivos porque, en vez de tener un ambiente de laboratorio, emplean la inferencia a partir de un marco teórico, el cual da lugar a una pregunta o problema. De hecho, concuerdo con Dennett: algunos experimentos mentales son analizables como argumentos rigurosos, los cuales toman la forma de reducciones al absurdo de una teoría (retomo este punto abajo). En consecuencia, hay herramientas de trabajo simples, como los experimentos mentales, que son típicamente argumentativos.

No obstante, coincido con Dennett también en que hay otros experimentos mentales que son mucho menos rigurosos, pero igualmente efectivos. Las bombas de intuición son ese tipo de experimentos mentales, pues tienen como objetivo generar el efecto Ajá! O bien el del pulgar hacia arriba del lector, quien está presto a decir: ¡Claro, cómo no se me había ocurrido!

En relación con estos dos efectos, el primero postulado por mí y el segundo por Dennett, es importante destacar que este sostiene que: “algunos pensadores han concluido que quiero decir algo despectivo o desdeñoso con el término. ¡Por el contrario, *amo las bombas de*

intuiciones!” (Dennett 2013, p. 11, énfasis mío). Es curiosa esta afirmación, dada la postura original de Dennett en que se considera a algunas bombas de intuiciones como aduladores de masas y como narrativas que desbordan los límites de la imaginación.

Agrega en su libro que las bombas de intuiciones han jugado un rol central en la Historia de la Filosofía, que uno no podría olvidar. Esto lo corroboran ejemplos como La Caverna de Platón, Menón y el niño esclavo, el genio maligno de Descartes, el Estado de la Naturaleza de Hobbes y muchos otros.

Independientemente de la curiosidad que suscita usar el término “adulador de masas” cuando se acuña el concepto bomba de intuiciones, y de cómo parece coincidir la primera caracterización con el término denettiano, es claro que este filósofo le da un valor epistemológico a errar en la filosofía, y ello justamente lo destaca en su libro. Errar es una herramienta para pensar, según él. Desde mi perspectiva, agregaría que no es el errar mismo la herramienta para pensar, sino el que somos conscientes de que erramos. Más aún, los filósofos somos inseguros de una manera obsesiva, cuestión que es dramática porque tenemos la tendencia a equivocarnos y a errar. Aquí vale la pena citar nuevamente el ejemplo de Babbage, quien consideraba que era necesario erradicar el error humano de las matemáticas. Como dice el refrán *errar es humano*. Y agregaría, hacernos conscientes de nuestros errores, que es el punto de Dennett, es cuasi divino, porque nos incentiva a pensar mejor, de manera más clara y precisa, un problema que de suyo es resbaladizo. Errar y dar en el clavo son parte de la discusión crítico racional, del aprendizaje necesario para llegar a la verdad. En relación con la necesidad de cometer errores, Dennett afirma lo siguiente:

A veces Ud. no quiere correr el riesgo de cometer errores; Ud. en verdad quiere cometerlos – si solo se le diera a Ud. algo claro y detallado que fijar. Cometer errores es la clave para progresar. Por supuesto hay momentos en que es realmente

importante no cometer errores – pregúntele al cirujano o al piloto de avión. Pero hay momentos en que cometer errores es la única manera de ir hacia adelante [...]

Nosotros los filósofos somos especialistas en cometer errores. (Sé, suena como un mal chiste, pero escúcheme). Mientras que otras disciplinas se especializan en obtener la respuesta correcta a sus preguntas definicionales, nosotros los filósofos nos especializamos en maneras de hacer que las cosas se mezclen de tal forma, tan erradamente, que nadie está del todo seguro acerca de cuáles son las preguntas, menos las respuestas. (Dennett 2013, p. 18)

Creo que Dennett nuevamente está en lo cierto. Hay veces en que tentamos diferentes soluciones a riesgo de errar, solo con el fin de orientarnos acerca de un problema. Por ejemplo, los casos de algoritmos que no llevan a la solución de un problema, y entonces seguimos los pasos, erramos y al fin nos decimos a nosotros mismos: “esto no lleva a ninguna parte”. A quien le quepa alguna duda, le sugiero hacer el ejercicio de sumar dos números impares para llegar a un número impar de manera algorítmica, es decir, de forma recursiva. Quien lo haga, y yerre, se dará cuenta que los errores lo han hecho llegar a la conclusión de que el problema no tiene solución. Por tanto, ha avanzado, pero a diferencia de lo que piensa Dennett, estimo que lo ha hecho al hacerse consciente del error que ha cometido. Tal es el poder de la razón, paso necesario para la discusión crítica: requiere de una serie de pasos, algunos dados en falso, para aprender de los errores y así progresar.

En general, los filósofos nos metemos en embrollos de tal calibre que por momentos se pierde del todo la perspectiva de lo que se está investigando. Especialmente, se difumina la pregunta a la que llevó

a tal embrollo. Pero, cuidado. Esto no quiere decir que los filósofos somos una especie de “confusionistas” de los problemas, por lo intrincado de estos, y por el embrollo en que nos hemos metido. Por el contrario, los filósofos, al menos cuando se hace análisis, tenemos la intención de salir del embrollo. Tratamos de tomar la hebra de este y desenrollar la madeja, y por esto tentamos caminos que incluso son erróneos.

En vista de la perspectiva de Dennett acerca de cometer errores como una herramienta de pensamiento, queda más clara su posición, ahora positiva, con relación a las bombas de intuiciones. Estas, en efecto, a veces nos hacen cometer errores, nos meten en un tremendo embrollo, a tal punto que la pregunta o el problema a investigar parecen desaparecer. Por esto, creo, se explica la postura negativa de Dennett sobre las bombas de intuiciones relacionadas con el problema de la conciencia. El lío que este término provoca en la filosofía y la diversidad de definiciones que hay sobre el fenómeno es a no dudarlo una advertencia para quien quiera meterse ahí. Por ello alaba tanto sus bombas de intuiciones a propósito de los qualia, o de la experiencia consciente. Ellas supuestamente refinan el ámbito de discusión, al mostrar la madeja en la que nos han metido algunos filósofos de la conciencia.

No obstante, considero que esa no es la única herramienta dennettiana de pensamiento que redime las bombas de intuiciones. Piénsese en lo que mencioné más arriba acerca de las reducciones al absurdo, en relación con los experimentos mentales, en este caso, negativos. Se les llama así porque su objetivo es refutar una teoría y se lo hace más o menos así: en aras de la argumentación, se asume la teoría T. De esta se desprenden mediante inferencia enunciados hasta que se llegue a dos que se contradigan flagrantemente. Si ello ocurre, entonces la teoría T debe ser falsa.

Mucho se ha escrito acerca de esta propiedad de algunos experimentos mentales. Un caso clásico es el de Galileo y los objetos

que caen a la misma velocidad en el vacío. En este experimento mental se asume, en aras de argumentar, la teoría aristotélica. De esta se infiere que si una piedra A, más pesada, se amarra a una B, más liviana, y ambas forman una unidad, deberían caer primero que la piedra A sola (esto se sigue de que A+B se supone más pesado que A). Pero también es posible inferir que la inercia que B provoca a A hace que A + B caiga luego de A sola. Luego, la teoría aristotélica debe ser falsa. Más aún, se puede concluir una cuestión positiva, a saber, que los objetos caen a la misma velocidad en el vacío. Pero, ¿pasa esto con las bombas de intuiciones?

Estas son narrativas pintorescas que llevan a conclusiones de manera más o menos informal, aunque es posible formalizar aquellas como argumentos. Me inclino a pensar que no son herramientas de reducción al absurdo, y ello lo hago pese a que Searle declara abiertamente que la Habitación China es justamente eso (1980). Lo declara así porque se asume una teoría T, el cognitivismo, y según él se analiza de qué manera funcionaría la mente según tal teoría. Y de ahí se construye todo el experimento mental: Searle está en la habitación con el libro de reglas y los bancos de datos y manipula símbolos o ideogramas en chino. Sin embargo, no entiende nada de lo que manipula. Creo que, puesto de esta manera, la Habitación China podría describirse como una reducción al absurdo. Aunque estimo que la teoría T es más bien socavada, al brindarse razones para *dudar* acerca de su poder explicativo.

Al parecer Dennett y compañía tienen razón. Girar las perillas del experimento hace que este ya no se vea de una manera tan clara como una reducción al absurdo. Tómese en cuenta la respuesta del robot del propio Dennett y Fodor o la respuesta del sistema. Ambas producen el efecto Ajá! del que hablaba antes, pero solo lo hacen en una parte de la comunidad. Esto es fundamental, porque una reducción al absurdo es, de alguna forma, un argumento tipo *modus tollens* que no suscita grados de disenso, como sí ocurre con algunas bombas de intuiciones. Por ejemplo, en el caso del experimento mental de Galileo que he

citado, no hay disenso con relación a la conclusión, y por lo mismo nadie comienza diciendo: ¿y qué hubiera pasado si una cadenita que une a los objetos A y B se soltara? Tal pregunta simplemente no cabe. Luego, no es tan claro que la Habitación China cuente como una reducción al absurdo, al menos tal como se entienden estas de manera clásica. ¿Y a qué otras herramientas recurren las bombas de intuiciones?

Diría que varias bombas de intuiciones recurren a la navaja de Occam, es decir, a la herramienta de pensamiento que cercena todo aquello que atenta contra la elegancia y la parsimonia. Por ejemplo, el experimento de los antipodeanos de Rorty utiliza tal maniobra, y lo hace del siguiente modo: supóngase que en un extremo de la galaxia viven seres como nosotros, esto es, que lanzan bombas y escriben poemas, que sienten deseo sexual, se casan y se enamoran. Sin embargo, estos seres, los antipodeanos, no tienen conceptos como mente, ni tienen estados mentales. Y ello ocurre porque entre los primeros conocimientos que adquieren están los de la neurología y la fisiología, a tal extremo que cuando un bebé antipodeano llora cuando nace, ellos dicen: ¡Miren, se han activado sus fibras C! De este modo, la bomba de intuiciones descrita *prescinde* del vocabulario asociado a la mente, a un grado, que simplemente hablar de la mente entre los antipodeanos resultaría absurdo.

Diferentes bombas de intuiciones emplean diferentes herramientas de pensamiento, como las descritas por Dennett (2013). Una última, negativa, es la que él llama el bloque mental [*the mental block*]. Consiste en el uso del adverbio “Ciertamente” o *Surely*, en inglés. Según él, es una herramienta más bien retórica que sirve para evaluar o desestimar argumentos o teorías. Ned Block la usaría para desestimar su teoría de la conciencia del siguiente modo:

Pero con certeza [*surely*] no es nada más que un hecho biológico acerca de la gente — no una construcción cultural— que algunas

representaciones del cerebro perseveran suficientemente para afectar la memoria, el control de la conducta, etc. (Block en Dennett 2013, p. 37)

Aquí Dennett parece errar el tiro, pues lo que Block propone es mucho más que una estrategia retórica. Lo que postula Block es que la teoría de la conciencia de Dennett *debe* adecuarse a ciertos hechos biológicos relacionados con cómo el cerebro representa. Una argumentación parecida puede emplearse contra la teoría de la conciencia de Dennett y Kinsbourne: los fallos acerca de la conciencia que ellos atribuyen son más bien problemas relacionados con la memoria y la atención, dos instancias de la cognición. En realidad, lo que Dennett y Kinsbourne hacen es, desde mi perspectiva, inflar la conciencia con la memoria y la atención para mostrar que es falible, algo similar a lo que el primero hace con los qualia. Con relación a ellos, él considera que los filósofos de la conciencia piensan que son infables, cuestión discutible. Hasta donde sé, nadie ha dicho que tener un *quale* es estar en una suerte de estado cuasi místico, con una ausencia total de palabras. Y tampoco nadie parece sostener que lo epistémico, el juicio acerca de los *qualia*, es idéntico a *tener* dichos estados mentales.

De todas formas, es claro que la nueva posición de Dennett en relación con las bombas de intuiciones le permite un análisis caso a caso de ellas, evitando así tener un criterio más o menos delimitado y, por lo mismo, excesivamente rígido. Ello recuerda a un fan de los “caso a caso” que no tenía un criterio con el cual discernir si A era B. Al final, este “caso a caso” arriesga sumirlo a uno en lo arbitrario. Considero que algo muy similar sucede en la epistemología de los experimentos mentales: un experimento puede ser negativo, explicativo, negativo y explicativo, pero ello no requiere de una interpretación caso a caso sin tener, justamente, un criterio para determinar cuándo un experimento mental es tal y tal cosa. No hay, a mi entender, un progreso genuino en la epistemología de los experimentos mentales sin que hayan

etiquetas claras para poder, entonces, juzgar mediante el caso a caso. En el fondo, si la herramienta de errar es efectivamente positiva, las bombas de intuiciones nunca serían del todo negativas, pese a lo que el mismo Dennett originalmente declara.

Conclusión

En este ensayo he hecho una revisión del término “bomba de intuición”, introducido por Dennett en 1980 y luego por él y Hofstadter en 1981. La motivación de dicha revisión ha sido mostrar que el término ha evolucionado, teniendo un cariz negativo entre 1980 y 1984, y luego más positivo. El hecho de que trate a algunos experimentos mentales como “aduladores de masas” cuenta como evidencia del juicio negativo que hubo alguna vez con respecto a las bombas de intuiciones. Al comienzo no parece identificarse ninguna virtud epistémica asociada a las bombas de intuición. Que estas hayan sido acuñadas gracias a la controvertida y vapuleada Habitación China confirma la primera parte de mi análisis. Parte de mi examen previo a este ensayo se concentró en asociar las bombas de intuiciones, especialmente las de la filosofía de la mente, al problema mente-cuerpo (González 2013).

Sin embargo, considero que la evolución en el pensamiento de Dennett, especialmente a propósito de las herramientas de pensamiento, lo ha llevado a tener una visión más positiva de las bombas de intuiciones. En efecto, que podamos errar, rasurar con la navaja de Occam, usar la escoba de este, utilizar el bloque mental, entre otras, muestra que, incluso en el caso de las bombas de intuiciones, es más lo que aprendemos, filosóficamente hablando, que lo que fallamos. La imaginación y la narrativa de las bombas de intuiciones, he argumentado en la sección 2 de este trabajo, es positiva, incluso si nos embarca en un viaje sin destino claro. Tal como no hay martillos sin propósito o función, no hay bombas de intuiciones sin criterio claro para la búsqueda de conocimiento mediante discusión racional. Para concluir llamo la atención sobre la última línea del ensayo de

Dennett en esta compilación, su gran *finale*: “Las bombas de intuición son parte de la caja de herramientas del filósofo, y si usted quiere ser un trabajador confiable, usted debe entender cómo funcionan sus herramientas” (Dennett, en esta compilación, p. 121).

De alguna manera la finalidad de cada herramienta es normativa. Puede destapar un lavabo con un destornillador, pero no es conveniente hacerlo. Lo hará mucho mejor con una llave inglesa y con paciencia, con mucha paciencia, al igual que con los problemas de la filosofía analizados a la luz de las llamadas bombas de intuiciones. Considero que estas, por sus enseñanzas y por la discusión racional a la que invitan, son connaturales al ejercicio mismo de nuestra actividad: el filosofar.

Referencias

- Churchland, P.M. y P.S. (1990): Could a machine think? *Scientific American* 262 (1), 32-37.
- Dennett, D. (1984): *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: Bradford Books.
- _____ (1988): “Quining qualia”. M. y E. Bisiach (eds.) *Consciousness in Contemporary Science*. Oxford: OUP, pp. 43-77.
- _____ (1991): *Consciousness Explained*. New York: Little Brown.
- _____ (2005): *Sweet Dreams*. Cambridge, MA: MIT Press.
- _____ (2013): *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Co.
- Dennett, D. y Hofstadter, D. (1981): *The Mind's I: Fantasies and Reflections on Mind and Soul*. New York: Basic Books.
- Farrel, B. (1950): “Experience”. *Mind* 59, 170–198.
- González, R. (2013): “Experimentos mentales: en torno a la categoría de “bombas de intuiciones” en la discusión Searle, Dennett y Hofstadter”. *Revista Observaciones Filosóficas No. 15*: <http://www.observacionesfilosoficas.net/>

[experimentosmentales.htm](#)

- Goodman, N. y Putnam, H. (1983): *Fact, Fiction and Forecast*. Cambridge, MA: Harvard University Press.
- Jackson, F. (1986): "What Mary Didn't Know". *The Journal of Philosophy* 83 (5), 291-5.
- Putnam, H. (1975): *The meaning of 'meaning'*. En: *Mind, Language and Reality. Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press.
- Quine, W. V. O. (1960): "Translation and meaning". En: *Word and Object*. Cambridge, MA: MIT Press, pp. 26-73.
- Rawls, J. (1971): *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Searle, J. (1980): "Minds, Brains and Programs". *Behavioral and Brain Sciences* 3, 417-24.
- _____ (1990): "Is the Brain's Mind a Computer Program?". *Scientific American* 262 (1), 26-31.
- Wittgenstein, L. (1953): *Philosophical Investigations*. Oxford: Blackwell.

Conocimiento Subjetivo⁸⁹

Tim Crane

It is obvious that a man who can see knows things which a blind man cannot know, but a blind man can know the whole physics. Thus the knowledge which other men have and he has not a part of physics.

Bertrand Russell (1927, p. 389)

1. Introducción

Nuestro conocimiento depende del contexto de diversas formas. Algunos contextualistas acerca del conocimiento, por ejemplo, creen que las aserciones de conocimiento [*knowledge claims*] deberían examinarse en relación con cierta clase de contextos dialécticos. Aquí trataré con otro tipo de dependencia contextual: la forma en que tenemos conocimiento solo es *accesible* o *disponible* desde el interior de ciertos contextos.⁹⁰ Estos son los contextos de la experiencia de un sujeto y de su posición en el mundo; por esta razón, llamo a la clase de conocimiento que es dependiente de contextos de esta manera, *conocimiento subjetivo*. En este trabajo argumentaré que parte del perenne interés del famoso “argumento del conocimiento” de Frank Jackson, y el experimento mental en el que se apoya —reside en su defensa de la idea de que existe tal conocimiento (Jackson 1982).

⁸⁹ Una parte de este trabajo fue previamente publicado como “Subjective facts” en Lillehammer y Rodríguez-Pereyra (2003). Estoy agradecido a Brian McLaughlin por extensos comentarios y por haberme sugerido el epígrafe de Russell; y a Katalin Farkas, Carsten Hansen, Penélope Mackie, Greg McCulloch, Harold Noonan, Alex Oliver y James Tartaglia por discusión y críticas. Me gustaría dedicar este trabajo a la memoria de mi amigo Greg McCulloch (1951-2001).

⁹⁰ Véase por ejemplo David Lewis (1999).

Mucho se ha escrito acerca del argumento de Jackson; mi duda respecto de añadir un poco más al debate ha sido superada a causa de mi convicción de que la lección correcta del argumento tiene que todavía ser captada.

El argumento del conocimiento está oficialmente diseñado para mostrar, desde un experimento mental, mediante premisas aparentemente no controvertidas y razonamiento simple, que la concepción fisicista del mundo es falsa. D. H. Mellor correctamente señala que si es sólido, el argumento mostraría más que eso: mostraría que algunos hechos son *subjetivos*, y así que una postura según la cual todos los hechos son *objetivos* sería falsa.⁹¹ Por lo tanto, si el argumento tiene éxito, la visión de que la ciencia objetiva puede establecer todos los hechos es falsa también. Mellor, junto con otros, piensa que tiene que mostrar que el argumento no es sólido, debido a que considera que no puede aceptar su conclusión. Él por lo tanto adopta la hipótesis de la habilidad de Lewis y Nemirow, la cual tiene como objetivo mostrar que el argumento del conocimiento es falaz, y que descansa en una noción equívoca de “conocimiento”.⁹²

Argumentaré aquí, contra Lewis, Nemirow y Mellor, que la hipótesis de la habilidad es errónea, y que todo lo otro que intenta el fisicista en aras de rechazar el argumento (o como inválido o como no sólido) está igualmente equivocado. El argumento del conocimiento es un argumento sólido respecto de la conclusión de que existen hechos subjetivos: hechos acerca del carácter subjetivo de la experiencia. No obstante, a diferencia de otros defensores del argumento,⁹³ no estimo que la conclusión

⁹¹ Véase Mellor (1992-3). Dado que mucho de lo que sigue depende de qué significan “objetivo” y “subjetivo”, el lector tendrá que esperar para un enunciado más preciso de lo que aquí pongo. Estoy en deuda con Mellor por haberme alertado de la generalidad de la conclusión del argumento del conocimiento; discuto las propias ideas de Mellor en detalle en Crane (2003).

⁹² Véase Lewis (1990), quien reconoce su deuda con Lawrence Nemirow (1990).

⁹³ Véase Robinson (1982) y el artículo original de Jackson. Él ha cambiado su postura (Jackson 1995 y 1998). Si estoy en lo correcto en lo que afirmo aquí, él no necesitó variar su posición acerca de la solidez del argumento, incluso luego de su conversión al fisicismo; el solo necesitaba re-describir la conclusión. Entre las múltiples discusiones de Jackson respecto de las cuales tengo deuda, debo destacar Horgan (1984).

amenace alguna versión plausible del fisicismo. Los fisicistas deberían aceptar que hay hechos subjetivos, y deberían negar, entonces, que todos los hechos son objetivos, en el sentido que elucidaré. Ni la ciencia física, ni ninguna otra ciencia objetiva, puede establecer todos los hechos, pero esto no puede ser la base de una crítica al fisicismo.

2. El argumento del conocimiento

El famoso argumento de Jackson no se mueve desde una aserción acerca de la existencia de la experiencia a la refutación del fisicismo; se mueve desde la aserción acerca de cómo conocemos acerca de nuestra experiencia a la negación del fisicismo; de ahí su nombre.⁹⁴ El argumento empieza con un experimento mental en que Mary, quien ha estado toda su vida confinada a una pieza en blanco y negro, nunca ha visto ningún otro color, salvo el blanco y el negro. Ahora imagínese que Mary ha realizado un estudio intensivo de la ciencia del color en todos sus aspectos —la física, al fisiología, la psicología y así sucesivamente. De hecho, supongamos que ella sabe todos los hechos físicos acerca del color. Ahora supóngase que Mary un día abandona su pieza blanco y negra, y la primera cosa que ve es un tomate rojo. Es natural sostener que ella ahora sabe algo que no sabía en la pieza en blanco y negro: cómo es ver rojo. Sin embargo, esto que ahora sabe no es un hecho físico, debido a que por hipótesis sabía todos los hechos físicos en la pieza en blanco y negro. Así, si un nuevo conocimiento es un hecho novedoso, entonces Mary aprende un hecho novedoso cuando deja la pieza blanco y negro. Si el fisicismo es (como parece suficientemente plausible) la visión de que *todos los hechos son hechos físicos*, entonces parece que el fisicismo es falso.

El argumento del conocimiento no comete una petición de principio contra el fisicismo. Esto es claro si representamos sus premisas y la conclusión tal como a continuación:

⁹⁴ Véanse las referencias al enunciado del argumento de Jackson y Robinson en la nota anterior. En esencia el argumento tiene una historia más larga que esta, por supuesto. Fuentes de principios del siglo XX son Herbert Feigl (1958, p. 68) y Broad (1926, p. 71).

- (1) En la pieza, Mary sabe todos los hechos físicos acerca del color
- (2) Luego de dejar la pieza, Mary aprende algo nuevo sobre el color
- (3) Luego, no todos los hechos son hechos físicos

Esto, en esencia, es el argumento —aunque se requieren de algunas suposiciones extras para demostrar su validez de modo apropiado. Pero, es claro que ni la premisa (1) ni la premisa (2) de manera obvia cometen una petición de principio en contra del fisicismo. Un fisicista no podría objetar la idea de que alguien que aprende todos los hechos físicos comete una petición de principio en contra del fisicismo. Y (2) parece una cosa irresistible y simple de sostener acerca de cómo la historia se describe arriba. Tal vez, cuando estas premisas se revisan, mostrarán una profunda incoherencia —pero, como el argumento se describe, obviamente no comete una petición de principio.

Los fisicistas han tratado de resistir la conclusión impugnando o la validez del argumento, o la verdad de las premisas. Pienso que se equivocan. Estimo que el argumento es válido, y que los fisicistas deberían aceptar sus premisas. Entonces deberían aceptar su conclusión. Aún así, argumentaré que no deberían preocuparse por su conclusión; pues esta no puede ser que el fisicismo, debidamente entendido, es falso.

En la sección 3 propondré la idea de que el argumento del conocimiento o depende de, o implica la existencia de *qualia* —en el sentido en el que los *qualia* son objeto de discusión. En la sección 4, examinaré la objeción de que el argumento es inválido, y en la quinta examinaré las objeciones a sus premisas. En la sexta sección postularé lo que considero que el argumento realmente muestra: que hay hechos subjetivos. En la sección 7 examinaré las consecuencias de esta conclusión para el fisicismo.

3. El argumento del conocimiento y los qualia

A veces se afirma que (a) el argumento del conocimiento asume la existencia de qualia, considerados como cualidades de la experiencia (estados mentales o eventos) “intrínsecas, no representacionales, introspectibles”; o que (b) el argumento del conocimiento establece que hay tales qualia. Ninguna de estas afirmaciones me parece correcta. De hecho, me parece que la cuestión de los qualia es *irrelevante* al argumento del conocimiento, en el sentido de que para aceptar el argumento, uno ni necesita asumir la existencia de qualia ni necesita aceptar que el argumento muestra que existen qualia — incluso si el argumento es del todo exitoso. Las suposiciones (a) y (b) son falsas. (Por supuesto, a veces los “qualia” se toman como un término alusivo a la experiencia mental consciente; en este sentido, el argumento asume los qualia. Pero este no es el sentido de “qualia” en que hay debate acerca de la existencia de qualia).⁹⁵

La forma simple de ver que (a) es falsa es observar que si uno fuera un seguidor de la intencionalidad con relación a la conciencia fenoménica, y entonces uno negase los qualia, uno no tendría una vía de escape simple para el argumento del conocimiento.⁹⁶ Los seguidores de la intencionalidad acerca de la conciencia fenoménica no piensan que su teoría de la intencionalidad como tal les proporcione una solución al argumento del conocimiento; ellos aún piensan que necesitan refutar el argumento (si desean defender el fisicismo, como lo entiende el argumento).

Pero, tampoco debemos aceptar (b): que el argumento del conocimiento *establece* que hay qualia en el sentido arriba declarado. Solo establecería que hay qualia si (i) la única manera de interpretar el nuevo conocimiento que Mary tiene es en términos del conocimiento acerca de la experiencia, y (ii) la única forma de interpretar su

⁹⁵ Véase mi trabajo (Crane 2000) para una discusión de los sentidos de qualia.

⁹⁶ Para el intencionalismo de la conciencia véase Tye (1995).

conocimiento acerca de la experiencia es en términos del conocimiento de qualia. Pero ninguna de estas movidas es obligatoria. Tomándolas en orden inverso: (ii) Un seguidor de la intencionalidad puede decir que el conocimiento que Mary adquiere es conocimiento sobre la naturaleza de algunos de sus estados intencionales. Al haber visto rojo por primera vez, Mary ahora sabe que es cómo estar en un estado intencional de una cierta clase especial (ver rojo). De acuerdo con las visiones intencionalistas, este estado no involucra qualia, y nada en la historia descrita requiere que uno afirme que lo hace. (i) Una respuesta más inusual pero aún coherente es que Mary aprende algo acerca del mundo cuando ve rojo por primera vez. Ella aprende acerca de las propiedades de las cosas rojas, a saber, aquellas propiedades que solo pueden conocerse al *mirar* las cosas rojas. Una respuesta aparentemente coherente con el argumento del conocimiento es sostener que Mary aprende algo sobre los colores, acerca de las propiedades físicas en el mundo. No estoy tratando de argüir que una u otra de estas visiones es correcta; solo trato de señalar que uno puede aceptar la conclusión del argumento del conocimiento sin aceptar que existen qualia.

Lo que es verdadero es que si uno tuviera alguna razón independiente para creer en los qualia —por ejemplo, que uno fuera persuadido por el argumento de Ned Block “la tierra invertida”— entonces uno podría desear usar una apelación a los qualia al tratar de entender cuál es la lección a partir del argumento del conocimiento: Mary aprende hechos acerca de los qualia (cómo es tener experiencias que los involucren). Ahora, creo que afirmar esto no es correcto. Pero, todo lo que necesito enfatizar por el momento es que *si es verdadero o no*, esta visión no se sigue del argumento como lo he presentado.

Concluyo que el argumento del conocimiento es independiente de la pregunta por los qualia.

4. Desafiando la validez del argumento: “la hipótesis de la habilidad”

Aquellos que desafían la validez del argumento, usualmente argumentan que existe un equívoco con respecto a “saber”.⁹⁷ En la primera premisa, “saber” se usa para expresar conocimiento proposicional, pero (ellos dicen) en la segunda premisa se usa para expresar saber-como o conocimiento mediante habilidad. Deberíamos aceptar que Mary aprende algo nuevo, pero lo que ella aprende cuando ella ve rojo por primera vez, es cómo reconocer el rojo, imaginar rojo y recordar experiencias de rojo.⁹⁸ Al haber visto algo rojo, ella ahora reconoce el color de los carros de bomberos, ella puede considerar si quiere pintar su pieza roja, y ella puede recordar la experiencia decisiva de encontrarse con un tomate. Estas son habilidades cognitivas, no aserciones sobre conocimiento proposicional, y es una visión ampliamente considerada que no existe reducción de conocimiento por habilidad a conocimiento proposicional. Entonces, Mary puede aprender algo nuevo —en el sentido de adquirir una habilidad— pero no es un nuevo conocimiento proposicional. Saber cómo es ver el rojo, es saber-como. Luego, el argumento del conocimiento es inválido, porque involucra una falacia de ambigüedad: “saber” significa algo distinto en las dos premisas. Debido a que es solo en el caso del conocimiento proposicional que los objetos de conocimiento son hechos —si sé como andar en una bicicleta, *sé cómo andar en bicicleta* no es un hecho— se concluye que Mary no llega a saber ningún hecho novedoso y que el fisicismo está a salvo.

Esta respuesta, conocida como “la hipótesis de la habilidad”, presupone dos cosas: (i) que el saber-cómo es un conocimiento de habilidad, y que es completamente diferente, e irreductible al conocimiento proposicional; y (ii) que a pesar de la habilidad que adquiere, Mary no llega a saber ninguna nueva proposición en lo

⁹⁷ Para un listado de respuestas al argumento del conocimiento, véase Robert Van Gulick (1997).

⁹⁸ Véase Lewis (1990), Nemirow (1990) y Mellor (1992-3).

absoluto. La primera aseveración (i) es una consideración teórica general sobre la relación entre el saber-como, las habilidades y el conocimiento proposicional. Esta aseveración es incluso más dudosa de lo que normalmente se asume; pero por economía no entraré a analizar esto aquí.⁹⁹ En vez, me concentraré en la segunda aseveración (ii).

Los defensores de la hipótesis de la habilidad, sostienen que Mary no aprende ningún conocimiento proposicional del todo. Pero esta suposición es en verdad bastante implausible. Pues existe una forma bastante natural para Mary de expresar su conocimiento de cómo es ver rojo. “Ajá, el Rojo se ve así”. (Supongamos en aras de la simplicidad que Mary sabe que los tomates son rojos, y sabe que está viendo un tomate; estas son suposiciones inocuas). Ahora “el Rojo se ve así” es una oración en modo indicativo; en un contexto particular, con seguridad expresa una proposición; y en el contexto descrito, la proposición es verdadera. (Podría haber sido falsa. Supóngase que a Mary se le muestra un tomate falso pintado de azul; la proposición expresada por “el Rojo se ve así” sería falsa. El Rojo no se ve de esa manera). Y es una proposición que Mary no sabía antes. Todo esto asume que una oración que incluye un demostrativo puede usarse para expresar una proposición; pero esta asunción es inocua y debería aceptarse por todos los participantes del debate (podemos ver su completa relevancia más abajo). Así, *incluso si* Mary adquiriera un montón de saber-como, *e incluso si* el saber-como es esencialmente diferente del conocimiento proposicional, aún así existe algo que ella aprende que no podría haber aprendido antes. Y eso es suficiente para que el argumento tenga éxito.

Se apoya la visión de que hay una proposición que se aprende mediante la observación de Brian Loar de que se puede razonar usando la oración “El Rojo se ve así”: puede incluirse en un condicional, por ejemplo, “Si el Rojo se ve así, entonces el rojo se ve así para los perros o no lo hace”. Superficialmente, esto es un condicional de la

⁹⁹ Para una excelente discusión de este punto, ver A. W. Moore (1997, p. 171) y Paul Snowdon (2003).

forma “Si p entonces q”; los sustituyentes para p y q son portadores de valores de verdad y entonces objetos posibles de conocimiento proposicional (Loar 2002, p. 607).¹⁰⁰ La hipótesis de la habilidad tiene que despejar esto si apoya la conclusión de que nada proposicional se aprende. Dudo que esto pueda ser llevado a cabo. Por todas estas razones, rechazo la hipótesis de la habilidad.

Una forma alternativa de cuestionar la validez del argumento es afirmar que el conocimiento que se adquiere es conocimiento por familiaridad.¹⁰¹ Mary se familiariza con alguna característica de la rojura (cómo se ve) o con alguna característica de su experiencia (qualia, como podría ser). El conocimiento por familiaridad no es reductible a conocimiento proposicional; pero estas características (de la rojura, o de las experiencias) podría no obstante ser física. A esta objeción, mi respuesta es esencialmente la misma que mi réplica a la hipótesis de la habilidad: a menos que el crítico pueda mostrar que Mary *no aprende* ningún conocimiento proposicional, entonces el hecho de que adquiera familiaridad es irrelevante a la conclusión del argumento. Y tenemos un ejemplo perfectamente claro del tipo de proposición que aprende: la proposición expresada por la oración “el Rojo se ve así”.

Mellor piensa que “la hipótesis de la habilidad” refuta el argumento del conocimiento; él también sostiene que Nagel se equivoca acerca de los límites del conocimiento objetivo:

Estos no son los de otro modo misteriosos hechos que la teoría del saber-como explica. También explica la misteriosa inhabilidad de la ciencia, la cual tanto impresiona a Nagel, saber cómo son las experiencias de un murciélago con radar. Pero, de acuerdo con la teoría del saber-como

¹⁰⁰ Debo ignorar aquí el asidero que este punto tiene respecto del famoso problema Frege-Geach.

¹⁰¹ Esta línea toma Paul M. Churchland (1995).

esto no es misterioso, ni una limitación sobre el alcance factual de la ciencia objetiva. Pues el único conocimiento que cualquier ciencia nos da es conocimiento de hechos. E incluso si muchas habilidades dependiesen de saber hechos, siempre hay más en saber aquellas habilidades que en saber tales hechos (Mellor 1992-3, p. 7).

Pero, si la hipótesis de la habilidad es falsa, entonces no puede explicar por qué Nagel está equivocado respecto del “alcance factual de la ciencia objetiva”. Ciertamente, parece más bien que existen hechos acerca de la experiencia del murciélago (asumiendo que tiene experiencias) que están más allá del alcance de la ciencia objetiva: los hechos que verdaderamente se expresarían (*per impossibile*) al decir “Experimentar el mundo desde el punto de vista de un sonar es como *esto*”. O tómese un ejemplo de la vida diaria, el hecho que puedo expresar cuando digo “El Rojo se ve así” es un hecho que una persona ciega no puede conocer. Aún así, como Russell señala, una persona ciega puede saber un montón de física. Y no hay nada que detenga a la persona ciega a aprender la totalidad de la ciencia objetiva. Es cierto que la persona con vista tiene habilidades que la persona ciega no tiene, y Mellor está en lo correcto en que no hay ciencia suficiente que pueda otorgar estas habilidades. Pero, esto es irrelevante. El punto que importa no es si existen habilidades que alguien sabe cuando tiene saber-como; el punto importante es que alguien que tiene saber-como sabe que *ciertas cosas son el caso*. Esto es conocimiento proposicional que una persona con vista tiene y que un no-vidente desconoce, además de cualquier otra habilidad que pueda tener.

5. Desafiando las premisas

Rechazo, por lo tanto, estos intentos de disputar la validez del argumento; este es válido. Pero, ¿qué hay con las premisas? Pocos fisicistas desean desafiar la primera premisa, de como se narra en la historia, Mary sabe todos los hechos físicos acerca de la visión

en colores.¹⁰² Pues supóngase que el fisicista negara esto. Entonces, tendría que aceptar que existen ciertos *hechos físicos* que en principio no pueden saberse sin haber tenido ciertas experiencias. La física, la ciencia que establece *los hechos físicos*, sería en principio *no completable* hasta que experiencias muy específicas se tuvieran. Ahora, sería cierto que tener conocimiento en general requiere tener experiencias de algún tipo. Aún así, ¿cómo puede el fisicismo, que basa su postura epistemológica en la ciencia física, requerir que la ciencia necesite experiencias *específicas*? La sugerencia tiene poca plausibilidad.

Así, en vez de eso la mayoría de las respuestas al argumento han desafiado la segunda premisa, y han argumentado que Mary no aprende ningún hecho novedoso. En una investigación reciente, Güven Güzeldere describe este carácter de la respuesta dominante como sigue:

El problema pivote aquí es si el tener una experiencia constituye una clase especial de irreductibilidad a “hechos de la primera persona” o si lo que Mary carece tiene que ver con “el modo de acceso” a hechos con los cuales ella ya está familiarizada (Güzeldere 2002, p. 38).

La idea parece ser que Mary ya sabe todos los hechos en cuestión, ella simplemente gana un nuevo “modo de acceso” (lo que sea que es) a un hecho que ya sabía. Si esta respuesta fuera cierta, entonces el argumento colapsaría. Pero me parece que, a pesar de su popularidad, la respuesta no puede ser correcta.

La idea central es que Mary aprehende o encuentra novedosamente lo que ella ya sabía. La frase “modo de acceso” a menudo se usa para describir qué es esta nueva forma de encuentro. Pero, ¿qué son

¹⁰² Pero, véase Churchland (1995). Dennett (1991) lanza un ataque generalizado contra la metodología de los experimentos mentales como una manera de aprender sobre la conciencia.

“los modos de acceso”? Una forma de entenderlos es en términos de nuevo *modo de presentación fregeana* de los objetos y propiedades ya conocidos bajo otros modos de presentación. De acuerdo con esta interpretación, el puzle del argumento es una pieza más dentro de otros puzles de la intencionalidad, y muchos otros autores han hecho esta comparación explícitamente. Vladimir podría saber que Hesperus brilla en la tarde, pero no saber que Phosphorus no brilla en la tarde. Nosotros no concluimos a partir de esto que Hesperus no es Phosphorus, debido a que como bien se sabe “X conoce p” no es un contexto extensional. De acuerdo con esta postura, el hecho de que Hesperus brille en la tarde es el mismo hecho que el hecho de que Phosphorus brille en la tarde —después de todo, ¿son la misma estrella, el mismo brillo, la misma tarde! Así, aunque Mary sabe que el rojo se ve así, esto no es un hecho novedoso que haya aprendido, sino análogamente, un nuevo modo de presentación de un hecho que sabía antes.

Pero, ¿cuál es el hecho? Necesitamos identificar algo que pueda ser referido en más de una forma, el hecho relevante con relación a cuál puede aprenderse en la pieza en blanco y negro. Una manera de ponerlo es así. Cuando abandona la pieza blanco y negro, Mary juzga que saber *ver rojo es como esto*. El físico dice que ver rojo es estar en un estado neural B, entonces supongamos que Mary sabía esto en la pieza en blanco y negro. Mary puede entonces inferir que estar en un estado neural B es así. Nosotros entonces tenemos dos términos, “ver rojo”, “estar en un estado neural B” el cual identifica la misma cosa, y un predicado “como este” que puede solo usarse cuando uno tiene la experiencia. Pero aún así la experiencia es el estado neural para todo eso.

Hasta aquí vamos bien. Pero recuérdese que la distinción para distintos modos de presentación de la misma cosa supone mostrar que la segunda premisa del argumento es falsa: *Mary no aprende nada nuevo*. Pero no puede mostrar esto. Porque si esta construcción del caso de Mary y el caso de Hesperus y Phosphorus realmente son paralelos, entonces esto implica que alguien que llegue a creer que

Phosphorus brilla en la tarde, debido a su creencia de que Hesperus es Phosphorus, no aprende nada nuevo, pero solo llega a apreciar un hecho previamente conocido bajo un nuevo modo de presentación. Y esto no puede ser correcto: el punto original de la distinción entre sentido y denotación fue hacer justicia al hecho de que el descubrimiento que Hesperus es Phosphorus puede ser un avance significativo en el conocimiento de alguien. Fue un *descubrimiento* acerca de los cielos que Hesperus es Phosphorus, es un nuevo conocimiento que los antiguos aprendieron. Entonces, de la misma manera, el conocimiento de que Phosphorus brilla en la tarde es un nuevo conocimiento. Si son hechos lo que Ud. aprende cuando adquiere conocimiento, entonces la aproximación estándar a la distinción entre sentido y denotación implica que lo que los astrónomos antiguos aprendieron cuando ellos aprendieron que Hesperus es Phosphorus es un hecho novedoso.

Por supuesto, hay *algo* que es lo mismo antes y después de este particular descubrimiento: cómo son las cosas en el mundo, la referencia de los términos, las entidades. Nadie disputa esto sobre el caso de Hesperus-Phosphorus. Así, uno puede decir: “En un sentido los hechos son los mismos, en un sentido son diferentes”. Pero la pregunta relevante, si algo *se aprende* cuando alguien adquiere la creencia de que Hesperus es Phosphorus, si hay nuevo conocimiento del todo. Y si hay un sentido en el cual el hecho aprendido *es* un hecho novedoso (incluso si hay un sentido en el cual las cosas son idénticas también). Entonces, hay nuevo conocimiento. Esto en verdad no puede negarse. Nótese que si uno niega esto, uno tiene que negar *al menos* que existe nuevo conocimiento en el siguiente sentido: el conocimiento de que dos modos de presentación son modos de presentación de la misma cosa (Chalmers 1996). Pero esto hace imposible siquiera establecer qué es lo que los antiguos aprendieron.

Debido a que introdujeron el paralelo sería infructuoso para el fisicismo tratar y deducir alguna diferencia en principio entre el caso de Mary y el caso de Hesperus y Phosphorus. Luego, o el fisicismo dice que nada nuevo se aprende en ambos casos —lo cual es una cosa

no esperanzadora— o afirma que algo se aprende en ambos casos. Esta es la única cosa plausible que puede decirse. Pero, entonces Mary aprende algo nuevo. Las premisas del argumento son verdaderas, y ya hemos decidido que era válido. Luego, ¿se refuta el fisicismo?

6. Hechos físicos y hechos subjetivos

Esto depende, por supuesto, de qué es el fisicismo. Lo que se refuta es la doctrina de que todos los hechos son hechos físicos — dado cierto entendimiento de “lo físico” y “hecho”. El argumento asume una cierta comprensión de qué son “los hechos físicos”.

¿Qué son los hechos? Los filósofos no han acordado sobre la naturaleza de los hechos, y si hay tales cosas. Algunos afirman que los hechos son proposiciones verdaderas, otros que corresponden uno-a-uno con proposiciones verdaderas,¹⁰³ y otros dicen que debido a que es lo que hace que las proposiciones sean verdaderas (son “hacedores-de-verdad [*truth-makers*]) no necesitan adecuarse uno-a-uno con proposiciones verdaderas. ¿Qué postura sobre los hechos asume el argumento del conocimiento? Es obvio, creo, que el argumento del conocimiento tiene que asumir que los hechos son *objetos de conocimiento proposicional* —donde un estado de conocimiento proposicional es descrito en enunciados de la forma “X conoce que *p*” donde X es un agente cognitivo y “*p*” se reemplaza por una oración. Así, para que algo sea un hecho novedoso es que al menos sea un nuevo conocimiento, un avance en el conocimiento de alguien, un nuevo conocimiento que antes no poseía.

¿Significa esto que el argumento del conocimiento veladamente hace una petición de principio contra el fisicismo, al asumir una concepción de *hecho* que el fisicismo rechazaría? No. Si o no el fisicismo decide llamar objetos de conocimiento proposicional a los “hechos”, debería ciertamente aceptar que hay objetos de conocimiento

¹⁰³ Para estas posturas véase Frege (1967), Mellor (1995), Austin (1961) y Davidson (1984).

proposicional, y que los estados de conocimiento se individúan en parte por sus objetos. Todos aceptan que hay tales objetos de conocimiento proposicional, *si o no* ellos también aceptan que hay hechos en algún otro sentido. Luego, pienso que es un error afirmar que necesitamos establecer qué teoría de los hechos es correcta antes de cerrar el debate sobre si el argumento del conocimiento funciona. Esto sería sostener que el argumento debe poseer, como premisa oculta, que una teoría particular de los hechos es la correcta. Pero, esto no es el caso; todos tienen que aceptar que hay objetos de conocimiento proposicional.

La concepción de los hechos, del argumento del conocimiento, no comete ninguna petición de principio. Lo que dice es que una pieza distintiva de conocimiento proposicional es conocimiento distintivo de un hecho. Esto es con seguridad una idea muy natural y poco debatible. Podemos aprender habilidades o nuevos conocimientos; cuando aprendemos nuevos conocimientos, lo que aprendemos son hechos. Pero, a veces se afirma que hay dos nociones de información (o hecho): una noción de grano-grueso [*coarse-grained*] y una de grano-fino [*fine-grained*].¹⁰⁴ Según la última, los hechos son individuados en el nivel del sentido; para la primera, los hechos se individúan en el nivel de la referencia. Nótese que este punto a veces se pone al servicio de la idea errónea (desestimada arriba) de que Mary no aprende nada nuevo, sino que solo adquiere un nuevo “modo de acceso” a lo que ella ya sabía. Si uno usa la distinción entre hechos de grano-fino y de grano-grueso para apoyar esta idea errónea, entonces uno está obligado a decir que *solo* la visión de grano-grueso es relevante a la individuación del conocimiento. Pero, esto es claramente falso, y no algo que un físico deba asumir, por todas las razones expuestas en la sección previa.

En *Facts of Causation*, Mellor (1995) hace una distinción entre hechos y lo que llama *facta*. Los hechos son las “sombras” de las verdades —si es verdadero que *p*, es un hecho que *p*. Los *facta* son los

¹⁰⁴ Véase Van Gulick (1997, pp. 562-3).

hacedores de verdad para las verdades; es una pregunta empírica qué hechos hay, tal como es una pregunta empírica qué propiedades hay. Entonces, no debemos inferir diferencia entre *facta* desde la diferencia entre hechos; los *facta* y los hechos no tienen una correspondencia uno-a-uno. Para los propósitos presentes esta distinción corresponde a la distinción entre los hechos de grano-fino y de grano-grueso. Pienso que debemos estar de acuerdo con Mellor de que ambas nociones de los hechos (o las nociones de hecho y *factum*) tienen su lugar. Esto es consistente con lo que dije arriba, a saber, que los objetos de conocimiento usualmente se individuán de manera de grano-fino. Puede ser que a veces individúemos los objetos de conocimiento de una manera de grano-grueso, esto es perfectamente aceptable también. Pero en tanto lo hacemos de una manera de grano-fino, entonces debemos aceptar la conclusión de que Mary aprende un hecho novedoso.

Habiendo dicho lo que el argumento quiere decir por “hecho”, nos podemos mover a “lo físico”. Lo que se requiere que imaginemos es que el conocimiento que uno adquiere de los colores en la visión de Jackson de la pieza en blanco y negro se establece en el lenguaje de la física. Pero, no ayudaría a Mary si ella aprendiese cosas en la pieza que estuvieran en el lenguaje de la psicología y la fisiología. Ni le ayudaría tampoco si ella aprendiera un discurso totalmente desarrollado en términos psicológicos dualistas (si hubiese algo como eso) acerca de los estados de conciencia, mientras se hace explícitamente consciente de su naturaleza totalmente no-física. Ninguna de estas teorías le serviría a ella respecto de cómo es ver el rojo. El punto no es que el tipo de conocimiento que ella adquiere en la pieza en blanco y negro es conocimiento físico; más bien, el punto estriba en que es el tipo de conocimiento que puede expresarse en una forma u otra: es conocimiento “aprendido de libros”. Como David Lewis lo pone “el punto intuitivo de partida no era solo que las lecciones *de física* no podrían ayudar a quien no ha tenido experiencia a saber cómo es ver. Es que las *lecciones* no pueden ayudar” (Lewis 1990, p. 281).¹⁰⁵

¹⁰⁵ También Mellor (1992-3).

Así, aunque el fisicismo —entendido como el enfoque de que todos los hechos son hechos físicos— es uno de los objetivos del argumento, es realmente una instancia de un objetivo más general: la visión de que todo el conocimiento del mundo es el tipo que puede impartirse en lecciones, sin presuponer ningún tipo de experiencia. De este modo, cualquier visión que se comprometa con esta postura sobre el conocimiento entra en el alcance del argumento del conocimiento. Similarmente, con el dualismo cartesiano —uno no podría saber cómo es ver rojo, el argumento implica, incluso si uno aprendiese toda la teoría cartesiana de la mente.

Paul Churchland ha argumentado que esta característica del argumento muestra que prueba demasiado (Churchland 1995 y 1997).¹⁰⁶ Piensa que el argumento de Jackson involucra una “patología lógica”: “hace cualquier explicación científica de nuestra experiencia sensorial totalmente imposible, sin importar qué ontología se emplee”. Pero esto es claramente un *non-sequitur*: todo lo que se sigue del argumento del conocimiento es que si uno *supiese* toda la explicación científica de nuestra experiencia sensorial, no se seguiría que uno sabría cómo era tener esa experiencia. Esto no implica nada respecto de si tal visión científica completa del funcionamiento de nuestros sentidos puede darse. Ahora, Churchland mismo identifica esto como el problema central, en un punto:

Si funciona del todo, el argumento de Jackson funciona en contra del fisicismo, no porque haya defecto que es único al fisicismo: *funciona porque no hay cantidad suficiente de conocimiento discursivo, sobre cualquier tópico, constituirá el conocimiento no discursivo que Mary carece* (Churchland 1997, p. 574).

¹⁰⁶ Jackson (2002) intenta una respuesta, pero sobre la base implausible de que *hay una diferencia entre la clase de conocimiento que una psicología dualista daría y la clase que una teoría psicológica daría.*

Pero él toma esto como conectado a la aseveración de que cualquier explicación científica de la experiencia deberá ser imposible. Esto, pienso, es un error, por la razón expuesta arriba. (Nótese que debido a que pienso que Mary adquiere conocimiento proposicional, no identificaría “discursivo” con “proposicional”).

Es verdad que lo que Mellor (en la cita arriba) llama “el alcance factual de la ciencia objetiva” se muestra estar restringido por el argumento del conocimiento. Pues ninguna explicación científica de la visión le dirá al ciego cómo es ver, y he argumentado que lo que el ciego carece aquí es (además de saber-como y conocimiento por familiaridad) conocimiento proposicional. Estos conocimientos proposicionales —estas clases de hecho— son lo que la ciencia objetiva no puede expresar. Pero nadie esperaría que lo hiciera; esto no debe verse como una “restricción” misteriosa sobre el poder de la ciencia.

Concluyo que no hay falacia en el argumento del conocimiento; pero tal vez ahora estamos comenzando a ver que su conclusión se establece de una manera errónea, i.e., como una objeción al *fisicismo*. Porque incluso si este es la visión de que todos los hechos son hechos físicos, el argumento del conocimiento es una objeción más que esto (hasta aquí Churchland está en lo correcto). Es realmente una objeción a la postura de que todos los hechos son, por decirlo así, hechos “aprendidos de libro”. *Los hechos cuyo aprendizaje no requiere que Ud. tenga una cierta clase de experiencia o que ocupe una cierta posición en el mundo.* (Como Jackson dice, “Ud. no necesita televisión en colores para aprender física o psicología funcionalista” (Churchland 1997, p. 569)). “Objetivos” sería un buen nombre para estos hechos. Y “subjetivos” sería en consecuencia un buen nombre para *esos hechos cuyo aprendizaje requiere que uno tenga cierto tipo de experiencia, o que ocupe cierta posición en el mundo, etc.* Esto es por qué digo que el argumento del conocimiento es un argumento para la visión de que hay *hechos subjetivos*. Es un argumento que muestra que para adquirir nuevo conocimiento de un cierto tipo —aprender nuevos hechos— Ud. tiene que tener experiencias de un cierto tipo.

Que hay hechos subjetivos en este sentido no debería ser sorprendente. Para otro ejemplo de hecho cuya aprehensión depende de la locación específica del sujeto en el tiempo y espacio, considérese el caso del conocimiento indexical. Tómese en consideración, por ejemplo, Vladimir perdido en el bosque; él consulta su brújula y un mapa y exclama con alivio “¡Estoy aquí!” señalando un punto en el mapa. Cuando Vladimir exclama “¡Estoy aquí!” apuntando el mapa es algo que él aprende. Él ahora sabe dónde está, y no sabía antes. En un artículo clásico, John Perry se describe a sí mismo siguiendo una huella de azúcar alrededor de un supermercado, con el objeto de señalarle al cliente desde quien vino el azúcar que está haciendo el desaguisado (Perry 1979). Cuando Perry se da cuenta de que él estaba haciendo el desaguisado, él aprendió algo que expresa al decir “¡Soy yo! ¡Estoy haciendo un desaguisado!”. Y este conocimiento es distinto del conocimiento que hubiera expresado al afirmar “El cliente con una bolsa perforada de azúcar está haciendo un desaguisado”. Ambos ejemplos de nuevo conocimiento requieren que uno tenga una cierta posición en el mundo: Vladimir y Perry no pueden aprender lo que ellos aprendieron sin ocupar ciertas posiciones o ser las personas que son. En particular, no pueden aprender estos conocimientos, estos hechos, desde libros. ¿Cómo podrían? Algunos escritores han notado aquí la analogía con el argumento del conocimiento. Discutiré esto con profundidad más abajo.¹⁰⁷ Lo que Mary, Vladimir y Perry han todos aprendido son *hechos subjetivos*.

Alguien podría tratar de neutralizar esta conclusión en este punto al apelar a la distinción entre hechos *como verdades* y hechos como *hacedores de verdad*. Tal vez, tal teórico podría admitir que hay hechos subjetivos en el sentido de que hay *verdades* subjetivas, o en el sentido de que hay objetos de conocimiento (en tanto que los objetos de conocimiento son individuados por el sentido en vez de solo por la referencia). Esto es, incluso si el teórico se persuadiera con mi

¹⁰⁷ Para el uso del paralelo con los indexicales como una respuesta al argumento del conocimiento, véase Rey (1997).

argumento de que Mary aprende un hecho novedoso y que su situación es relevantemente como el caso indexical, ellos aún podrían afirmar que esto es solo otra forma de decir que hay *verdades* subjetivas. Lo que realmente importa es que hay negación de *hacedores de verdades* subjetivas (o en la terminología de Mellor, *facta*). Y esto, como muestra la analogía indexical, no es tocado por el argumento del conocimiento.

Pero, ¿qué sería un *hacedor de verdad* subjetivo? Un hecho subjetivo, como lo he definido arriba, es un hecho cuyo aprendizaje requiere que el aprendiz tenga cierta experiencia o que ocupe cierta posición en el mundo. Los *hacedores de verdad*, en cambio, no se aprenden: son lo que hacen *verdadero* a las verdades que son aprendidas. Entonces, tal vez podríamos decir esto: un *hacedor de verdad* subjetivo es el *hacedor de verdad* para un hecho o verdad subjetiva. O: un *hacedor de verdad* subjetiva es lo que tiene que existir para que un hecho subjetivo sea aprendido. (Esto es *grosso modo*, pero no se requiere de algo más preciso aquí). Luego, ¿qué es lo que se necesita que sea el caso para que Mary aprenda que el rojo se ve así? Una parte obvia de la respuesta es: una experiencia visual de rojo. La experiencia visual de Mary del rojo necesita existir si ella va a aprender que el rojo se ve así. Ahora, si un *hacedor de verdad* subjetiva es una experiencia, entonces nadie debería negar la existencia de *hacedores de verdades* subjetivas; pues el problema no es sobre la existencia de *experiencias*. Estas son subjetivas en el sentido de que dependen de la existencia de *sujetos* experimentadores; pero nadie en esta discusión niega la existencia de sujetos experimentadores (e.g. Mary) tampoco. Luego, ¿qué negaría alguien al negar que hay *hacedores de verdades* subjetivas?

La distinción objetiva/subjetiva que hice arriba fue entre distintos tipos de conocimiento. Se admite, es difícil ver cómo claramente se aplica a tipos de entidades. El físico debería ciertamente decir que una de las entidades que constituye el *hacedor de verdad* para el conocimiento de Mary de que *el rojo se ve así* es: la experiencia de Mary del tomate. Y esta experiencia podría llamarse una entidad

subjetiva en el sentido de que es una entidad que depende de un sujeto que experimenta. La experiencia podría denominarse un hacedor de verdad subjetivo, entonces. Así, parece que todos deberían aceptar que hay hechos subjetivos (verdades) y que (en tanto que la idea tiene sentido), hay hacedores de verdades subjetivas también, debido a que hay experiencias. La distinción verdad/hacedor de verdad no ayuda al fisicista a escapar a la conclusión del argumento del conocimiento.

He argumentado que el fisicista debería aceptar que hay hechos subjetivos. La pregunta ahora es cómo esto puede ser compatible con versiones más plausibles del fisicismo; esto es, con versiones que no afirmen que todos los hechos son físicos u objetivos.

7. El fisicismo revisitado y re-descrito

El argumento del conocimiento considera que el fisicismo es la teoría de acuerdo con la cual todos los hechos son físicos. Dado lo que significa “hecho”, esto significa que todo el conocimiento proposicional es conocimiento físico. Y dado lo que significa “físico”, esto significa que todo el conocimiento es la clase de conocimiento que puede aprenderse dentro del escenario de la pieza en blanco y negro —esto es, sin tener que tener cualquier tipo de experiencia particular. Así, el objetivo del argumento es que todos los hechos son hechos “objetivos”. Y esta es la postura que refuta el argumento del conocimiento. Conclusivamente.

Pero, ¿por qué los fisicistas tienen que decir que todo el *conocimiento* es físico en este sentido? Ciertamente, ¿por qué el fisicismo tiene que ser una postura sobre el conocimiento, del todo? El fisicismo es una postura acerca de lo que hay, y derivativamente acerca de cómo lo conocemos. La motivación más fuerte y más clara del fisicismo viene de su tesis de explicar la causación mental (Papineau 2001 y Loewer 2001). Para hacerlo, el fisicismo no requiere comprometerse con la visión de que todo el conocimiento debe ser expresable sin que quien expresa tenga que tener cualquier tipo de experiencias. Solo necesita comprometerse con la idea de que la física

es *causalmente cerrada*, ni siquiera con la postura de que la física es *explicativamente adecuada*.¹⁰⁸ Entonces, el fisicismo no necesita sostener que *la física debe establecer todos los hechos*. (La idea de que debería derivar de la imagen del libro del mundo, con todas las verdades escritas en este en la única historia correcta de la realidad. Pero, la imagen es incorrecta; si lo que digo aquí es correcto, nunca podría existir este libro. Porque no puede expresar la proposición que Vladimir expresa cuando dice: “¡Estoy aquí!” y que Mary enuncia cuando afirma “¡El Rojo se ve así!”).

Es en este punto —en vez de en el intento fallido de disputar la segunda premisa del argumento— que el fisicismo debe apelar al paralelo con la indexicalidad. La idea de que Vladimir y Perry adquieren nuevo conocimiento —conocimiento de hechos nuevos— es compatible con todo objeto y propiedad involucrada en estas historias, siendo físicas, en el sentido del *problema central de la ciencia física*. Y es compatible con todo objeto y propiedad siendo objetivo, en el siguiente sentido: *el problema central de la ciencia objetiva*. El hecho de que estos conocimientos son solo disponibles desde cierta perspectiva no implica que hay otros objetos y propiedades no-físicos/no-objetivos involucrados en estas situaciones. Lo subjetivo son los hechos.

Ahora bien, muchos han conectado la indexicalidad y el argumento del conocimiento. Pero es importante enfatizar que para apreciarlo, no necesitamos entrar en el debate sobre qué teoría es la correcta sobre los hechos o resolver el problema de cómo se deben individuar las proposiciones.¹⁰⁹ Y no tenemos que hacer la movida implausible de que Mary no aprende nada que sea realmente nuevo. Todo lo que

¹⁰⁸ Lewis argumenta que la física tiene “adecuación explicativa” (Lewis 1990), pero el argumento desde la causación mental al fisicismo solo requiere de la tesis de que la física es causalmente cerrada, no que es explicativamente adecuada: véase Crane (2001, sección 12).

¹⁰⁹ Así no estoy de acuerdo con Van Gulick (1997) de que esta es la línea más fructífera a seguir.

necesitamos es reconocer que existe conocimiento que solo puede tenerse desde ciertos puntos de vista. Conocimiento de hechos subjetivos. Este conocimiento no será conocimiento físico en el sentido del argumento del conocimiento. Pero esto no debe preocupar al físico. Sorprendente como podría parecer, un físico puede (y debería) inteligentemente negar que todo el conocimiento es (en un sentido relevante) conocimiento físico.¹¹⁰ Y ellos entonces deberían negar que todos los hechos son físicos. Esto se explica porque no todo el conocimiento es (en un sentido relevante) objetivo —esto es, conocimiento de hechos objetivos. Por lo tanto, no todos los hechos son hechos objetivos.

Un número de escritores han llamado la atención a que el argumento se mueve desde premisas epistemológicas a una conclusión metafísica.¹¹¹ Mellor dice que “la existencia de hechos subjetivos ha sido falsamente inferida de ciertas clases de conocimiento” (Mellor 1992-3, p. 1). Al considerar este problema de esta forma, estos filósofos han tratado de encontrar algo incorrecto en el argumento. Pero, como he tratado de mostrar, no hay inferencia falsa. Ciertamente, demostrar exactamente lo que el argumento logra debería en sí mismo mostrarnos por qué no deberíamos preocuparnos con él. En tanto que los físicos no sostienen que todo el conocimiento es físico u objetivo, que todos los hechos son físicos u objetivos, o que la física debe ser “explicativamente adecuada” —o que la ciencia objetiva puede establecer todos los hechos— entonces el argumento del conocimiento no objeta al físico. Nos dice, más bien, algo importante acerca de nuestro propio conocimiento, algo que incluso los físicos deben aceptar.

¹¹⁰ En esto estoy de acuerdo con Tye (1995).

¹¹¹ Véase Jackson (1995), Lewis (1990), Levine (1993), Horgan (1984) y muchos otros.

Referencias

- Austin, J. L. (1961): “Unfair to facts”. En: *Philosophical Papers*. Oxford: Clarendon Press.
- Broad (1926): *The Mind and its Place in Nature*. London: Routledge y Kegan Paul.
- Chalmers, D. (1996): *The Conscious Mind*. Oxford: OUP.
- Churchland, P. M. (1995): “Reduction, qualia and the direct of introspection”. *The Journal of Philosophy*, Vol. 82, 1, 8-28.
- _____ (1997): “Knowing qualia: A reply to Jackson”. En: Block *et al.* (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Crane, T. (2000): “The origins of qualia”. En: T. Crane y S. Patterson (eds.) *History of Mind-Body Problem*. London: Routledge.
- _____ (2001): *Elements of Mind*. Oxford: OUP.
- _____ (2003): “Subjective facts”. En: H. Lillehammer y G. Rodríguez-Pereyra (eds.) *Real Metaphysics*. London: Routledge.
- Davidson, D. (1984): “True to the facts”. *Inquiries into Truth and Interpretation*. Oxford: OUP.
- Dennett, D. (1991): *Consciousness Explained*. London: Allen Jane.
- Feigl, H. (1958): *The “mental” and the “physical”*. Minneapolis: University of Minnesota Press.
- Frege, G. (1967): “The thought: A logical enquiry”. En: P. F. Strawson (ed.) *Philosophical Logic*. Oxford: OUP.
- Güzeldere, G. (1997): “Understanding the Phenomenal Mind”. En: N. Block, O. Flanagan y G. Güzeldere (eds.) *The Nature of Consciousness*. Cambridge, Mass.: MIT Press.
- _____ (2002): “Approaching consciousness”. En: Block *et al.* (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Horgan, T. (1984): “Jackson on physical information and

- qualia*". *Philosophical Quarterly* 34, 147-152.
- Jackson, F. (1982): "Epiphenomenal qualia". *Philosophical Quarterly* 32, 127-136.
 - _____ (1995): "Postscript to what Mary didn't know". En: P. Moser y J. D. Trout (eds.) *Contemporary Materialism*. London: Routledge.
 - _____ (1998): *From Metaphysics to Materialism*. Oxford: OUP.
 - _____ (2002): "What Mary Did Not Know". En Block et al (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.
 - Levine, J. (1993): "On leaving out what it's like". En: M. Davies y G. Humphreys (eds.) *Consciousness*. Oxford: OUP.
 - Lewis, D. (1990): "What experience teaches". En: W. Lycan (ed.) *Mind and Cognition*. Oxford: Blackwell.
 - _____ (1999): "Elusive knowledge". *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
 - Lillehammer, H. y Rodriguez-Pereyra, G. (2003): *Real Metaphysics: Essays in Honour of D.H. Mellor*. London: Routledge.
 - Loar, B. (2002): "Phenomenal states". En: Block et al. (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.
 - Loewer, B. (2001): "From physics to Physicalism". En: C. Gillet y B. Loewer (eds.) *Physicalism and its Discontent*. Cambridge: Cambridge University Press.
 - Mellor, D. H. (1992-3): "Nothing like experience". *Proceedings of the Aristotelian Society* 93, 1-16.
 - _____ (1995): *The Facts of Causation*. London: Routledge.
 - Moore, A. W. (1997): *Points of View*. Oxford: OUP.
 - Nemirow, L. (1990): "Physicalism and the subjective quality of experience". : W. Lycan (ed.) *Mind and Cognition*. Oxford: Blackwell.
 - Papineau, D. (2001): "The rise of Physicalism". En: C. Gillet y B. Loewer (eds.) *Physicalism and its Discontent*. Cambridge:

Cambridge University Press.

- Perry, J. (1979): “*The problem of essential indexicals*”. *Nous* 13, 3-21.
- Rey, G. (1997): “*Sensational sentences switched*”. *Philosophical Studies* 68, 3, 289–319.
- Robinson, H. (1982): *Matter and Sense*. Cambridge: Cambridge University Press.
- Russell, B. (1927): *The Analysis of Matter*. London: George Allen y Unwin.
- Snowdon, P. (2003): “*Knowing how and knowing that: a distinction and its uses reconsidered*”. *Proceedings of the Aristotelian Society*, 104, 1, 1-29.
- Tye, M. (1995): *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- Van Gulick, R. (1997): “*Understanding the phenomenal mind*”. En N. Block, et al. (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.

