

UNIVERSIDAD DE CHILE
Facultad de Filosofía y Humanidades
Escuela de Postgrado

EL PROBLEMA EPISTEMOLÓGICO DE LOS JUICIOS DE POSIBILIDAD COMO PROBLEMA METAFILOSÓFICO Y SOBRE EL ENTENDIMIENTO

Tesis para optar al grado de Magister en Filosofía
con mención en Epistemología

Autor: Felipe Morales Carbonell
Profesor Guía: Rodrigo González Fernández

Santiago de Chile
2013

Le monde réel a ses bornes, le monde imaginaire est infini; ne pouvant élargir l'un, rétrécissons l'autre; car c'est de leur seule différence que naissent toutes les peines qui nous rendent vraiment malheureux.

Emile, II

JEAN-JACQUES ROUSSEAU

Agradecimientos

Tengo que agradecer a las personas que han estado junto a mi durante el tiempo que he tratado de realizar este trabajo. A mi madre, sobre todo, quien es la que persona que más ha tenido que soportarme (en más de un sentido) durante el tiempo en que he trabajado en esta investigación; he tratado de ayudarla en la medida de lo posible, pero siempre quedaré, creo, con la impresión de no haber hecho suficiente. A mi ex-pareja, Marisol. A mis amigos Mario, Paula y Soledad, que me dieron apoyo. Al profesor Rodrigo González, quien ha guiado esta investigación.

Esta tesis recibió financiamiento por parte del proyecto Fondecyt 11080020, “Lo Imaginable y lo Posible: el Estudio de lo Mental desde la Experiencia Concebible”.

Índice general

El problema epistemológico de los juicios de posibilidad como problema meta- filosófico y sobre el entendimiento	VI
Resumen	VI
Introducción	1
1 Descartes	7
1.1. Argumentos de separación	7
1.1.1. El caso de K. Argumentos de separación.	7
1.1.2. Algunos aspectos formales sobre los argumentos de separación.	10
1.1.3. El problema fenomenológico de los casos cotidianos.	11
1.2. El argumento cartesiano a favor del dualismo	12
1.2.1. El argumento de la VI <i>Meditación</i>	12
1.2.2. Las dos fases del argumento cartesiano	13
1.2.2.1. La primera fase: el argumento general	13
1.2.2.2. El criterio de verdad cartesiano	14
1.2.2.3. La segunda fase: el argumento aplicado	21
1.3. La epistemología modal cartesiana versus su metafísica modal	23
1.4. Algunas perspectivas sobre la opción del dualismo como ilusión metafísica	26
1.5. Otros problemas de la metafísica cartesiana	27
1.5.1. Existencia, cantidad y la categoría de substancia	28
1.5.2. El soñador cartesiano como sujeto <i>posible</i>	32
2 Kripke	34
2.1. Modalidades y mundos posibles	37
2.1.1. La idea de la teoría de modelos	38
2.1.2. <i>Consideraciones semánticas sobre la lógica modal</i> (1963)	38

2.2.	La tesis de la rigidez de los nombres	42
2.2.1.	El problema del significado de los nombres	42
2.2.2.	El argumento modal contra la teoría descriptivista y la rigidez de los nombres	45
2.3.	La rigidez de los términos de clase natural	48
2.4.	La separabilidad de la mente y el cuerpo, nuevamente	49
2.5.	La actitud kripkeana sobre los juicios de posibilidad	53
3	Entendimiento y juicios de posibilidad	56
3.1.	El aspecto metafilosófico	56
3.1.1.	Distinciones fundamentales.	56
3.1.2.	Una postura metafilosófica	60
3.2.	El entendimiento y el juicio de posibilidad	66
3.2.1.	Caracterizando el entendimiento	66
3.2.2.	Explicación y entendimiento	74
3.2.3.	La constitución del juicio de posibilidad en el entendimiento	89
3.2.3.1.	Un antecedente: El confiabilismo modal de Bealer	90
3.2.3.2.	Unificación y juicios de posibilidad	95
3.3.	Observaciones finales	100
	Bibliografía	103

El problema epistemológico
de los juicios de posibilidad
como problema metafilosófico
y sobre el entendimiento

Resumen

En esta tesis se explora el problema epistemológico de los juicios de posibilidad desde el punto de vista de sus implicaciones para la actividad filosófica. Para mostrar de qué manera los problemas de la epistemología del juicio de posibilidad cumplen un rol en una filosofía, se examinan las obras de Descartes (en especial, las *Meditaciones Metafísicas*) y Kripke (*El Nombrar y la Necesidad*). Se propone que la epistemología del juicio de posibilidad puede atacarse desde el ángulo de la cuestión del entendimiento, y se esboza la hipótesis de que juzgar que algo es posible consiste en una forma de entenderlo. El modelo del entendimiento que se adopta para ello utiliza el concepto de unificación: un juicio de posibilidad correcto está unificado, en un sentido específico, en una estructura que representa una manera en que podría haber sido el mundo. Adicionalmente, se muestra cómo esta propuesta sobre el juicio de posibilidad se enlaza con una cierta comprensión de la filosofía.

Introducción

La intención de quien escribe esta tesis es explorar en ella el problema epistemológico de los juicios de posibilidad filosóficos. En primera instancia, para mostrar que si se ponen en duda los juicios de posibilidad a los que se apela en filosofía, se ponen automáticamente en duda muchos de los juicios filosóficos en general. En segundo lugar, para hacer visibles las complicaciones que presenta la epistemología modal en general.

La manera en que se va a tratar de hacer lo primero no requiere mayor introducción que lo siguiente: si se examina la manera en que un filósofo trata con situaciones posibles o argumentos que se sostienen sobre que situaciones imaginarias o concebidas sean posibles verdaderamente, y se conectan los conceptos que involucran esas posibilidades con los conceptos que normalmente se consideran como los principales *temas* que el filósofo aborda, se ve que importantes tesis filosóficas que mantenga sólo pueden sostenerse sobre la base de juicios de posibilidad. En los capítulos primero y segundo de esta tesis se mostrará que esto es el caso sobre dos filósofos cuya influencia ha sido históricamente considerable (Descartes y Kripke), y se mostrará, se cree, que sus filosofías (entendidas estas como el conjunto de creencias que estos filósofos proponen como verdaderas) dependen de varios compromisos epistemológicos sobre el juicio acerca de posibilidades.

Ahora vale la pena detenerse sobre los problemas *generales* de la epistemología modal; no se volverá a ellos explícitamente hasta el tercer capítulo, y no obstante conviene tenerlos presentes como trasfondo de la discusión previa a aquél.

Considérense los siguientes enunciados:

- 1) El autor de este texto es Felipe Morales,
- 2) Este texto pudo ser escrito por Felipe Morales,
- 3) Felipe Morales pudo no haber escrito este texto,
- 4) Este texto pudo no haber sido escrito,
- 5) Este texto pudo ser escrito por alguien más que Felipe Morales¹,
- 6) Este texto podría haberse titulado “Escepticismo y Juicios de Posibilidad”

¹ Así como en el cuento “Pierre Menard, autor del Quijote”, de Borges, el susodicho Pierre Menard escribe un texto original línea por línea idéntico al *Quijote* de Cervantes.

7) Este texto pudo haber tenido un contenido distinto.

¿Cuál de ellos es correcto? ¿Qué significa que enunciados de como estos sean correctos? Examinense uno por uno:

1) es correcto, porque el autor de este texto sí es Felipe Morales; la corrección de un enunciado como este depende de cómo sean de hecho las cosas.

2) es correcto, porque es intuitivo que si Felipe Morales *es* el autor de este texto, también *pudo* serlo.

3) también parece ser correcto; es un hecho contingente sobre Felipe Morales el haber sido el autor de este texto (quizás Felipe Morales podría no haber aprendido nunca a escribir, quizás podría no haberse interesado nunca por la filosofía, etc.).

4) también parece ser correcto. Al menos, si 3 es correcto, parece posible que si Felipe Morales no hubiese escrito el texto, nadie más lo habría escrito (por ejemplo, se puede imaginar un escenario en que el texto no fue escrito porque Felipe Morales no lo escribió). ¿Es ese escenario contingente o necesario, es decir, podría haber sido de otro modo o no? Si fuera contingente, 5) tendría que ser correcto: decir que es contingente que Felipe Morales no haya escrito el texto (si el texto no fue escrito, se entiende), *es lo mismo* que decir que alguien más pudo haberlo escrito (es decir, 5). Por otro lado, si fuera necesario, 5) tendría que ser incorrecto. ¿Es entonces, contingente o necesario que, de ser el caso que este texto fuera escrito, Felipe Morales lo escribiera? Esta pregunta no es trivial. ¿Cómo se puede saber su respuesta? ¿Cómo se puede saber 5)? ¿Se puede saber 5? Formulando estas preguntas debería tenerse ya una imagen de la clase de problemas epistemológicos que rodean a los juicios de posibilidad en general.

Uno puede encontrar los mismos problemas en juicios como 6) y 7. Incluso si tenemos intuiciones sobre qué es lo correcto en esos casos (por ejemplo, incluso si creemos que el título de este texto podría haber sido otro porque no es necesario que un único título sea aplicable a un texto), cabe preguntarse si estas intuiciones apuntan a las respuestas correctas (o incluso si se puede confiar en que se tendrán intuiciones en todos los casos que puedan aparecer).

No parece que puedan tratarse todos los enunciados que se han examinado de la misma manera. ¿Cómo se explican las diferencias entre unos y otros casos?

Las preguntas centrales en la epistemología de los juicios de posibilidad (cf. van Inwagen (2001), pág. 69) son

- a) ¿Tenemos conocimiento efectivo de posibilidades?
- b) ¿Qué alcance tendría este conocimiento?
- c) ¿Cómo sabemos que una afirmación de posibilidad es verdadera o falsa?
- d) ¿Qué clase de justificación puede darse para enunciados de la forma ‘tal y tal es un estado de cosas posible’?
- e) Si sabemos que enunciados de ese tipo son verdaderos o falsos, ¿de dónde proviene este conocimiento, qué fuente (o fuentes) tiene)?

La respuesta a la primera pregunta parece tener que ser positiva: hay muchos casos en que no es dudoso que se pueda afirmar que se tiene conocimiento sobre si algo es posible o no. No lo es, por ejemplo, que se diga que algo que ha sido hecho en el pasado puede ser hecho nuevamente en el futuro, a menos que haya indicios de lo contrario (e incluso en ese caso, no se duda tanto la posibilidad de que ocurra como el que vaya realmente a suceder). No lo es tampoco cuando las posibilidades pueden inferirse a partir de creencias firmemente establecidas, como las del sentido común; así pasa, por ejemplo, cuando decimos que podríamos mover un mueble a otra parte de la habitación.

Puede decirse que lo mismo sucede con las predicciones e hipótesis teóricas. Por ejemplo, Hill (2012) propone extensiones a la teoría de la relatividad especial de Einstein que permiten velocidades mayores a la de la luz (efectivamente argumentando que tales velocidades son posibles). Pero aquí aparece un problema conceptual: aunque se acepte la posibilidad *lógica* de velocidades mayores a la de la luz, aún cabe preguntarse si esta posibilidad es concreta (si en el mundo real podría pasar que un objeto viaje a una velocidad mayor a la de la luz). Ya que no se tiene evidencia de que eso ocurre, ¿cómo se puede saber si es efectivamente posible? ¿Basta el puro argumento *a priori*? La discusión sobre la posibilidad de desplazamiento de objetos y comunicación de información más rápidos que la velocidad de la luz es un claro indicio de que el problema no es trivial.

Así, pues, ¿qué alcance tiene el conocimiento de posibilidades? La respuesta a esta pregunta depende de cómo se conciben las fuentes y la justificación de los juicios sobre posibilidades. Quizás las fuentes del conocimiento modal no son confiables (Nozick (2001), por ejemplo, propone que nuestras capacidades para el conocimiento modal están sub-desarrolladas si se las compara con las que tenemos para obtener conocimiento de nuestro entorno). O quizás haya que decir que la justificación para un juicio de posibilidad tiene que satisfacer un estándar menos estricto que la que se requiere para juicios perceptuales (así como, por ejemplo, parece requerirse menos esfuerzo para proponer una hipótesis que para verificarla), y que en consecuencia el alcance del conocimiento modal es más extenso que el del conocimiento empírico. Quizás el conocimiento de posibilidades depende del conocimiento que se tenga de cómo sean cosas de naturaleza similar a aquellas de las que ya tenemos conocimiento (por ejemplo, que sólo se conciben posibilidades aceptables mediante combinación de representaciones del mundo que conocemos), o quizás esta restricción no es necesaria. Las opciones teóricas son abundantes; sin embargo no todas son igualmente plausibles.

La teoría clásica sobre el conocimiento modal es la teoría de la concebibilidad. Esencialmente, ella se asienta sobre la idea de que concebir algo es al menos un indicio de que lo que se concibe es posible. En este tipo de teoría, la concebibilidad se considera como una fuente fundamental de conocimiento modal. Las variaciones de esta teoría especifican qué es lo que se entiende por concebir (por ejemplo, podría construirse el concepto como una forma de estado mental en que se tienen representaciones figurativas, o bien proposicionales, o bien de ambos tipos) y cuál es el grado de evidencia sobre la posibilidad que se obtiene mediante este mecanismo (concebir algo podría

ser un mero indicio de la posibilidad, o bien podría ser equivalente a que ella es efectiva, mediante una forma de implicación). Las teorías que adoptan a la concebibilidad como fuente primaria del conocimiento modal sostienen todas que la concebibilidad no es un mero indicio de la posibilidad, pues si así fuera no podría contar como justificación suficiente o “final” de los juicios de posibilidad. Por otro lado, la opinión de que la concebibilidad de algo entalla su posibilidad sin restricciones ha sido casi unánimamente rechazada, aunque es probable que haya personas que crean que la concebibilidad sin restricciones basta para justificar sus juicios modales.

Por ejemplo, Yablo (1993) acepta que la concebibilidad justifica la creencia en la posibilidad, pero define “concebible” de esta manera:

- i) P es concebible para un sujeto S si S puede imaginar un mundo tal que S considere que verifique P .

De acuerdo a esta condición, para, por ejemplo, afirmar que es posible que exista una vaca azul es necesario imaginar un mundo en que se verifique que hay vacas azules. No es claro que se pueda imaginar un mundo como ese, porque no es claro qué mecanismos podrían hacer que una vaca tuviese un color azulado manteniendo la condición de que sea una vaca (¿es necesario que las vacas *no* sean azules?). Asimismo, Menzies (1998) propone una restricción al alcance de la noción de concebibilidad cuando propone que

- ii) P es posible si y sólo si un sujeto ideal puede concebir que P

pues la noción de sujeto ideal implica restricciones sobre las condiciones que debe satisfacer un sujeto que considere P . Chalmers (2002) sugiere que para que la concebibilidad implique posibilidad debe 1) consistir en construir un escenario que se verifique el enunciado que se considera (por ejemplo, un escenario en que haya vacas azules cuando se considera un enunciado sobre vacas azules), y no meramente en ser incapaces de descartarlo (en la terminología de Chalmers, la concebibilidad debe ser *positiva*), 2) no poder ser derrotada por otros razonamientos o mecanismos epistémicos (así como puede serlo un razonamiento producto de una ilusión; Chalmers dice en este caso que la concebibilidad debe ser *ideal*, en oposición a *prima facie*).²

Existen otros enfoques sobre la epistemología del juicio de posibilidad. Por ejemplo, en vez de tomar la concebibilidad como fuente primaria del conocimiento de posibilidades, puede tomarse el

²Chalmers además requiere que la concebibilidad remita a lo que llama la “intensión primaria” de los enunciados en cuestión. Se trata en este caso de si el escenario “posible” que se considera sería verdadero o falso si el mundo real fuera el mundo del escenario, en cuyo caso se trataría de la intención primaria del enunciado, o si sería verdadero o falso en un mundo contrafáctico dadas las condiciones del mundo real, en cuyo caso se trataría de la intención secundaria. Chalmers propone que la concebibilidad sólo puede implicar una posibilidad primaria. Por ejemplo, si se concibe un mundo en que hay un elemento que tiene todas las propiedades del agua en el mundo real, pero que tiene una composición química distinta, resulta que si se considera la intención primaria el escenario es posible (el agua en ese mundo sería XYZ, por ejemplo), pero imposible si se considera la secundaria (porque en el mundo real la composición del agua es H_2O).

entendimiento como mecanismo primario. La idea esencial de este punto de vista es que la capacidad de hacer juicios modales depende de la comprensión de parte de un sujeto de los conceptos que tiene y de las relaciones que hay entre ellos. Por ejemplo, se podría justificar el juicio de que no es posible cuadrar el círculo porque si se entienden los conceptos de círculo y de cuadratura, la noción de la cuadratura del círculo no es admisible. Para dar un análisis acabado de esta estrategia es necesario especificar qué tipo de comprensión se requiere para que se satisfaga esta condición. Dos filósofos que han desarrollado una teoría de este tipo son Bealer (2002) y Peacocke (1998). El primero enfatiza la relación entre comprender un concepto y tener intuiciones sobre los casos en que el concepto se aplica (así, que el concepto de círculo no se puede aplicar en el caso que se tomaba como ejemplo), mientras que el segundo piensa que hay conocimiento tácito fundamental sobre ciertos principios que gobiernan la noción de modalidad, de los que los sujetos derivan su aplicación de los conceptos modales. Se volverá a este tipo de teorías en el tercer capítulo, donde se esbozará una postura sobre el juicio de posibilidad que ocupa como concepto central el de entendimiento por unificación (la idea de que entender algo implica ponerlo en un marco conceptual que lo cubra). Los límites al conocimiento modal que implica una postura como esta radican en los límites de la comprensión de los conceptos que se ven involucrados en cada caso; pero por lo mismo implican que los juicios de posibilidad tienen un aspecto fundamentalmente epistémico.

Una tercera alternativa sobre la epistemología del juicio de posibilidad es reducirla a la epistemología del juicio contrafáctico. Esta es la estrategia de, por ejemplo, Williamson (2007). El argumento es que si se pueden reducir los conceptos modales como el de necesidad lógica a una formulación contrafáctica, se puede reducir la epistemología del juicio modal al del juicio contrafáctico. Según esta propuesta, uno estaría justificado para creer que algo es posible si, asumiendo que es el caso y explorando qué es lo que se sigue lógicamente de ello, no se llega a una contradicción; los juicios de posibilidad justificados deberían poder reducirse a razonamientos de este tipo. Del mismo modo, establecer la necesidad de algo sería simplemente el resultado de verse llevado a una contradicción al asumir que no sea el caso.

Como se ha mencionado, en este trabajo se va a esbozar una teoría del segundo tipo. La decisión de examinar esta perspectiva deriva de la creencia de que esta solución al problema epistemológico calza de manera más precisa con una postura correcta sobre la naturaleza de la actividad filosófica. En primer lugar, porque el así llamado “conocimiento filosófico” debe concebirse mejor como una forma de entendimiento que como una forma de conocimiento sobre el mundo. Si bien una teoría filosófica puede proponerse como una teoría sobre la realidad, es claro que la mayoría de las teorías filosóficas son materialmente falsas. Sin embargo, su función no se limita a presentar una visión correcta sobre cómo son las cosas; en muchos casos las teorías y proposiciones filosóficas son valiosas porque muestran que las cosas no pueden ser como se pensaba que eran, y por lo tanto no entregan conocimiento positivo sino entendimiento. En segundo lugar, porque más allá de este primer punto, permite adoptar una postura escéptica crítica sobre el valor de verdad de varias hipótesis filosófi-

cas, en particular aquellas que se sostienen directamente en juicios de posibilidad. Es plausible que algunas propuestas filosóficas sean simplemente incomprensibles, a pesar de los juicios *prima facie* que puedan hacerse sobre su comprensibilidad, y que esto signifique que las posibilidades que son presentadas en ellas no puedan ser afirmables. Así como Wittgenstein (1958) advierte que

When philosophers use a word—“knowledge”, “being”, “object”, “I”, “proposition”, “name”—and try to grasp the essence of the thing, one must always ask oneself: is the word ever actually used in this way in the language-game which is its original home?

116

es necesario que se examine si los juicios filosóficos de posibilidad son aceptables y no solamente producto de una epistemología errónea. van Inwagen (2001) levanta precisamente esa sospecha (aunque quizás un poco bruscamente):

Many philosophers have a far more sanguine view of [la capacidad de hacer juicios modales] than I. They are very confident that they can make philosophically interesting modal judgments about concerns remote from everyday life, and they are impatient with anyone who challenges their claim to have this ability. Richard Gale is an excellent example of such a philosopher. . . He has said that if my own capacity to make modal judgments is really as limited as I say it is, I must be regarded as modally challenged, as someone who “cannot modalize like normal people do”. But I *can* “modalize” like normal people, and so can he. What he *can't* do, apparently, is to discriminate those cases in which his modal judgments are products of his ordinary human powers of “modalization” from those that are based on his immersion in a certain philosophical environment—an environment composed of philosophers who unthinkingly make all sorts of fanciful modal judgments because *they've* always been surrounded by philosophers who unthinkingly make the same sorts of fanciful modal judgments. . . He is unaware that the modal beliefs he expresses or presupposes when he says “We’d have more room if we’d moved the table up against the wall” . . . and the modal beliefs he gives such confident expression to in his writings on philosophical theology have quite different sources. The former have their source in our ordinary human powers of “modalization” . . . ; the latter have their source in his professional socialization, in “what his peers will let him get away with saying”. (págs. 72–73)

Descartes

En el caso de Descartes se puede encontrar uno de los ejemplos más claros del uso filosófico de tesis modales. Su conocido argumento a favor del dualismo tiene como núcleo el juicio de que es posible que la mente exista separadamente del cuerpo, de lo que se sigue que son sustancialmente distintos. Cabe preguntarse si este juicio cartesiano está justificado; cómo se lo puede justificar dentro del marco conceptual de la epistemología cartesiana, y si Descartes, o al menos una interpretación de sus razonamientos, puede entregarnos nociones que permitan clarificar otras instancias de argumentos similares, o en general, los problemas de la epistemología modal.

Este capítulo está dedicado a estas interrogantes. No interesa tomar a Descartes aquí mas que como ejemplo, y en efecto, al argumento sobre el dualismo como prototipo de cierta clase de razonamiento. Para ponerlo en perspectiva, se examinará primero un ejemplo de naturaleza extrafilosófica, no teórica. Como se verá., incluso en estos casos se encuentran dificultades, pero se tiene de manera más clara la intuición de que son correctos, y que los juicios modales son exitosos, que las afirmaciones que se hacen sobre ellos son acertados. Sobre los argumentos teóricos de este tipo existe mucha menos seguridad, y en especial, respecto a los filosóficos.¹ ¿Por qué se da esta diferencia?

1.1. ARGUMENTOS DE SEPARACIÓN

1.1.1. EL CASO DE K. ARGUMENTOS DE SEPARACIÓN.

Considérese el siguiente caso: durante el viaje a su trabajo, el señor K recuerda que dejó su lápiz favorito en casa, y aunque le fastidia un poco, piensa “bueno, puedo escribir sin él por hoy, hay otros lápices en la oficina”. En efecto, esa tarde escribe sin problemas utilizando otro lápiz.

Pienso que sobre este escenario habría que decir que K juzgó correctamente. ¿Qué fue, sin embargo, lo que juzgó, y cómo razonó? Lo primero puede extraerse directamente del relato: que hoy podía trabajar sin su lápiz favorito. Lo segundo requiere de cierta reconstrucción, llenando ciertos

¹Esto es, hay gente que está convencida de que los argumentos son válidos y gente que está convencida de que no lo son. ¿Qué convence a unos y otros? ¿Por qué otros sin estar convencidos de ninguna de estas opciones?

espacios indeterminados del relato. Una reconstrucción plausible del relato podría ser: K recordó que dejó su lápiz en casa; esto lo molestó, porque tenía la intención de usarlo, y dado al cambio de circunstancias tendría que actuar de manera distinta a como deseaba; se preguntó qué podría hacer; tal vez pensó por un momento en volver por el lápiz, lo que presentaba otros inconvenientes y fue descartado; pensó que lo que tenía que hacer podía hacerlo con un lápiz sin su lápiz favorito, pues cualquier lápiz podía satisfacer su función; recordó que había otros lápices que podía usar en la oficina (puede que haya imaginado o incluso recordado usarlos en ese momento); finalmente decidió ir a la oficina, y usar los lápices que tenía allá.

Es posible incrementar el nivel de abstracción de la reconstrucción, considerando el razonamiento de K al nivel de *conceptos*. Podemos suponer que K tenía dos conceptos, el de una tarea (“lo que tenía que hacer”) y el de un plan (“usar su lápiz favorito”). Inicialmente los creía unidos: el plan satisfaría la tarea. Sin embargo, pudo concebir un caso en que la tarea quedaría satisfecha sin ejecutar el plan (ejecutando otro distinto), de modo que llegó a creer que no había necesidad en el vínculo entre la tarea y el plan, logrando separar los conceptos.

Si esta reconstrucción es correcta, el razonamiento de K habría seguido el siguiente patrón:

- 1) *A* y *B* están unidos *de facto* o *prima facie*.
- 2) Es posible un caso en que *A* y *B* no estén unidos.
- 3) *A* y *B* son separables.

Llamaremos a un argumento que sigue este esquema un *argumento de separabilidad* o de *separación*. Como puede entenderse, el concepto de separabilidad puede variar según el contexto (aquí, por ejemplo, que *A* sea “separable” de *B* significa que *A* puede llevarse a cabo mediante algo distinto a *B*), pero en general debe satisfacer la condición lógica de que no es necesario que si *A*, entonces *B*. Ahora bien, esto se deriva directamente² de que sea posible que *A* y no *B*, que queda implicado en 2). 1) no cumple entonces ningún rol *lógico* dentro del argumento.³ Sin embargo, vale la pena enunciarlo porque sólo en el caso en que se cumple es significativo evaluar el resto de las premisas. Si *A* y *B* estuvieran *de facto* separados, podría derivarse inmediatamente que son separables. La cuestión de si *A* y *B* son separables no surge a menudo tampoco en los casos en que *A* y *B* están *prima facie* separados, aunque en estos casos puede resultar que *A* y *B* estén unidos, contra las apariencias. En este caso la pregunta, desde un punto de vista epistémico, es si son escrutables casos en que *A* y *B* no estén unidos, con la dificultad de que no se puede descartar a priori que, incluso si no hay casos escrutables, podría ser que si los haya inescrutables. Para probar que *A* y *B* están unidos se necesitan otros recursos (las cuestiones que puedan aparecer sobre este punto no son pertinentes aquí). No puede dejar de mencionarse que el juicio de que *A* y *B* estén separados también puede

²Esto es, bajo las asunciones habituales sobre las modalidades. Pueden darse conceptos de lo posible y lo necesario bajo los cuales esto no es válido.

³Nótese que aquí queda implícito un argumento de separabilidad.

presentar dificultades; quizás bajo ciertas condiciones habría que decir que los juicios de este tipo son determinados meramente *prima facie*, es decir, admitir la posibilidad de que las cosas no sean así, a pesar de las apariencias. Esta posibilidad de duda puede ser explotada a la manera escéptica; pero si se aplica esta estrategia, *todo* argumento colapsa.

El peso del argumento recae, entonces, enteramente sobre 2). Dado 1) como presuposición, 2) tiene que soportarse sobre un juicio contrafáctico, a saber, que hay una situación meramente posible en que A se satisface y B no (o viceversa). El problema epistemológico que esto implica se presenta en varios grados de dificultad. En la situación de K, creo, es patentemente plausible que él pueda hacer lo que tiene que hacer usando lápices que no sean su favorito. Otros casos son menos plausibles. Por ejemplo, Kripke (1980) propone el siguiente escenario:

Consider a counterfactual situation in which, let us say, fool's gold or iron pyrites was actually found in various mountains in the United States, or in areas of South Africa and the Soviet Union. Suppose that all the ores which actually contain gold now, contained pyrites instead, or some other substance which counterfeited the superficial properties of gold but lacked its atomic structure.

¿Es plausible este escenario? ¿Es este efectivamente una situación posible? Ciertamente, uno puede determinar que en ese escenario tal o cual cosa podría ser el caso. Así, por ejemplo, Kripke se pregunta si diríamos que en esta situación diríamos que el oro no habría sido un elemento (porque la pirita no es un elemento), a lo que responde que no, creo que correctamente. Para sus propósitos no es relevante determinar si este escenario es posible o no. No se nos dice cómo es posible este mundo, por lo que no deberíamos decir si es posible o imposible. La descripción que se nos da de este mundo no permite, por lo demás, identificarlo únicamente, de modo que tenemos que admitir que nos enfrentamos en realidad a una familia de mundos posibles, entendiéndolos como conjuntos máximamente coherentes de proposiciones (la manera en que Kripke parece entenderlos). Que se refiera a “este mundo posible” lleva a una confusión sobre de qué en realidad se habla. Dicho esto, creo que si se consideraran descripciones más completas de estos mundos resultaría que en algunos casos estos supuestos mundos no son realmente posibles. Por ejemplo, considérese la cuestión de los cambios geológicos que implicaría que no hubiese oro en cierta región, sino pirita. Esto pone en tensión la primera parte del escenario con la segunda: podría ser que no pudiera decirse que existan las montañas que en este mundo contienen oro en “ese mundo”. A menos que se de una razón para descartar estas posibilidades, hay que decir entonces que no sabemos si ese escenario propuesto es posible. Por otro lado, puede verse que es necesario distinguir entre los conceptos epistémicos y metafísicos de posibilidad. Las distinciones necesarias para atacar estos problemas serán tratadas en el próximo capítulo, cuando se examinen las tesis kripkeanas sobre los juicios modales. Lo dicho es suficiente para apreciar las ramificaciones de las dificultades que pueden aparecer en la evaluación de tesis del tipo de 2), de los contrafácticos en general, y luego, de los juicios modales en general.

¿Qué es lo que hace la diferencia entre unos casos y otros?

1.1.2. ALGUNOS ASPECTOS FORMALES SOBRE LOS ARGUMENTOS DE SEPARACIÓN.

Debe ser patente ya que los argumentos de separación no se limitan a seguir el patrón de 1), 2), y 3). Por lo general, antes de 2), y como antecedente para este se coloca una tesis epistémica que sirve como base evidencial. En el mismo caso de K, habría que introducir una tesis que caiga bajo el esquema:

2_0 K imagina un caso en que A y B no estén unidos.

donde K representa a un sujeto cualquiera, y A y B son dos ítemes, ya sea de la misma o distinta clase (por ejemplo, un par de conceptos distintos, como en “José imagina que la justicia no implica equitatividad”), o un concepto y un individuo, como en “José imagina a Danae siendo injusta”,⁴ etc.)

De forma más general, se va a tener una tesis del tipo:

$2_{0'}$ Un sujeto S tiene una actitud de tipo A respecto a la representación R, en la que α y β no están unidos

(aquí α y β cumplen la función de A y B en el esquema anterior), o simplemente

$2_{0''}$ Un sujeto S se representa a modo de A que α y β no están unidos.

A en estos esquemas puede ser reemplazado por conceptos como imaginar, concebir, sus derivados, etc. En cualquier caso se pretende que los juicios de cierto tipo sirven de base evidencial para los juicios del tipo 2, en que se afirma la posibilidad de las situaciones representadas. Este es por lo general el punto más contencioso de los argumentos de separación, desde el punto de vista epistemológico. Se entiende que el problema aplica asimismo a todos los juicios de posibilidad que se funden en que las situaciones putativamente posibles sean o puedan ser representadas de cierta manera.

En algunos casos, puede querer utilizarse juicios del tipo de

$2_{0'''}$ Es posible que un sujeto S se represente a modo de A que α y β no están unidos

como base evidencial de los juicios de separación, lo que no hace sino introducir un nuevo nivel al problema. En muchos casos puede pasar que un juicio del tipo $2_{0'''}$ quiera ser empleado como un juicio del tipo $2_{0'}$ o $2_{0''}$. Por ejemplo, aquí mismo decimos que en el caso de K es posible decir algo que cae bajo el esquema $2_{0'}$.⁵

⁴Aquí se omiten algunos detalles sobre el análisis de estos enunciados, que no son cruciales; por ejemplo, en qué sentido el individuo “Danae” no está “unido” al concepto “justicia”, que es lo que se implica en el enunciado). Los juicios que se examinarán en lo que sigue no presentan estas dificultades.

⁵Habría que ver de qué manera hay aquí un juicio *problemático*, en el sentido de Kant.

1.1.3. EL PROBLEMA FENOMENOLÓGICO DE LOS CASOS COTIDIANOS.

Lo dicho basta de momento sobre los aspectos formales de los argumentos de separación. Hay que hacer algunas observaciones adicionales sobre el caso de K, que se utilizó para introducir el tema de los argumento de separación.

Primero, es importante responder a una posible objeción: que desde la forma original del caso sólo se llega al argumento de separación mediante cierta abstracción que puede resultar artificiosa. Esta abstracción se fundamenta en una serie de presuposiciones que no son triviales; en particular, en la idea de que el razonamiento opera sobre conceptos, y que en el juicio son conceptos lo que se pone en relación. En la reconstrucción del razonamiento de K se identificaron primero los conceptos operativos, y luego se describieron las operaciones que se efectuaron sobre los conceptos; en un segundo nivel, se aislaron las relaciones puramente lógicas de los aspectos psicológicos. La suposición que se hace es que a algún nivel se llevan a cabo estas operaciones lógicas sobre estos juicios. Si fuera así, habría que pensar que el juicio en general no tiene como aspecto esencial un elemento consciente. Ciertamente, es claro que alguien en la posición de K no necesita estar consciente de que juzga que es posible un caso en que la tarea y el plan estén separados. Aquí hay una diferencia importante entre casos como el de K y casos en que se enuncia explícitamente un argumento de separación. Consideraré irrelevante esta diferencia para la evaluación de la *corrección* del razonamiento.

Ahora bien, podría proponerse la hipótesis de que en el contexto de los razonamientos cotidianos, la separación de conceptos se realiza de manera indirecta en vez de explícita. Llevando esto a un extremo podría decirse que *los juicios de posibilidad no son cotidianos*. De cualquier modo el razonamiento modal pertenecería al andamiaje que constituye nuestras formas de pensar.

Otro caso en que se presenta cierta separación entre el juicio de posibilidad y el proceso epistémico que se realice efectivamente es aquél en que un argumento de separación (o más bien, la separación conceptual) es inducido apelando a la mera consideración de casos contrafácticos (así, por ejemplo, cuando se enseña a alguien que lo que le frustra no depende de ella). En estos casos nos encontramos ya con las formas dialécticas de las que ciertas formas de argumentar filosóficas, como los experimentos mentales, son una tipo particular.

En definitiva, el paso de la forma original del caso de K al argumento de separación es coherente con algunas asunciones que son razonables, y es al menos tan justificado como ellas. Solo el que la reconstrucción dada valga como una descripción correcta del razonamiento que se llevó acabo *de facto* es una cuestión que va más allá de lo que podemos afirmar.

Otras preguntas: ¿ha de decirse que cuando *entendemos* el escenario de K y vemos que su razonamiento es correcto, nosotros mismos llevamos a cabo el razonamiento propuesto? Si no, ¿cómo evaluamos la corrección del razonamiento? Y si lo hacemos, ¿no valdría esto como evidencia (nuevamente) de que el juicio no tiene como elemento esencial un aspecto fenomenal (es decir, que

no es esencial que seamos conscientes de lo que juzgamos)? Tengo que admitir que al comprender el escenario, no estoy consciente de que de hecho juzgue según mi reconstrucción del razonamiento de K. Quizás la explicación de esto sea que los actos que realicé mientras proponía el escenario, y los que realizo ahora no sean equivalentes, y que en el segundo caso me aprovecho de la memoria para evaluarlo. Contra está hipótesis, está el hecho de que en principio podría llevar a cabo los actos de cuando diseñé el escenario.

Como puede inferirse de lo dicho hasta aquí, la solución de estas dificultades no implica que se solucione la cuestión de qué justifica el razonamiento de K, y los razonamientos similares. En la sección anterior se ha indicado en qué punto se encuentran las mayores dificultades.

1.2. EL ARGUMENTO CARTESIANO A FAVOR DEL DUALISMO

1.2.1. EL ARGUMENTO DE LA VI *Meditación*

El argumento de separación *filosófico* más notorio y prototípico es, probablemente, el argumento a favor del dualismo de sustancias que Descartes desarrolla en la VI *Meditación Metafísica* (Descartes (2007)):

Primeramente, puesto que ya sé que todas las cosas que concibo clara y distintamente pueden ser producidas por Dios tales como las concibo, bastará que pueda concebir clara y distintamente una cosa sin otra, para que esté cierto de que la una es distinta o diferente de la otra, ya que pueden estar separadas, al menos por la omnipotencia de Dios; y no importa cuál sea la potencia que verifique esta separación, para que sea forzoso el juzgarlas diferentes. Por lo tanto, puesto que sé de cierto que existo, y sin embargo, no advierto que a mi naturaleza o a mi esencia le convenga necesariamente otra cosa, sino que yo soy algo que piensa, o en ser una sustancia cuya esencia o naturaleza toda es sólo pensar. Y aún cuando, acaso, o más ciertamente, como luego diré, tengo yo un cuerpo al que estoy estrechamente unido, sin embargo, puesto que por una parte tengo una idea clara y distinta de mí mismo, según la cual soy algo que piensa y no extenso, y por otra parte, tengo una idea distinta del cuerpo, según la cual éste es una cosa extensa, que no piensa, resulta cierto que yo, es decir, mi alma, por la cual soy lo que soy, es entera y verdaderamente distinta de mi cuerpo, pudiendo ser y existir sin el cuerpo. (pag. 178-179)

En las siguientes secciones de este capítulo se tratará de la reconstrucción formal del argumento de Descartes, qué hace que el argumento funcione dentro del contexto de la filosofía cartesiana (para que se aprecie cómo pueden ser tratadas las cuestiones que se han delineado sobre los juicios de posibilidad), qué función tiene dentro de ella (para que se aprecie la significatividad de la cuestión de la epistemología modal para una filosofía), y se criticará la epistemología modal cartesiana, y

(TODO) será demostrado que es insuficiente/para mostrar que es insuficiente. Por otro lado, durante el transcurso de estas se discutirán varios aspectos problemáticos de la filosofía cartesiana.

1.2.2. LAS DOS FASES DEL ARGUMENTO CARTESIANO

El argumento de la VI *Meditación* transcurre en dos fases: en la primera, se establece el principio que conecta la concebibilidad y la posibilidad, y en la segunda, se aplica el principio para separar las ideas de *res cogitans* y *res extensa*.

1.2.2.1. La primera fase: el argumento general

¶ El argumento de la primera fase es hecho desde un punto de vista epistémico. Descartes argumenta que *dado que sabe que*

C₁ todo lo que se concibe clara y distintamente ₂ puede ser creado por Dios tal y como se lo concibe.
se sigue que

b₁ si concibo clara y distintamente una cosa sin otra, ₂ puedo tener certeza de que son distintas o diferentes

Es decir, lo que Descartes propone es que el que sepa *C* le habilita para tener certeza sobre el juicio de que *b*. Si no lo supiera, no podría tener certeza en ese juicio. En efecto, como se discutirá luego, *C* es un elemento habilitador crucial en la epistemología racionalista de Descartes; sin este principio no podría utilizar premisas basadas en concepciones para determinar la naturaleza de cosas existentes. Lo fundamental es que según *C* lo que alguien concibe clara y distintamente puede existir en el orden natural. Esto manifiesta el alcance del criterio cartesiano de que lo que se piensa clara y distintamente es verdadero más allá de las meras actualidades en el campo modal de lo posible y lo necesario.

El paso de *C* a *b* no es directo en un sentido formal. Si concibo clara y distintamente una cosa sin otra (*b*₁), se sigue de *C* sólo que

b₂₀ Dios puede hacer que esas cosas existan separadamente.

de lo que se deriva (bajo la asunción de que lo que Dios puede crear es posible)⁶ que

b_{20'} es posible que esas cosas existan separadamente.

*b*₂ se sigue de *b*_{20'} empleando el principio de conexión entre la separabilidad y la diferencia (CSD)

CSD si es posible que *a* y *b* existan separadamente, es necesario que $a \neq b$

⁶Descartes no pone en cuestión esta asunción. Sin embargo, parece problemática porque Descartes cree además que Dios puede hacer cualquier cosa, independientemente de si podemos concebirla o no, lo que lleva a hacer una doble distinción entre lo posible *para Dios* y lo posible *para nosotros*. Se volverá a este punto luego.

Ahora bien, desde un punto de vista meramente intuitivo, el paso de b_{2_0} a b_2 es “obvio” (lo que quiere aquí decir tan sólo que CSD es una parte tan esencial de las maneras en que entendemos y juzgamos cotidianamente que cosas pueden ser diferentes que no somos conscientes necesariamente de su aplicación), lo que explica que Descartes no haya explicitado este punto de la conexión entre C y b . ¿Qué justifica CSD, sin embargo, más allá de su cualidad de “ser de sentido común”? ¿Podría CSD no ser verdadera? ¿Podemos pensar en algún caso en que CSD sea falsa, en el sentido de que incluso si dos cosas (a y b) existen separadamente, $a = b$? Como veremos con detalle en el próximo capítulo (cuando examinemos el problema de la identidad desde la perspectiva de Kripke), no; esto se sigue directamente de la conjunción del principio de necesidad de la diferencia ($a \neq b \rightarrow \Box(a \neq b)$) y la tesis, razonable, de que si algo puede existir separadamente de otra cosa, entonces son cosas diferentes (que a su vez es una aplicación del principio de discernibilidad, $Fa \wedge \neg Fb \rightarrow a \neq b$). Finalmente, para llegar a b_2 basta aceptar que se puede tener certeza sobre lo que es necesariamente verdadero.

El paso de C a b es, así pues, válido bajo ciertas asunciones adicionales que parecen razonables. Si logra mostrar que b es verdadero, Descartes garantiza para sí el empleo de argumentos de separabilidad basados en tesis sobre la concebibilidad de escenarios contrafácticos en la segunda fase del argumento sobre el dualismo. Sin embargo, para hacer esto necesita mostrar que C es verdadero. ¿De qué manera defiende esta tesis?

1.2.2.2. *El criterio de verdad cartesiano*

El presupuesto C dice algo sobre las capacidades de Dios: que cualquier cosa que se conciba clara y distintamente está en ellas. En la práctica, esto es equivalente a decir que cualquier cosa que se conciba clara y distintamente puede existir (es posible, en cierto sentido). Desde este punto es posible avanzar en varias direcciones para entender el significado sistemático de C en la epistemología cartesiana.

Si nos preocupamos de la cuestión de la justificación de C (como se ha expresado) podemos dirigirnos a la manera en que Descartes trata de forma general el problema de la justificación de los juicios.

En las *Meditaciones* lo que Descartes hace es, fundamentalmente, el análisis de una serie de clases juicios que tienen cierta importancia epistemológica (representan sendas clases de conocimiento sobre el mundo, los sujetos que viven e él, etc.). A primera vista, la perspectiva cartesiana inicial es la de quien pregunta ‘¿de cuáles de mis juicios puedo estar seguro?’, pero en realidad lo que le interesa es responder la pregunta ‘¿de qué puedo estar seguro?’. En las *Meditaciones*, pasa de la primera cuestión a la segunda. Así es como procede primero a atacar la certeza aparente de varias clases de juicios, presentando escenarios posibles que socavan su seguridad, hasta llegar a un punto en que ninguna de las clases de juicios previamente aceptadas se mantiene en pie. Sólo una vez que ha desmontado la certeza primitiva sobre esas clases de juicios procede a resolver positivamente la

pregunta, y entonces sólo desde una postura diferente; sólo si no se tiene certeza sobre nada (o al menos, si no se tiene certeza sobre las cosas sobre las que uno creía tener certeza) se puede plantear con completa convicción y libertad la segunda cuestión. Descartes no busca necesariamente entre sus juicios originales aquellos de los que puede tener certeza, sino que se dedica a construir nuevos juicios siguiendo el principio de la veracidad de la intuición clara y distinta. Sólo una vez que ha establecido este principio puede decir que ha llegado a una base epistemológica sólida (claro está, en sus propios términos; uno puede permanecer neutral respecto a la cuestión de si la manera en que Descartes fundamenta su epistemología es exitosa).

La primera verdad a la que llega Descartes en el transcurso de las *Meditaciones* es la del *cogito*, y ella le sirve como modelo de juicio sobre el que puede tener certeza, o afirmar que sabe. Descartes considera que la diferencia entre ella y las creencias que tuvo que abandonar es que la percepción de ella es clara y distinta, sin que le sea posible proponer un caso en que no sea verdadera (con lo que se manifiesta su “necesidad”). Esto le permite establecer como regla, dice, el principio:

CD: Todas las cosas que se conciben muy clara y distintamente son verdaderas

En el *Discurso*, Descartes resume su razonamiento así:

... habiendo notado que en la proposición “yo pienso, luego soy” no hay nada que me asegure que digo verdad, sino que veo muy claramente que para pensar es preciso ser, juzgué que podía admitir esta regla general: que las cosas que concebimos muy clara y distintamente son todas verdaderas, y que sólo hay alguna dificultad en notar cuáles son las que concebimos claramente. (p. 66)

En las *Meditaciones*, sin embargo, da cuenta de una dificultad:

Estoy seguro de que soy una cosas que piensa; pero ¿no sé también cuáles son los requisitos precisos para estar seguro de algo? Desde luego, en este mi primer conocimiento no hay nada que me asegure su verdad, si no es la percepción clara y distinta de lo que digo, *lo cual no sería, por cierto, suficiente para asegurar que lo que digo es verdad, si pudiese ocurrir alguna vez que fuese falsa alguna cosa concebida así por mí de ese modo claro y distinto*; por lo cual me parece que ya puedo establecer esta regla general... (III, pp. 137–138)

A causa de esta dificultad resulta que en este punto de las *Meditaciones* CD no puede establecerse aún, al contrario de lo que aparentemente afirma Descartes; al menos no como una regla de alcance universal, porque justamente la dificultad muestra que su alcance no está realmente demostrado. En realidad, es debido a este problema que tiene que avanzar los argumentos que siguen, primero sobre la existencia de Dios, luego sobre su bondad, y luego sobre su rol como garante de nuestra facultad de concebir. En efecto, ya en el “Resumen” Descartes dice que no se ha probado que todo lo que concebimos clara y distintamente sea verdadero sino hasta la cuarta meditación (cuando

discute por qué en el texto principal no se trata explícitamente el problema de la inmortalidad del alma). Si no se ha demostrado la validez de CD como regla general, ¿qué validez pueden tener los juicios que Descartes considere claros y distintos, incluido el *cogito*? Esto conduce al problema del así llamado “círculo cartesiano”, que no es más que una forma especial del problema del criterio (¿qué criterio puede justificar nuestros criterios?).⁷

CD es efectivamente el criterio de verdad cartesiano. Si CD no es verdadero, y en el punto en que se propone aún no es cierto que lo sea, a causa de la dificultad observada, Descartes no tiene manera alguna de proceder y argumentar a favor de la verdad de CD mismo mediante ningún juicio; tendría que proceder argumentando sólo con juicios no-afirmativos; pero de estos no podría derivar CD, que es él mismo afirmativo. Si Descartes emplea CD para argumentar a favor de él, caerá en el círculo (o al menos aparentemente); y en efecto lo hará a menos que tenga otra opción. Una solución al problema del círculo ha de consistir, entonces, en “dar esta otra opción”.⁸

Si se lo deja sin responder, el problema del criterio ataca a todas las teorías epistemológicas de igual manera, y el resultado de ello es la caída en el escepticismo más global (ningún juicio puede sostenerse). Si Descartes o cualquier otro tienen efectivamente conocimiento, el problema debe resolverse del algún modo. Es posible, por esto, emplear la estrategia “naturalista” de considerar el conocimiento como un *factum*, para avanzar a examinar la manera en que se “prueba” CD, independientemente de que se tenga una solución al problema del círculo. El escéptico objetará *especialmente* esta movida; simplemente pasa por alto el problema escéptico que *al parecer* se quería contrarrestar. De cualquier manera, detenerse en espera de una solución al problema evitaría pasar a varios puntos de las *Meditaciones* que están estrechamente vinculados con el tema de la modalidad, que es lo que nos preocupa aquí.⁹

⁷Para una discusión de varias maneras en que podría tratar de resolverse el círculo cartesiano, véase van Cleeve (1998). Para una discusión general del problema del criterio, Chisholm (1973).

⁸Quizás Descartes consideró que el que no encontrara una opción a caer en el círculo indicaba que CD era verdadero necesariamente, y que por lo tanto, literalmente, no había otra opción. Considérese la segunda máxima de la moral provisional de Descartes:

...ser en mis acciones lo más firme y resuelto que pudiera y seguir tan constante en las más dudosas opiniones, una vez determinado a ellas, como si fuesen segurísimas... Y así, puesto que muchas veces las acciones de la vida no admiten demora, es verdad muy cierta que si no está en nuestro poder discernir las mejores opiniones, debemos seguir las más probables...

que da pie a una interpretación metodológica de la (posible) caída de Descartes en el círculo. Una interpretación de este tipo, sin embargo, no parece plenamente satisfactoria, porque no se ve cómo de una consideración meramente metodológica se puede pasar a la fundamentación de un principio epistémico como CD. Véanse las objeciones de van Cleeve (*op.cit*) a Gewirth y Feldman.

⁹La solución de van Cleeve (1998, 110 y ss.) consiste efectivamente en una movida de este tipo. Ella se basa en la distinción entre estas proposiciones:

A: para todo *p*, si percibo clara y distintamente que *p*, estoy seguro de que *p*

Reiterando: Descartes requiere de una prueba de CD. Ahora bien, una prueba para esta tesis tienen que mostrar que no puede engañarse sobre lo que concibe, *ni ser engañado*. Los párrafos que siguen a la cita exploran la posibilidad de que un Dios le engañe incluso en las cosas que le parecen más ciertas, al mismo tiempo que expresan la incapacidad de, en esos casos, concebir las cosas de otro modo. Descartes llega a la conclusión de que debe probar la existencia de Dios y de que Dios no es engañoso. La apelación a Dios no es arbitraria aquí, como se verá: si Dios no es engañoso, las facultades de los sujetos, como cosas creadas, no pueden ser engañosas, porque eso manifestaría una imperfección en Dios, y de este modo se prueba de la misma manera que el sujeto cartesiano no se engaña en lo que concibe ni es engañado. Con esto se tiene la prueba cartesiana de CD, esencialmente. Sin embargo, hay que detenerse en varios detalles del transcurso del argumento, y recorrerlo con mayor detalle.

Descartes quiere mostrar aquí que poseer la idea de Dios implica con seguridad la existencia de Dios.¹⁰ El argumento es presentado de forma sintética, avanzando desde sus antecedentes a las tesis que conducen lógicamente a su conclusión; Descartes dice que “para encontrar ocasión de indagar [si hay Dios y si puede ser engañador], sin interrumpir el orden que me he propuesto en estas meditaciones, que es pasar gradualmente de las primeras nociones que halle en mi espíritu a las que pueda encontrar, *debo dividir aquí mis pensamientos todos en ciertos géneros, y considerar en cuáles de estos hay propiedad o error*” (p. 139). ¿Por qué debe hacer *precisamente* esto? Es implausible que Descartes argumente a tientas; es plausible en cambio que el rumbo de la tercera meditación esté diseñado de modo que permita satisfacer su motivación (a saber, la resolución de la dificultad sobre la regla CD).

Sean cuales sean las razones por las que Descartes sigue el rumbo de la tercera meditación, Descartes hace una serie de distinciones de gran importancia. Primero, distingue tres clases de cogitaciones: las ideas, los juicios y las afecciones. Las ideas son “como las imágenes de las cosas”,

B: tengo certeza de que (para todo p , si percibo clara y distintamente que p , entonces p)

Lo que propondría Descartes en la cita que se examinaba de la tercera meditación es a). Lo que Descartes probaría en la cuarta meditación sería b). a) dice que cuando concibo una proposición clara y distintamente, tendré certeza de *ella*, mientras que b) dice que tengo certeza de un principio epistémico determinado, a saber, que si percibo clara y distintamente algo, es verdadero (sic). a) puede ser verdad del alguien sin que b) sea verdad de él. En esta propuesta, resulta que no hay realmente un círculo. Si A es verdadera para Descartes, puede avanzar mediante las ideas que perciba clara y distintamente, y llegar a probar CD. Esto hace la propuesta eminentemente interesante, incluso tal vez correcta.

Ahora bien, van Cleeve propone que A es un *hecho* que permite a Descartes avanzar su argumento a favor de CD. Si A es un hecho, no obstante, el conocimiento es un hecho también, de modo que esta solución esta expuesta de igual modo a la crítica escéptica.

¹⁰En la *Quinta Meditación* Descartes propone un argumento alternativo para demostrar la existencia de Dios. El punto de Descartes es que es integral a la esencia de un ser perfecto y supremo el existir necesariamente. Hay que notar que en ese punto del texto la posición epistémica de Descartes es distinta a la que tiene al comienzo de la *Tercera Meditación*: ha establecido ya CD, lo que le permite afirmar con certeza sus intuiciones sobre la esencia de Dios. Para el sujeto de la *Quinta Meditación*, la existencia de Dios es auto-evidente: no puede concebir a Dios salvo existiendo.

mientras que las afecciones y los juicios son el resultado de actos que “añaden algo” a las ideas de las cosas que son el tema de la mente (sic). Así, por ejemplo, en un juicio se afirma algo sobre una cosa, si quiero algo tengo una disposición especial sobre esa cosa, etc. La diferencia crucial está en que, en términos más contemporáneos, las afecciones y los juicios son actos *esencialmente* intencionales, mientras que las ideas no lo son, al menos no esencialmente.¹¹ El problema sobre las ideas es que es preciso determinar cuáles tienen cierta relación con cosas externas, dado que es sobre esta relación que, según Descartes, se cae en error más frecuentemente al hacer juicios. Al plantear esto, Descartes abre el problema del conocimiento del mundo externo en su manera más directa: ¿qué razones se puede tener para creer que las cosas que se creen ser externas son en realidad tales, y no meramente producto de la actividad de la mente? Se da pie a las posibilidades del idealismo solipsista, o la más cercana al programa cartesiano, de que no se logre validar una física que no sea una pura explicación de los “fenómenos” (cf. Kant, donde este problema es más obvio); incluso si logra probar la existencia de Dios, tiene luego que ganar el mundo. En la tercera meditación el problema queda pendiente: Descartes sólo nota que no le parece convincente, por un parte, que sea intuitivo que las cosas de las cuales las ideas son supuestamente imágenes sean externas, y por otro, que esas ideas sean independientes de su voluntad (el argumento que propone es una variación del argumento del sueño).

Nótese, por lo demás, que en esta teoría no se puede errar ni en las afecciones ni en las ideas: si quiero algo, dice Descartes, no puede dejar de ser verdad que lo quiero; y si imagino algo, aunque no sea más que una quimera, no deja de ser verdad que la imagino. Esto será reafirmado en la cuarta meditación, y como se verá, es central a la epistemología cartesiana en lo que respecta al conocimiento modal.

Avanzando, Descartes examina qué sería el caso *si* las ideas que se tienen fueran (o pudieran ser) imágenes efectivas de cosas externas. En este caso, dice, y en oposición a lo que pasaría si no lo fueran, habría una diferencia entre ellas: no todas tendrían la misma “realidad objetiva”, es decir, no tendrían el mismo grado de ser o perfección por representación. Mediante juicios correctos sobre el grado de realidad objetiva de una idea podríamos determinar que algo es una sustancia o un modo (pues Descartes dice que la idea de una sustancia tiene más realidad objetiva que la de un modo), y es probable que Descartes esté pensando aquí en cierto tipo de juicio relacionado con el que se puede hacer sobre si algo pertenece a alguna de estas categorías: aquél en que no solamente se ve que se pueden atribuir a una cosa tales o cuales cosas, sino que estos atributos entallan tal o cual grado de perfección. Descartes emplea aquí el término “realidad formal” o “actual”, que no tiene

¹¹ Este es un punto contencioso, tal vez; sin embargo, no creo que pueda decirse *prima facie* ni que las ideas cartesianas no puedan tener cierta intencionalidad, porque algunas de ellas son imágenes de cosas existentes, ni que no puedan no ser intencionales, porque algunas de ellas no son imágenes de nada que exista (ellas *son* aquello *de* lo que se supone que son). Quizás alguien querría construir este último caso de modo que tales ideas sean en realidad montones de juicios que tiene la apariencia de una idea. Si esto fuera posible, podría tal vez también ser el caso que todas las ideas no fuesen sino montones de juicios.

que ver tanto con la cuestión de la perfección sino con la de la finitud: las sustancias infinitas tiene mayor realidad formal que las sustancias finitas, y estas a su vez mayor realidad formal que las propiedades. Sin embargo, en el caso de la existencia de Dios la cuestión de la perfección hace más comprensible la distinción. Hay que tener presente además que la realidad objetiva es una propiedad exclusiva de las ideas, mientras que la realidad formal es una propiedad de cualquier cosa (aunque el planeta Tierra tiene realidad formal, no tiene realidad objetiva).

Aquí vale la pena observar que según Descartes, desde un punto de vista epistémico, no es posible determinar que algo es una sustancia salvo mediante las atribuciones que se hacen sobre la cosa en cuestión, y que, desde el punto de vista metafísico, las sustancias se manifiestan en sus propiedades.¹² Así, por ejemplo, se determina que hay una mesa porque se determina que tal o cual figura, color, tamaño, etc. corresponden a algo.

La idea que tiene mayor realidad objetiva es la de Dios, pues Dios contiene en sí todas las perfecciones y es infinito, y por lo tanto tiene la mayor realidad formal (como se apuntaba antes, el grado de realidad objetiva de una idea debe ser equivalente al de realidad formal del objeto que representa si este existe). Tal idea, propone Descartes, no puede provenir de sí mismo (esto es, del sujeto cartesiano), pues es finito, y es imposible que algo con menor realidad formal sea causa de algo de mayor realidad formal. En particular, las varias propiedades que le corresponden a Dios según la idea que se tiene de él no pueden provenir de cosas de las que tenemos idea, pero que reconocemos como finitas (así, Descartes dice que 1) las propiedades de Dios no son meramente la negación de las limitaciones que ve en sí mismo, 2) el que creamos que Dios las tiene no es el producto de nuestra naturaleza finita y falible, y 3) esas propiedades no son una proyección de la perfectibilidad que ve en sí). Dadas estas condiciones, Descartes argumenta que la idea de Dios es innata, así como la idea de sí mismo; no podría ser adventicia ni una mera “creación del espíritu”. Más aún: habría sido producida en él por Dios mismo, pues si no fuera así habría que explicar de dónde ha provenido en quien la pusiera en su mente, en cuyo caso podría caerse en una regresión al infinito (cf. pp. 150–151).

Descartes resume:

... toda la fuerza del argumento que he empleado hasta aquí para probar la existencia de Dios consiste en que reconozco que no podría ser mi naturaleza lo que es, es decir, que no podría tener yo en mí mismo la idea de Dios, si Dios no existiese verdaderamente, ese mismo Dios, digo, cuya idea está en mí, es decir que posee todas esas elevadas perfecciones, de las cuales puede nuestro espíritu tener una ligera idea, sin poder, sin

¹²Otros intérpretes construyen la noción de “realidad objetiva” como equivalente a la de intencionalidad. Como se ha notado antes, no creo que la intencionalidad sea una propiedad esencial de las ideas cartesianas, mientras que la realidad objetiva debe serlo. En la propuesta que se hace aquí, la realidad objetiva de una idea es una propiedad lógica de ella, provisto que tenga sentido decir que algo tenga o no perfecciones. Por otro lado, la noción de realidad objetiva admite grados, y es difícil ver cómo puede ser así para la intencionalidad.

embargo, comprenderlas, y que no tiene ningún defecto... por donde resulta que no puede ser engañador... (pp. 152–153)

Es decir, concebido Dios como la cosa que reúne todas las perfecciones, no se puede pensar que Dios sea engañador. Se tienen ya todos los elementos para la prueba de CD.

Dios es garante de la facultad de concebir porque tal facultad no pudo haber sido otorgada al sujeto cartesiano con el único propósito de conducirlo al error: debe pues ser *capaz* de conducir a la verdad. No puede ser engañado, entonces, por Dios, ni puede engañarse él mismo del todo. Sin embargo, puede errar, sin que esto implique contradicción con la tesis de que Dios es perfecto (pues no es comprensible que el errar sea un bien, etc.)

El paso siguiente es el crucial. Según Descartes, el error es producto de la interacción de la facultad del entendimiento y la voluntad. Mediante el uso del entendimiento, nos dice, *no se afirma ni niega cosa alguna, sino sólo se forman las ideas que pueden afirmarse o negarse* (sic, p. 158). No hay, por tanto, error posible en el uso del entendimiento. La voluntad, por su lado, consiste “tan sólo en que podemos hacer una cosa o no hacerla, es decir, afirmar o negar, buscar o evitar una misma cosa”. El juicio depende de la voluntad, por lo consiguiente; en ella debe buscarse la fuente del error. Este consiste en la operación de la voluntad por sobre los límites del entendimiento. En su operación regulada, la claridad en el entendimiento guía a la voluntad a juzgar correctamente, pero ya que aquél es más limitado que esta, que puede considerarse infinita, ella puede obrar en independencia de las condiciones del entendimiento. Si el juicio coincide entre la voluntad y el entendimiento, lo que se concibe ha de ser verdadero, ya que las ideas que el entendimiento concibe no pueden ser engañadoras (lo que se sigue de que Dios mismo no es engañador, como se ha visto antes):

mientras contengo mi voluntad dentro de los límites de mi conocimiento, sin juzgar más que de aquellas cosas que el entendimiento representa claras y distintas, no puede suceder que me equivoque, porque toda concepción clara y distinta es, sin duda, algo, y por lo tanto, no puede provenir de la nada, y debe ser necesariamente obra de Dios, quien siendo sumamente perfecto, no puede ser causa de error; y por consiguiente, hay que concluir que esa concepción o ese juicio ha de ser verdadero.

Una vez probado CD, es sencillo ver como C (descrito en la sección anterior) calza en el sistema cartesiano. Si todo lo que concibo clara y distintamente es verdadero, tiene por fuerza poder ser creado por Dios, y *es posible* (el problema que implica la introducción de Dios en la conceptualización de lo posible se tratará más adelante, como se ha mencionado antes).

El argumento de la cuarta meditación es rico en consecuencias para la epistemología modal cartesiana, como se verá a continuación.

1.2.2.3. *La segunda fase: el argumento aplicado*

La segunda fase del argumento sobre el dualismo cartesiano es meramente la aplicación de los principios establecidos en la primera a la cuestión de la distinción entre el alma y el cuerpo. Esta fase del argumento cartesiano tiene, a su vez, dos partes. En la primera reitera sus distinciones sobre la naturaleza del alma (que se basan en la certeza del *cogito*) y del cuerpo; en la segunda, aplica el esquema formal de *b* a esos conceptos.

Esta fase del argumento procede esquemáticamente así:

- 1) Me concibo como una cosa que piensa y que no tiene extensión, y
- 2) Concibo a mi cuerpo como algo extenso y que no piensa, de modo que,
- 3) Es posible que mi mente exista sin mi cuerpo, y viceversa, por lo tanto,
- 4) Soy [*mi mente es*] distinto de mi cuerpo

Ya se discutió la validez formal de argumentos de este tipo, y las estructuras formales que la filosofía cartesiana provee para ellos. Si las premisas son ciertas, la teoría cartesiana permite sostener la conclusión.

¿Qué significa “concebir” en la teoría cartesiana? Concebir es la operación del entendimiento, a saber, la formación de ideas. Como se ha visto, no puede decirse que el contenido de las ideas sea falso o errado; son los juicios sobre las ideas los que pueden fallar. ¿Qué significan entonces las tesis 1) y 2) del argumento? En lo que sigue se esbozará una propuesta de interpretación.

Así como una idea puede ser la imagen de una cosa, una idea puede ser también la representación de un “estado de cosas”. Supóngase la idea

- a) llueve en un determinado momento y lugar t, l

Como idea, esta representación no incluye que el contenido se afirme o niegue. Un sujeto que sólo tenga esa idea (pero que no esté en condiciones de hacer un juicio sobre ella, es decir, un sujeto “indiferente”, en la terminología que emplea Descartes) tiene, sin embargo, la capacidad de afirmar que:

- b) no puedo afirmar ni negar que llueva en t, l

esto es, está obligado lógicamente a adoptar sobre p una actitud distinta a la de la aserción. No puede, por tanto afirmar, p ni $\neg p$, pero quizás puede afirmar algo que no implique ninguna de estas opciones, pero que sin embargo mantenga relación con el contenido de p . Una función del juicio (en el sentido de forma de aplicarse el acto del juicio sobre un contenido) que puede generar tal afirmación es la de la posibilidad: que algo sea el caso implica que sea posible, pero no viceversa, y que algo no sea el caso no implica que no sea posible. De este modo, a cada idea le correspondería un juicio de posibilidad *correcto*; para un sujeto del cual es verdad que

- c) S concibe p

es verdad que

d) S puede afirmar que $\Diamond p$

Este puede ser llamado el principio de la concebibilidad como guía de la posibilidad (CCP). La afirmabilidad de $\Diamond p$ se mantiene en principio; CCP no implica que S tenga que afirmar la posibilidad.¹³

En este modelo, los juicios de posibilidad son un caso extremo del juicio en que no es posible que la influencia de la voluntad haga caer en el error. En este sentido, los juicios de posibilidad son similares a los juicios sobre las apariencias (si veo rojo, no puedo decir que no veo rojo). Por otro lado, no requieren más fundamento que el mero hecho de que se tienen ciertas ideas; el conocimiento de lo posible como posible es, en cierto sentido, inmediato (intuitivo).

La confianza cartesiana aparente respecto a los juicios de posibilidad se funda, pues, en su teoría de las ideas; en particular, en la suposición de que las ideas no pueden ser erradas. La manera en que se justifica el juicio de posibilidad lo hace tan trivial que uno no puede dejar de pensar que la teoría tiene serias faltas. Por un lado, no hay ninguna noción de estratificación de la noción de posibilidad que se construye de esta manera. La única restricción que se hace sobre lo posible es una de composibilidad como idea, pero la manera en que se determina esto no es explícita. Podría ser equivalente a una forma de consistencia en términos puramente lógicos (i.e., no-contradicción), pero podría ser también algo más sustantivo. Por otro lado, esta teoría no permite, y esto tiene que ver con lo anterior, hacer un juicio sobre si los juicios de posibilidad establecidos de este modo corresponden a posibilidades en el sentido más amplio o en uno estrecho; esto se refleja en la doctrina de la doble significación de “posible” desde la perspectiva de Dios y la de los sujetos finitos, que se examinará en la próxima sección. Estas deficiencias de la teoría de los juicios de posibilidad cartesiana ponen en cuestión el verdadero alcance del argumento de separación (¿es la separación meramente lógica? ¿son la mente y el cuerpo *en concreto* separables? El argumento no lograría determinar esto; de la claridad y la distinción de la representación de esta posibilidad no se siguen distinciones sobre de qué nivel de posibilidad se trata—al menos, no parecen seguirse necesariamente).¹⁴

La premisa 3) del argumento se sigue mediante CCP, de la mera realización de que se tiene la idea de los escenarios propuestos. Una vez establecido esto, se sigue la conclusión del argumento,

¹³Incidentalmente, los filósofos pirrónicos parecen construir el juicio de posibilidad de la misma manera, reduciéndolo a una forma de no-afirmación (cf. Sexto Empírico, en los *Esbozos de Escepticismo*). Esto es metodológicamente significativo para ellos, porque argumentan precisamente la posibilidad de mostrar que tanto una tesis como su contraria son afirmables (*isostenia*) muestra que ninguna realmente lo es. Sería interesante desarrollar la idea del método cartesiano como una extensión positiva de la metodología escéptica: así como los pirrónicos se basan en no-afirmaciones para sostener su práctica de no-afirmación, Descartes se basaría en no-afirmaciones para sostener juicios sobre necesidades.

¹⁴Lo concretamente posible debe ser lógicamente posible, pero no es cierto que la mera posibilidad lógica deba ser relevante para la cuestión de la separación de la mente y el cuerpo. Me parece verdad que el mundo real podría no ser un mundo posible en que la mente y el cuerpo son concretamente separables. Esto no corre como crítica a Descartes, quien no hace esta distinción.

empleando el esquema conceptual establecido en el argumento general.

1.3. LA EPISTEMOLOGÍA MODAL CARTESIANA VERSUS SU METAFÍSICA MODAL

Cuando se discutió el núcleo de la epistemología modal cartesiana (a saber, las tesis que dan pie al argumento sobre el dualismo en su primera fase), se mencionó que Descartes hace una distinción entre lo que es posible para Dios y lo que es posible para nosotros. Esto trae problemas a la teoría cartesiana.¹⁵

Llévese el problema a un extremo. Considérese el juicio de posibilidad

a) Es posible que ninguna mente pueda separarse de un cuerpo.

(es decir, que es posible que la conclusión del argumento de separación sobre el dualismo sea falsa). Con base en lo que Descartes parece pensar sobre la relación entre la necesidad y la posibilidad (a saber, que si algo es necesario es necesariamente necesario), esta posibilidad implica que no es necesario que la mente sea separable del cuerpo, y esto a su vez significa que no es esencial a la mente ser distinta al cuerpo. Pero esto es precisamente algo que Descartes tendría que negar. Ahora bien, en la postura cartesiana esta posibilidad no puede negarse a Dios, en el sentido de que Dios debería poder hacerla verdadera. Pero si eso es el caso, entonces hay una manera de que no sea esencial a la mente ser separable del cuerpo, y Descartes tendría que aceptar el juicio de que no lo sea.

Esto no lleva, sin embargo, a Descartes directamente a una contradicción, porque puede apelar a la idea de que el concepto de posibilidad no es unívoco y decir que el juicio que uno, como ser finito, puede hacer sobre la esencia de la mente, y el juicio que Dios, como ser infinito, puede hacer sobre la esencia de la mente, no tienen el mismo contenido. El problema aquí es que al hacer este movimiento el concepto de *esencia* también realiza un deslizamiento, y se cae en la dificultad de que no hay una manera clara de distinguir en qué sentido pueden tener alguna relación semántica uno y otro juicio (y es claro que deben tener alguna; lo que se intenta hacer aquí, porque hay que hacerlo si no queremos quedarnos en el nivel del uso irreflexivo del sentido común, es explicar por qué es claro que es así).

¹⁵Encontramos un ejemplo de una doctrina cartesiana confusa (al menos *prima facie*) por esta dualidad en la teoría en el “Resumen” de las *Meditaciones*:

...todas las sustancias, es decir, todas las cosas que no pueden existir sin ser creadas por Dios, son por naturaleza incorruptibles, y no pueden nunca cesar de ser, como no las reduzca a la nada Dios...

¿Cómo puede Descartes dar cuenta de esta posibilidad condicional? Quiere decir que es posible que las cosas puedan dejar de ser, por un lado, desde la perspectiva de Dios. Pero ¿qué pueden decir los sujetos finitos sobre esta posibilidad? ¿Puedan afirmar que no pueden cesar de ser, o no pueden afirmar nada, dado que no saben si Dios hará que cesen de ser, y no pueden saberlo?

Supongamos que la mente tiene una esencia; esta reúne el conjunto de las cosas que *tienen* que ser verdaderas sobre ella. El juicio de que la mente es separable del cuerpo puede entenderse como un juicio sobre la esencia de la mente, y además, dentro del sistema cartesiano, como un juicio correcto sobre una idea clara y distinta. Si se toma esto desde un punto de vista realista, tenemos que decir que estos últimos aspectos del juicio significan que la separabilidad de la mente y el cuerpo es una propiedad esencial *objetiva* de ellos. Esto quiere decir fundamentalmente que no se puede emitir un juicio *correcto* que niegue la verdad de este juicio. Ahora bien, según el enfoque cartesiano *tenemos* que emitir el juicio de que la mente es separable del cuerpo, porque no podemos concebirla de otro modo, pero al mismo tiempo no podemos decir que Dios tiene que emitirlo al igual que nosotros. Esto nos pone en la posición de que no podemos negar que Dios pueda hacer que la mente sea inseparable del cuerpo, aunque *lógicamente* deberíamos poder hacerlo (porque que Dios pueda hacerlo implica que puede ser el caso, y tenemos razones para no aceptar que pueda ser el caso). Si aceptamos además que tenemos que decir que Dios no puede no poder hacer cualquier cosa, ya sea concebible como inconcebible por nosotros, entonces la esencia (y en consecuencia, la existencia) de Dios se vuelve incompatible con la esencia de la mente: si Dios existiera, no podría ser necesario (es decir, verdadero sin una posible excepción) que la mente sea separable del cuerpo. Tenemos que abandonar entonces, o bien que la esencia de Dios no implique restricción alguna sobre sus capacidades, o que la separabilidad de la mente y el cuerpo son necesarias.

Si se hace lo segundo pierde sentido el argumento de separabilidad a favor del dualismo. La tesis del dualismo se compone de tesis sustantivas sobre la esencia de la mente y el cuerpo, y es la pieza fundamental de la ontología construida por Descartes sobre la base de sus avances epistemológicos. Si las tesis sobre la esencia de la mente y el cuerpo se abandonan, la ontología cartesiana desaparece.

Si se hace lo primero en el contexto de la filosofía cartesiana se llega a varios problemas. Recuérdese que la bondad de Dios garantiza que el sujeto cartesiano no se engaña. Si Dios no pudiera hacer algo que parezca imposible no podría satisfacer este rol de garante del juicio. La hipótesis del genio maligno, que Descartes contrarresta con el juicio claro y distinto sobre la bondad de Dios, pone al sujeto cartesiano frente a la posibilidad de que todos los juicios, incluso aquellos que se basan en ideas claras y distintas, sean falsos. En esta condición, un sujeto cartesiano podría hacer el juicio que $1 + 1$ es igual a 2 (por ejemplo), y *equivocarse*—la negación del juicio debería ser verdadera; pero esto es algo que los sujetos cartesianos no pueden concebir, y por lo tanto, no pueden corregir su juicio (se asume que no hay manera de que los sujetos cartesianos sepan que hay un genio maligno que los engaña). Así, desde la perspectiva de Dios deberían ser posibles cosas que no son posibles desde la perspectiva de los seres finitos. Considérese ahora que los sujetos cartesianos pudieran corregir sus juicios en esa situación (si bien esta posibilidad queda obstruida por la hipótesis de que los sujetos cartesianos no pueden concebir estas aparentes imposibilidades, ella podría abandonarse). Si eso es posible, el Dios cartesiano tiene que poder garantizar la verdad de

los juicios que expresan tanto posibilidades como imposibilidades (desde nuestro punto de vista).¹⁶

Ahora bien, si Dios no es omnipotente en el sentido de poder hacer cualquier cosa, ya sea posible o imposible desde la perspectiva de los sujetos cartesianos, un genio maligno podría en principio engañarnos incluso si Dios fuera benevolente y garantizara la mayoría de nuestros juicios, y lo que es peor aún, no podríamos saber cuáles de nuestros juicios son correctos, a menos que tuviéramos un principio para distinguirlos (en este caso el genio maligno sería enteramente inefectivo, de todas maneras; es más interesante considerar el caso en que la actividad del genio maligno es efectiva). Nada de esto es problema si Dios es omnipotente, porque en ese caso todos nuestros juicios claros y distintos son correctos necesariamente (o al menos pueden serlo).

Quizás la tesis del dualismo no deba tomarse en el sentido realista propuesto. Una alternativa que parece viable para construir la distinción entre la noción de posibilidad desde la perspectiva de Dios y la de un sujeto finito evitando las dificultades que se han señalado consiste en enfatizar la noción de afirmabilidad en lugar de la de verdad. La noción de afirmabilidad puede construirse de modo que algo puede *no* ser afirmable, y sin embargo ser verdadero (lo que, dicho de paso, parece ser correcto). Este principio puede usarse en la interpretación de la metafísica cartesiana de modo que a) sólo desde la perspectiva de Dios pueden afirmarse todas las posibilidades *reales*, es decir, objetivas, y b) el criterio de afirmabilidad de posibilidades para los seres finitos determina un concepto subjetivo de posibilidad, que sólo en virtud de la agencia divina coincide con una parte de la extensión del concepto de posibilidad real (es decir, que es contingente en el hecho de la creación). Las condiciones de afirmabilidad del concepto subjetivo de posibilidad no tienen que coincidir con las del concepto objetivo de posibilidad: las razones por las que Dios podría juzgar que tal o cual cosa es posible podrían ser enteramente distintas a cualquiera que un sujeto finito podría tener para afirmar que esa cosa es posible; se trataría de conceptos enteramente separables.

Una variante interesante de esta idea consiste en la tesis de que no sabemos qué es posible, pero no tenemos opción sino afirmar que tales y tales cosas son posibles. Con esto el problema del doble aspecto que introducen las consideraciones sobre la peculiaridad de la perspectiva de Dios se desliza a la cuestión del conocimiento; claramente esta es una dirección muy poco cartesiana, aunque él parece concebir el conocimiento como una variedad de condición de certeza inquebrantable en vez de una forma de creencia justificada (es decir, que tiene una concepción de la noción de conocimiento más subjetiva que la más habitual). Requiere, por lo demás, de una revisión de la noción cotidiana de afirmabilidad, que sí parece tener un vínculo estrecho con la de conocimiento.

¹⁶Dicho de otro modo: Dios garantiza la verdad de los juicios correctos en todos los mundos posibles. Si en un mundo posible m_1 un juicio que en un mundo posible m_2 sería un juicio correcto es un juicio correcto, Dios tiene que poder garantizar la corrección y la incorrección del juicio. Dios no podría hacer esto si no pudiera garantizar la verdad de juicios que expresan una imposibilidad relativa a mundos particulares.

1.4. ALGUNAS PERSPECTIVAS SOBRE LA OPCIÓN DEL DUALISMO COMO ILUSIÓN METAFÍSICA

¿Qué pasaría en el caso que el argumento cartesiano a favor del dualismo falle? (Es decir, si a pesar de todo, el argumento falla, sea cual sea la mejor manera de entender lo propuesto por Descartes.) Se puede preguntar por qué podría haber fallado, y qué consecuencias tendría este fallo. El argumento puede fallar en dos niveles diferentes: primero, puede fallar en probar que la mente *de hecho* está separada del cuerpo (y viceversa), y en segundo lugar, puede fallar en mostrar que una mente *puede* estar separada de un cuerpo. La recepción del argumento muestra que si bien muchos han considerado que ha fallado en el primer sentido, muchos han aceptado lo que propone en el segundo sentido (es probable que quienes lo han aceptado hayan creído de antemano en esta idea.)

Como (se espera) se ha visto, el argumento cartesiano tiene múltiples puntos en los que podría fallar (el “problema del círculo”, es decir, la cuestión de si Descartes asume su conclusión en su demostración de la existencia de Dios, es quizás el obstáculo más importante para el éxito del argumento); esto vale para el fallo del argumento tomado en cualquiera de los dos niveles descritos. Alguien podría querer rebatir cada una de las asunciones que Descartes hace, incluso las que parecen más neutrales. La opción más interesante, en el contexto de la investigación presente, es que podría haber fallado porque las tesis modales en que se sustenta no se satisfacen. Por un lado, podría ser materialmente falso que Descartes haya concebido los escenarios que constituyen las premisas del argumento de separación; podría ser incluso materialmente falso que pudiera haber concebido esas premisas. Esto no significaría, no obstante, que la conclusión tuviera que ser falsa. Supongamos ahora que nadie podría haber concebido esos escenarios; ¿qué sería el caso entonces? Si se los ha podido comprender en realidad (y es plausible que muchas personas hayan comprendido los escenarios cartesianos), todos quienes los han comprendido han estado equivocados sobre los escenarios cartesianos, y plausiblemente han estado equivocados sobre muchos otros casos. La cuestión entonces es qué tan profundo tendría que ser el cambio conceptual requerido para rectificar estas creencias. Por supuesto, no se puede determinar esto a menos que se disponga de todos los elementos que habrían hecho que el argumento fallase, y además, que sea posible evaluarlos desde una perspectiva externa que permita explicar el fallo. Es decir, el peso de la prueba se desliza desde los problemas internos a la teoría cartesiana a la cuestión externa de si la opción que propone es correcta o no. Las cuestiones externas son a menudo las más difíciles de justificar *filosóficamente* (desde una perspectiva puramente teórica; las cuestiones externas *son* resueltas constantemente por razones prácticas), por lo que se puede preferir no hacer más que explorar el espacio lógico de posibles respuestas a ellas.

Por ahora basta considerar la opción de que la posibilidad de la existencia separada de la mente y el cuerpo es una ilusión metafísica (es decir, que el argumento falla a ambos niveles). Con esto se quiere decir que se constituye desde una perspectiva que genera un error en el juicio, de modo que la posibilidad no es tal, el escenario solo aparentemente es válido. Para sostener esto es nece-

sario adoptar una postura que niegue varios sentidos en que puede ser tomado el principio de la concebibilidad como guía de la posibilidad; a saber, los que son compatibles con que el escenario cartesiano sea posible (entre ellos está el que se ha propuesto antes). Las motivaciones de una postura de este tipo pueden ser varias. Puede ser atractiva para el fiscalista/materialista porque bloquea el paso a uno de sus opositores históricos—aunque, por un lado, en la actualidad el dualismo de sustancias es el menos atractivo de sus opositores, y por otro, que el escenario sea espurio no le favorece en un sentido sustantivo, ya que igualmente puede ser utilizado para avanzar una postura idealista. A quien realmente compromete la cuestión es al monista, para quien no es satisfactorio admitir siquiera la posibilidad de un mundo en que la ontología dualista sea el caso.¹⁷

1.5. OTROS PROBLEMAS DE LA METAFÍSICA CARTESIANA

El argumento cartesiano a favor del dualismo y el problema epistemológico de los juicios posibles se ramifican en varias direcciones en la teoría cartesiana, aparte de las que ya se han considerado. En esta sección se revisarán dos ejemplos de tales direcciones, que refuerzan de manera particular la idea de la importancia sistemática del problema de los juicios de posibilidad para la filosofía cartesiana. Por un lado, está la cuestión del concepto de sustancia en la ontología cartesiana, y por otro, el de la posibilidad del escenario del soñador cartesiano.

¹⁷Se piensa aquí en una forma de monismo mínimo que se sostenga sobre la tesis epistémica:

m_1 no es posible dar sentido a la idea de que exista más de un tipo fundamental de entidad

Lo que esta tesis propone es que es forzoso ser monista (o bien no emitir juicios sobre qué tipos de entidades son fundamentales). Puede objetarse a un monista de este tipo que el concepto de que exista más de un tipo fundamental de entidad no es lógicamente contradictoria, a lo que puede replicar que aún así puede ser conceptualmente contradictoria, o incluso meramente *inacceptable*, en un sentido apropiado (que algo sea contradictorio puede construirse como un tipo de inaceptabilidad).

Un par de observaciones sobre esta postura:

- 1) El monista podría proponer que lo que se sabe sobre la posibilidad de que exista más de un tipo fundamental de entidad es que si fuera verdadera, tales y tales cosas serían verdaderas, pero nada más, y que esto no es suficiente para admitir una posibilidad (en este caso el monista debería proponer es una epistemología revisada del juicio de posibilidad).
- 2) Si el monista tiene razón en este punto, quizás se pueda aplicar esta estrategia al análisis de otros conceptos filosóficos. Por ejemplo, al de entidad supernatural (las razones para aplicarla al dualismo y al concepto de lo supernatural podrían ser incluso las mismas).

1.5.1. EXISTENCIA, CANTIDAD Y LA CATEGORÍA DE SUBSTANCIA

El argumento cartesiano a favor del dualismo puede tomarse en un segundo sentido, que denominaré *categorial*. En este sentido, el argumento tiene como resultado la justificación de la posibilidad de que digamos que tanto la *res cogitans* como la *res extensa* son tipos de *substancia* (es decir, que la *res cogitans* y la *res extensa* son sustancias, y por lo tanto, que la categoría de sustancia no está vacía).

Descartes mismo nunca habría propuesto que el argumento tiene esta función. La categoría de sustancia es un elemento conceptual que Descartes utiliza sin mayor justificación, y en efecto esto difícilmente podría haberse remediado. Por un lado, ya en su época el *término* estaba fuertemente identificado con la idea elemental de “algo existente”, y por esto mismo habría sido improductivo utilizar otra terminología. Por otro lado, la ontología cartesiana es tan minimalista (en un sentido que vamos a clarificar enseguida) que cualquier otro esquema conceptual sólo habría introducido complicaciones (como, dicho sea de paso, ocurriría después con las ontologías de Spinoza y Leibniz).

En las *Meditaciones*, el concepto de una sustancia es entendido básicamente en el sentido proto-gramatical¹⁸ aristotélico de aquello que tiene propiedades. Así, una sustancia puede ser extensa, pero de la extensión no puede decirse que sea ella misma una sustancia. Claro está que la extensión de algo puede tener ella misma propiedades (como se verá, en la ontología cartesiana esta observación tiene gran importancia), así que esta caracterización de la noción de sustancia es aún demasiado vaga. Una solución tentativa para esta dificultad consiste en proponer que una sustancia es aquello que puede tener propiedades, pero que no puede ser ello mismo propiedad de otra cosa. De cierta manera, los problemas ontológicos de la noción cartesiana de sustancia se derivan todos de la cuestión de por qué y cómo habría de ser eso el caso.

Un enfoque que Descartes mismo desarrolla en las *Meditaciones* es enlazar la noción de sustancia a la de independencia. Una sustancia es, en esta dirección, algo que puede existir independientemente. Con este movimiento conceptual se abre la posibilidad de interpretar el argumento de separación en un sentido categorial. Visto desde este punto de vista, él *prueba* la sustancialidad (en el sentido de existencia independiente) de tanto la mente como el cuerpo. Lo que propone es que si la mente puede existir separada del cuerpo, puede existir independientemente (es sustancia), y si el cuerpo puede existir separado de la mente, puede existir independientemente (ídem.).

El argumento tomado en este sentido es, sin embargo, incompleto. El paso crucial necesario para validar este argumento es el del concepto de “independiente de...” al de “...independiente en un sentido absoluto”, pero en ningún lugar del argumento de separación se lo hace explícitamente.

¹⁸Proto-gramaticales son aquellas distinciones que se basan en o son paralelas a distinciones gramaticales (sobre la estructura de una forma de habla) y que se presumen como fundamentales a ellas. En este caso, de la distinción gramatical entre sujeto y predicado se deriva la distinción ontológica entre sustancia y propiedad, que se presume básica para la gramatical.

Lo que se infiere directamente de las tesis de separabilidad de mente y cuerpo es que 1) la mente es independiente del cuerpo, y 2) el cuerpo es independiente de la mente, es decir, lo que se infiere es solamente la independencia *relativa* de estas posibles sustancias. Una entidad independiente relativamente a otra puede existir en un mundo en que la otra no exista. La formulación de la definición de sustancia como algo que puede existir independientemente lleva a pensar que la noción de independencia de que se trata en ella debe ser absoluta. Una entidad independientemente en sentido absoluto debería poder ser la única cosa existente. Llegamos en este punto a dos problemas: primero, por qué, si la mente y el cuerpo existen, existen *ambos* en vez de uno solamente, y segundo, y esto es lo más importante aquí, por qué en el caso de la mente y el cuerpo la independencia relativa y la absoluta habrían de coincidir.

Ahora bien, *hay* una manera en que la independencia relativa y la independencia absoluta puedan coincidir: es el caso en que aquello de lo que algo es independiente es necesariamente la única otra cosa que podría existir.¹⁹ ¿Puede la ontología cartesiana sostener esta posibilidad?

A primera vista, pareciera que lo hace, *si uno se limita a considerar que la ontología cartesiana sólo habla de la res cogitans y la res extensa*. Pero resulta que en el esquema cartesiano ninguna de ellas es algo necesariamente existente. Cuando Descartes habla de ellas se refiere a algo meramente contingente. Esto se infiere, de hecho, de la aceptación de la posibilidad de que la mente exista en un mundo en que no existe algo extenso, y que lo extenso exista en un mundo en que no existe una mente (que es lo que se busca defender en el argumento de separación). Así, si algo *tiene* que existir, *tiene* que existir algo que sea independiente de la mente y el cuerpo. Es claro que el lugar de este tercer elemento en la ontología cartesiana es ocupado por Dios.

La introducción de Dios a la ontología cartesiana tiene varias consecuencias para la noción de sustancia, pero la más importante es que, así como sucede en el caso de la noción de posibilidad (en que hay que distinguir entre lo posible para nosotros y lo posible para Dios), el concepto adquiere un aspecto dual. Si la existencia del cuerpo y la mente son contingentes mientras que la de Dios es necesaria, la existencia de la mente y el cuerpo ha de *explicarse* de algún modo por la existencia de Dios (recuérdense los argumentos de la *Tercera Meditación*). La consecuencia de esto es que la mente y el cuerpo son dependientes de Dios. ¿En qué sentido puede decirse entonces que son sustancias?

Como puede verse en una carta de Descartes a Hyperaspistes, de 1641 (Descartes 1991), este problema no le pasó de alto, e incluso trató de resolverlo:

No hay duda de que si Dios retrayese su concurrencia, todo lo que ha creado iría inmediatamente a la nada; pues todas las cosas no eran nada hasta que Dios las creó y les prestó su concurrencia. Esto no significa que no haya que llamarlas sustancias, porque cuando llamamos a una sustancia creada auto-subsistente no descartamos la concu-

¹⁹No se podría ser independiente de *algo* sin ser independiente de *esa* cosa.

rrencia divina que necesita a fin de subsistir. Tan sólo queremos decir que es el tipo de cosa que puede existir sin otra cosa creada; y esto es algo que no puede decirse de los modos de las cosas, como la figura y el número.

Es decir, hay que usar el término sustancia de distinta manera cuando hablamos de Dios y cuando hablamos de su creación. En los *Principios de Filosofía* (Descartes 1995), esto se hace explícito:

Por sustancia, no podemos entender otra cosa que lo que existe de tal forma que no tiene necesidad sino de si mismo para existir. [En la explicación de la frase “no tiene necesidad sino de si mismo” puede haber oscuridad, pues] hablando propiamente, no existe sino Dios que lo sea de este modo, y no hay ninguna cosa creada que pueda existir un solo momento sin el concurso de Aquel. Por esto se dice con razón en las Escuelas que el nombre de sustancia no es unívoco respecto a Dios y a las criaturas, esto es, que no hay ninguna significación de esta palabra concebida distintamente, que convenga a Él y a ellas. (51)

Volviendo al problema que se había trazado antes: ¿Puede sostener la ontología cartesiana la posibilidad en que la independencia relativa y la absoluta coinciden? Es claro, directamente de la cita, que no. Sólo podría hacerse si se restringe más la condición requerida, relativizando la noción de independencia absoluta al contexto de lo contingente o creado. El argumento de separación no puede ser una prueba de la sustancialidad ni de la *res cogitans* ni la *res extensa*, salvo con esa restricción que convierte la conclusión del argumento en la tesis de que estas sustancias son distintas sustancias *contingentes*.

Descartes mismo esboza una prueba distinta de la sustancialidad de la *res cogitans* y la *res extensa*, que revierte a la noción proto-gramatical. Se reconoce, según Descartes, que se está frente a una sustancia porque reconocemos atributos, que tenemos que adscribir a algo. En realidad, Descartes apunta más bien a que para cada propiedad que reconocemos como realizada, tenemos que asumir que hay una cadena de relaciones de dependencia que termina en algo que la sostiene. El blanco que aparece en el campo visual es el efecto de cierto modo del ser extenso de algo, por ejemplo: es un atributo de un atributo, o una propiedad de orden superior. Desde él se llega a una sustancia sólo mediante una cadena de relaciones de dependencia entre propiedades. Como sea, la prueba de la sustancialidad de la *res cogitans* y la *res extensa* puede resumirse en el recorrido a través de cadenas como estas desde cualquier juicio que se haga sobre las propiedades de algo que satisfaga la propiedad de ser pensante o extenso.

En la ontología cartesiana, la categoría principal es la de sustancia. Una sustancia tiene que tener propiedades, que son dependientes de la sustancia. Ahora bien, una propiedad puede, a su vez, tener propiedades o modos. La cuadrilateralidad es un modo de tener figura, que es un modo de la extensión, que es propiedad de la *res extensa*, por ejemplo. Estas propiedades de segundo orden son

casos especiales de ciertas propiedades que Descartes llama principales. Una propiedad principal de una sustancia constituye su esencia, es algo que la sustancia no puede dejar de realizar (así, por ejemplo, la extensión es la propiedad principal de la *res extensa*). Con estas distinciones en mente, podemos ver que en el argumento de separación se muestra que la *extensión* y el *pensamiento* son propiedades independientes que, si son propiedades principales, tienen que corresponder a sustancias de distinta esencia. Así es como, indirectamente, su separabilidad prueba la sustancialidad de aquello en que se realizan las propiedades que reconocemos de las cosas, o más precisamente, que la categoría de sustancia no sólo no está vacía, sino que no tiene un único elemento.

¿Qué consecuencias trae que la extensión y el pensamiento sean propiedades esenciales de la *res extensa* y la *res cogitans*? La más importante es que la ontología cartesiana, a pesar de ser minimalista en el sentido de operar con un número limitado de categorías y conceptos, tiene un carácter revisionista si se la contrasta con lo que puede llamarse la ontología de “sentido común”. Según esta última, como se la va a concebir aquí, por ejemplo, puede decirse que mi mesa y mi lápiz son “sustancias” distintas. En la reconstrucción que me parece más plausible de la teoría cartesiana, estos ítemes no son sino modos del ser extenso, es decir, no son sustancialmente distintos. Lo mismo podría decirse de, por ejemplo, “la mente de Aristóteles” y “la mente de Platón”, en el caso de la *res cogitans*.

Algunos filósofos (Bradley, por ejemplo) han visto en este aspecto de las sustancias en la ontología cartesiana una manera de promover cierta forma de idealismo, proponiendo que el *cogito* prueba, no la existencia del yo en el sentido personal, sino la existencia del pensamiento en términos neutrales (“se piensa”). En efecto, el *cogito* entrega los elementos para la prueba de la existencia de la *res cogitans*, pero no la finaliza. Sólo junto al argumento de separación se tiene una manera de defender que la existencia de la *res cogitans* es independiente, es decir, que es una sustancia. Incluso en este punto no es posible decir que el *ego* del *cogito* sea sustancial, pero esto no implica que no pueda atribuírsele existencia, o que deba abandonarse la noción de sentido común de existencia, que puede aplicarse en el caso de cosas ordinarias como rocas, plantas, animales, personas, océanos, etc.

La metafísica cartesiana nos dice poco en realidad sobre cómo lidiar teóricamente con las cosas ordinarias una vez que se establece la existencia de la *res cogitans* y la *res extensa*. Descartes parece más interesado en lo que se puede determinar *a priori* a partir de su esencia mediante un método de variación (explorando mediante el método de la “concepción clara y distinta” las posibilidades de especialización de la extensión y el pensamiento) que sobre el estatus ontológico de las cosas ordinarias.²⁰

²⁰Sobre la idea de que la física cartesiana es una teoría de la *res extensa* en sus especializaciones reales (una teoría que desarrolla *a priori* las posibilidades físicas compatibles con los datos de la experiencia), véase Garber (1996). Allí también una defensa de la interpretación propuesta sobre el lugar de las cosas cotidianas en la metafísica cartesiana.

1.5.2. EL SOÑADOR CARTESIANO COMO SUJETO *posible*

En la primera meditación Descartes propone que hay un problema sobre la cuestión de si alguien puede decir que está despierto o soñando. La cuestión puede formularse así:

r) ¿Puede un sujeto que tiene las mismas experiencias que tenemos cuando decimos que estamos despiertos estar *realmente* durmiendo y soñando esas experiencias?

o bien:

s) ¿Es posible la condición S: un sujeto que tiene las mismas experiencias que tiene alguien que diría que está despierto en realidad duerme y sueña esas experiencias?

El problema pone en oposición dos conceptos de vigilia (y del sueño): el de alguien que diría que está despierto teniendo ciertas experiencias y el de alguien que no diría que está despierto teniendo esas mismas experiencias. El primer concepto es el cotidiano, y representa “lo que se pensaría habitualmente”. El segundo se presenta como una posibilidad. El objetivo de proponer el segundo concepto es debilitar la convicción en la aplicabilidad universal del concepto cotidiano: si la condición S es posible, entonces muchos de los casos (si no todos) en que se diría que se está despierto son incorrectos. Esta posibilidad lleva a la hipótesis escéptica sobre el mundo externo: no podría saber si los objetos que percibo en la vigilia son realmente externos en vez de meras imaginaciones. Uno de los objetivos centrales de las *Meditaciones* es obtener certeza de que esto no es el caso.

Hay dos estrategias fundamentales que se pueden seguir para ello: en primer lugar, se puede mostrar que la posibilidad del soñador cartesiano no es relevante para la cuestión de si se tenga conocimiento sobre el mundo externo, o que no refuta la posibilidad de tal conocimiento, y en segundo lugar, se puede mostrar que la situación propuesta es sólo aparentemente posible. La primera de estas estrategias concede la posibilidad del soñador cartesiano, mientras la segunda la pone en duda.

La estrategia que se ha empleado más comúnmente para atacar el problema que plantea la posibilidad del soñador cartesiano ha sido la primera. Es, por ejemplo, la estrategia que emplea Descartes (el sujeto cartesiano no se engaña en creer en la existencia del mundo externo pues Dios no es engañador). La influencia de esta estrategia puede verse en Stroud (1984), que consiste en un intento de mostrar que el problema del conocimiento externo no puede descartarse, y que expresa un problema filosófico genuino desde una serie de perspectivas. Para una revisión de los problemas epistemológicos planteada desde la idea de esta estrategia, véase Hendricks (2005).

¿Puede mostrarse que la posibilidad del soñador cartesiano es sólo aparente? ¿Qué implicaría el éxito de esa estrategia?

En primer lugar, se mostraría que la existencia del soñador cartesiano es imposible. Esto requiere de un concepto de posibilidad que impida que el soñador cartesiano pueda caer bajo él, y además, que este concepto de posibilidad reemplace todos los conceptos de posibilidad que sean aceptables

(aplicando la primera estrategia también podría mostrarse que el soñador cartesiano es imposible en *algún* sentido, pero aún así se aceptaría la intuición cartesiana de la posibilidad del escenario).

La cuestión de cómo podría desarrollarse una teoría de este tipo queda como una pregunta abierta. La mayor dificultad que presenta es cómo debe sincronizar las variadas intuiciones que se tienen sobre otros escenarios supuestamente posibles: si el soñador cartesiano fuera imposible, ¿qué otras cosas tendrían que ser rechazadas?

Kripke

Las preocupaciones modales abundan en los argumentos filosóficos contemporáneos, a menudo de manera explícita. Esta situación se debe en gran parte al desarrollo sistemático de la lógica modal, es decir, a un avance técnico. En el lenguaje natural la modalidad adopta formas extremadamente complicadas que hacen difícil una aproximación clara a los problemas teóricos que revierten a cuestiones de modalidad. La exigencia (es decir, la necesidad) de precisar el significado de las expresiones modales condujo a filósofos y lógicos al planteamiento de diversas alternativas conceptuales, que sin embargo no lograron expurgar a las nociones modales de todo aire problemático. Así, por ejemplo, resulta ser que la noción aristotélica de posibilidad no logra capturar varias ideas

“naturales” en cuanto a qué debería aceptarse sobre la noción de posibilidad.¹ Ahora bien, con el

¹El concepto habitual de posibilidad metafísica se describe a menudo como el formalizado por la lógica modal normal S5. La posibilidad aristotélica es dual, en el sentido que si algo es posible, es posible también su negación:

$$D) \diamond p \equiv \diamond \neg p$$

Se asume como base inferencial el sistema modal S5 más D. Así, se tendrán como tesis:

1. $\Box p \equiv \neg \diamond \neg p$ [def \Box]
2. $\Box p \rightarrow p$ [A5]
3. $p \rightarrow \diamond p$ [T1, derivable de A5]
4. $\diamond p \rightarrow \Box \diamond p$ [A8; si tenemos esto tenemos S5]
5. $\Box p \rightarrow \Box \Box p$ [A7, derivable de A8; si tenemos esto y no A8, tenemos S4]
6. $\diamond \diamond p \rightarrow \diamond p$ [T18, derivable de A7 o A8]
7. $\neg \neg p \equiv p$ [doble negación, d.n.]

En este sistema puede probarse la tesis correspondiente a que si digo que algo es posiblemente necesario, tengo que admitir que, en vista de lo que sé, no es necesario:

1. $\diamond \Box p \rightarrow \diamond \neg \Box p$ (D $p/\Box p$)
2. $\diamond \neg \Box p \rightarrow \diamond \diamond \neg p$ (def NEC, d.n.)
3. $\diamond \diamond \neg p \rightarrow \diamond \neg p$ (T18 $p/\neg p$)
4. $\diamond \neg p \rightarrow \neg \Box p$ (d.n., def \Box)
5. $\diamond \Box p \rightarrow \neg \Box p$ (1, 2, 3, 4, trans.)

Con esto podría proponerse que no se pueden hacer juicios de necesidad, porque, por T1, si $Lp \rightarrow MLp$, y por lo tanto si $Lp \rightarrow \neg Lp$, lo que es contradictorio. Se prueba así que S4 + D es inconsistente (por lo tanto, que S5 + D es inconsistente). Un resultado análogo puede obtenerse respecto a los juicios de posibilidad:

1. $\diamond p \rightarrow \diamond \diamond p$ (T1 $p/\diamond p$)
2. $\diamond \diamond p \rightarrow \diamond \neg \diamond p$ (D $p/\diamond p$)
3. $\diamond \neg \diamond p \rightarrow \diamond \neg \diamond \neg p$ (D)
4. $\diamond p \rightarrow \diamond \neg \diamond \neg p$ (1, 2, 3, trans.)
5. $\diamond p \rightarrow \diamond \neg \diamond p$ (T1, D)
6. $\diamond \neg \diamond p \rightarrow \neg \Box \diamond p$ (def \Box , d.n.)
7. $\diamond p \rightarrow \neg \diamond p$ (4, 5, 6, A8)

[Una característica interesante del sistema T1+D es que en él lo necesario no es posible:

1. $\diamond p \rightarrow \diamond \neg \diamond \neg p$ (\uparrow)
2. $\diamond p \rightarrow \diamond \Box p$ (1, def \Box)
3. $\diamond p \rightarrow \neg \Box p$ ($\diamond \Box p \rightarrow \neg \Box p$, 5)

Este resultado entra obviamente en conflicto con A5 y T1 más transitividad ($\Box p \rightarrow \diamond p$).]

desarrollo de la lógica formal en la primera mitad del siglo XX, las ideas básicas sobre la modalidad (que lo necesario es el caso, que lo que es el caso es posible, que lo posible no es imposible, que lo necesario es posible, etc.) resultaron ser compatibles con varios sistemas lógicos, y por lo tanto, no lo suficientemente específicas para explicar ningún conjunto de modalidades en particular; y la posibilidad de abandonar estas ideas impulsó la creación de nuevas lógicas modales (como ejemplo de investigaciones en este sentido están las de Łukasiewicz, von Wright y Arthur Prior). Por otro lado, ciertos problemas en la lógica clásica (en particular, las paradojas del condicional material) motivaron el estudio de la modalidad de expresiones que no habían sido tratadas previamente como modales de manera explícita (aquí, el pionero fue C.I. Lewis, con sus lógicas de la implicación estricta).² El avance de la semántica matematizada, impulsado por el trabajo de Tarski, hizo precisa una solución a la cuestión del significado de los términos formales que se encontraban en las nuevas lógicas modales; entre los que se dedicaron a la cuestión de la semántica de la modalidad están Carnap y el mismo Prior. Quien finalmente daría expresión³ a una semántica adecuada para la modalidad fue Saul Kripke.

La aparición de la así llamada “semántica de los mundos posibles” tuvo como efecto a largo plazo la habilitación de la discusión plenamente modal de una serie de problemas filosóficos, e instauró una jerga común para ella (puede encontrarse una reseña de las aplicaciones de esta jerga en una multitud de campos de estudio en la introducción de Divers (2002)). Kripke mismo ha sido uno de los filósofos que mayor provecho ha sacado de la estrategia de considerar el aspecto modal de los problemas filosóficos, aplicándola a las cuestiones de la referencia y la identidad, entre otras. Por su influencia, vale la pena detenerse en él y los aspectos de su trabajo que se relacionan estrechamente con el problema de la modalidad.

La estructura de este capítulo es la siguiente: primero, se va a exponer en qué consiste y cómo opera la semántica de los mundos posibles, lo que brinda una manera consistente de tratar con las discusiones modales (de la que se prefirió no disponer en la discusión sobre Descartes, dicho sea de paso, para evitar sobrecargarla de tecnicismos; al hablar de Kripke es necesario introducir algunos, que van desde la jerga más o menos informal de los “mundos posibles” a ciertas cuestiones precisas de lógica modal). Luego, se delineará la teoría kripkeana de los nombres y las clases naturales, desarrollada en su famoso *Naming and Necessity* (Kripke (1980), desde ahora NN), para entonces discutir la *actitud* kripkeana ante el problema epistemológico de los juicios de posibilidad. El ob-

²En la lógica proposicional clásica, el consecuente de una implicación se sigue lógicamente de cualquier antecedente falso. Por ejemplo, el condicional “la luna es de queso, así que las computadoras son inútiles” es válido, porque la luna no está hecha de queso. C.I. Lewis propuso reemplazar el condicional material clásico por lo que él llamó implicación estricta, en que las expresiones condicionales se analizan como “es necesario que si A, entonces B”. En este caso, el condicional que se tomaba por ejemplo sería inválido (es claramente posible que las computadoras sean útiles incluso si la luna estuviera hecha de queso).

³“Daría expresión” y no “inventaría”, porque la idea fundamental a su trabajo aparece en el trabajo de Carnap, Prior, Tarski... Véase Copeland (2002).

jetivo de este capítulo es, como en el anterior, presentar de qué manera se ramifican las cuestiones relacionadas con los juicios de posibilidad dentro del cuerpo de una filosofía, y de qué manera esto puede presentar problemas en ella.

2.1. MODALIDADES Y MUNDOS POSIBLES

La semántica kripkeana para la lógica modal (y como extensión, para el lenguaje modal en general) se basa en unas pocas ideas bastante simples, pero llenas de consecuencias. Primero, que la verdad de un enunciado puede evaluarse en relación con lo que puede denominarse “mundo posible”, que puede ser descrito fundamentalmente como un conjunto coherente de proposiciones. En términos cotidianos, un mundo posible es una manera en que podría ser el mundo; “una manera cómo podría ser el mundo” puede ser descrita por un conjunto de proposiciones coherentes. Enunciados sobre cosas que no necesariamente pertenecen al mundo real (que es el mundo posible que mejor describe la situación actual) deben pertenecer a otros mundos posibles; que pueda enunciarse (en el mundo real, se entiende) que tal o cual cosa sea posible depende de que el mundo real y el mundo posible en que lo propuesto como posible sea verdadero estén en una cierta relación (a menudo se llama a esta relación, relación de accesibilidad). La intuición esencial de Kripke es que proporcionar la semántica de una lógica modal consiste en dar las condiciones de satisfacción de una relación de accesibilidad, es decir, de la relación que permite que en un mundo posible pueda decirse qué cosas son el caso en otros mundos posibles.

Esta idea se presta de forma natural a cierta clase de visualización, lo que ayuda enormemente a su comprensión. Por ejemplo: imagínese una serie de vitrinas separadas que contienen un montón de esferas etiquetadas (una esfera dice *a*, otra *b*, etc.) Muchas etiquetas se repiten en varias de las vitrinas. Inicialmente, cada etiqueta está pintada o bien roja, amarilla o azul. Cuando alguien observa por una vitrina, el color de las etiquetas puede cambiar o no, en comparación a la disposición de las etiquetas en las otras vitrinas. Imagínese ahora que se descubre que las etiquetas son rojas si son rojas en todas las vitrinas, que son amarillas si las mismas etiquetas están pintadas amarillas en otras vitrinas y que son azules si no se repiten en otras vitrinas. El color de cada etiqueta en cada vitrina está determinado relativamente al color de las etiquetas en otras vitrinas; en este escenario, el color de las etiquetas está regulado por una lógica modal, que puede especificarse de manera precisa si se quisiera.⁴

⁴Resulta que en este caso la lógica que norma los colores de las etiquetas es una lógica modal anormal, porque en ella existen tres modalidades básicas (rojo, amarillo y azul) en lugar de las dos habituales (“posible”, “necesario”). Rojo es análogo a “necesario”, y la unión de amarillo y azul es equivalente a “posible”, pero amarillo y azul por separado no son equivalentes a ninguna modalidad normal (si amarillo es equivalente a “posible”, la relación de accesibilidad no es reflexiva (lo que es el caso no es posible, como consecuencia), y si azul es equivalente a “posible”, se reduce a la mera actualidad). Nótese que, a pesar de esta anormalidad, las reglas que rigen los colores de las etiquetas en este escenario no son incoherentes.

2.1.1. LA IDEA DE LA TEORÍA DE MODELOS

Para comprender cómo opera la teoría semántica de los mundos posibles es necesario revisar primero algunos conceptos básicos.

Siguiendo el análisis de Frege, una proposición puede descomponerse en una función y una n -upla de argumentos (un conjunto ordenado). La función enlaza los miembros de un conjunto “fuente”, que en este caso es el conjunto al que pertenece la n -upla y otro conjunto “objeto” que en este caso es un conjunto de posibles valores de verdad (según Frege (1974), la referencia de una proposición *es* un valor de verdad). De una función de este tipo pueden darse un conjunto de condiciones de satisfacción, que son las condiciones en que la función dados ciertos argumentos resulta en una proposición verdadera. Por ejemplo, la condiciones de satisfacción de la función NO son tales que si toma como argumento una proposición p , “NO p ” es verdadero si y solo si “ p ” es falso.

Supóngase que se tiene una clase de proposiciones (por ejemplo, la clase de las proposiciones que son negaciones), N . Se puede obtener una clase de funciones N^1 si reemplazan las constantes de las proposiciones por variables (por ejemplo, reemplazamos todas las formas proposicionales por variables: así, “no llueve” es “no p ”, etc.). A una secuencia de objetos que satisfaga cada una de las funciones en la clase N^1 se la denomina un *modelo* de la clase N . Por ejemplo, el conjunto de todas las proposiciones que no son verdaderas es un modelo de la clase de las negaciones.

¿Cómo puede usarse la noción de modelo? Un ejemplo clásico es el de Tarski (1956), en que se propone una definición modelo-teorética de la consecuencia lógica. Según ella,

the sentence X follows logically from the sentences of the class K if and only if every model of the class K is also a model of the sentence X .

Por ejemplo, de $(p \rightarrow q)$ y p se sigue lógicamente q porque todos los modelos de las premisas son modelos de q . Los modelos de estas fórmulas son tuplas de valores booleanos (verdadero, falso; 1, 0). Los modelos de $p \rightarrow q$ son (1, 1), (0, 1), (0, 0); los de p son (1, 0), (1, 1); los de q son (0, 1), (1, 1). La única tupla que satisface a todas las premisas (i.e., su modelo) es (1,1), que es un modelo de q ; se puede decir entonces según la definición de consecuencia lógica que q es consecuencia lógica de $(p \rightarrow q) \wedge p$.

Para terminar esta breve exposición de la teoría de modelos, cabe añadir que se va a llamar estructura de modelo a la definición de los símbolos no lógicos que componen un modelo (las constantes); o dicho de otra manera, los objetos que pueden satisfacer la interpretación de una proposición. Por ejemplo, la estructura de modelo del ejemplo anterior es el par 0, 1.

2.1.2. Consideraciones semánticas sobre la lógica modal (1963)

La semántica de los mundos posibles fue expuesta por Kripke por primera vez en (1963), en un artículo llamado “Semantical considerations on modal logic”. Esta exposición de la teoría sigue

más o menos la manera en que se propone allí.

Primero, se especifica un lenguaje que incorpora modalidades en la forma de un par de operadores lógicos, \diamond y \square . El primero representa la “posibilidad” y el segundo la “necesidad”. Habitualmente se considera válido que estos operadores son inter-definibles, de modo que $\diamond p = \neg \square \neg p$ (si algo es posible, no es necesario que no sea el caso) $\square p = \neg \diamond \neg p$ (si algo es necesario, no es posible que no sea el caso). Estos operadores representan funciones que forman proposiciones sobre la base de proposiciones: si se aplica \diamond a la proposición “existen animales con cuatro patas”, como ejemplo, se obtiene la proposición “es posible que existan animales con cuatro patas”. Los operadores modales pueden aplicarse cuantas veces se quiera.

En un lenguaje en que existen esos operadores pueden hacerse varias inferencias que involucran a los operadores modales; por ejemplo, la inferencia desde “es necesario que si existen personas, existen no-personas” a “si es necesario que existan personas, es necesario que existan no-personas”. Esta inferencia no puede hacerse válida sin considerar la modalidad de la necesidad (en el cálculo proposicional clásico, tiene el esquema $p \rightarrow (q \rightarrow r)$, que no es lógicamente válido). Estas inferencias pueden codificarse en sistemas lógicos modales. Las que siguen conforman una familia de ellos:

- M: el cálculo proposicional clásico + ...
 - $A_1 \square A \rightarrow A$
 - $A_2 \square(A \rightarrow B) \rightarrow (\square A \rightarrow \square B)$
 - $R_1 A, A \rightarrow B | B$
 - $R_2 A / \square A$
- S4: M + ...
 - $A_4 \square A \rightarrow \square \square A$
- B: M + ...
 - $A_B A \rightarrow \square \diamond A$
- S5: M + ...
 - $A_5 \diamond A \rightarrow \square \diamond A$

Para definir las proposiciones lógicamente verdaderas en estos sistemas se especifican las siguientes estructuras de modelo y los modelos.

Una estructura de modelo para un sistema modal es una 3-upla ordenada $\langle G, K, R \rangle$. K es el conjunto de todos los mundos posibles, G es un elemento de K que representa el mundo real. R es una relación refleja sobre K (es decir, entre elementos de K) y puede especificarse como la relación de “posibilidad relativa”. Lo que esto quiere decir es que si H_1 y H_2 son elementos de K , $H_1 R H_2$

significa que H_2 es posible relativamente a H_1 , y esto a su vez significa que todas las proposiciones verdaderas en H_2 son posibles en H_1 .

Si asumimos A_1 en M , R en la estructura-modelo de M tiene que ser reflexiva. Todo mundo es posible relativamente a sí mismo, pues toda proposición verdadera en él es, por fuerza, verdadera en él. Kripke asume esto como un requisito natural sobre los modelos de los sistemas modales—los otros requisitos establecen ‘reducciones’ para las modalidades. Si una modalidad se entiende como una secuencia de operadores modales (que representan una función que forma proposiciones sobre proposiciones), una reducción es la simplificación de una secuencia a otra que se establece como de idéntico valor. Por ejemplo, la modalidad $\Box\Box$ puede reducirse a \Box si se introduce el axioma A_4 , porque entonces $\Box\Box p = \Box p$. Así, en M hay infinitas modalidades no reducibles, en S_4 hay 16 y en S_5 sólo hay 4.

Si R es transitiva (i.e., si H_1RH_2 y H_2RH_3 son válidas, H_1RH_3 es válida) la estructura-modelo es del tipo S_4 . Si es simétrica (i.e, $H_1RH_2 = H_2RH_1$), es del tipo Browniano (B). Si es una relación de equivalencia (es decir, es reflexiva, transitiva y simétrica), es del tipo S_5 .

La terminología de los mundos posibles permite explicar porqué es así. Por ejemplo: R es transitiva en S_4 porque:

- 1) A_4
- 2) $\neg\Diamond\neg A \rightarrow \neg\Diamond\neg\neg\Diamond\neg A$ (def. \Box)
- 3) $\neg\Diamond A \rightarrow \neg\Diamond\Diamond A$ (d.n, $A| - A$)
- 4) $\Diamond A \vee \neg\Diamond\Diamond A$ (def. \rightarrow)
- 5) $\Diamond\Diamond A \rightarrow \Diamond A$ (sim. \vee , def \rightarrow)

Donde 5) tiene que entenderse como si (si en un mundo posible w_3 A es el caso, en un mundo posible w_2 $\Diamond A$ es el caso) y (si en un mundo posible w_2 $\Diamond A$ es el caso entonces en un mundo posible w_1 $\Diamond\Diamond A$ es el caso), $\Diamond A$ es el caso en w_1 . En la figura 2.1 hay un esquema de las relaciones de accesibilidad entre estos mundos posibles.

Un modelo asigna a cada fórmula atómica P un valor de verdad en cada mundo H que pertenezca a K . Un modelo ϕ sobre una estructura $\langle G, K, R \rangle$ es una función binaria $\phi(P, H)$.

La asignación de valores de verdad a fórmulas no-atómicas se define por inducción. Si se ha definido $\phi(A, H)$ y $\phi(B, H)$ para todos $H \in K$:

- a) si $\phi(A, H) = \phi(B, H) = 1$, $\phi(A \wedge B) = 1$. En otro caso, se asigna 0.
- b) $\phi(\neg A, H) = 0$ ssi $\phi(A, H) = 1$.
- c) $\phi(\Box A, H) = 1$ ssi $\phi(A, H') = 1$ para todo $H' \in K$ tal que HRH' . De otro modo, $\phi(\Box A, H) = 0$. O: A es necesario en H ssi A es verdadero en todos los mundos posibles relativamente (accesibles) a H .

Podría reemplazarse c) por

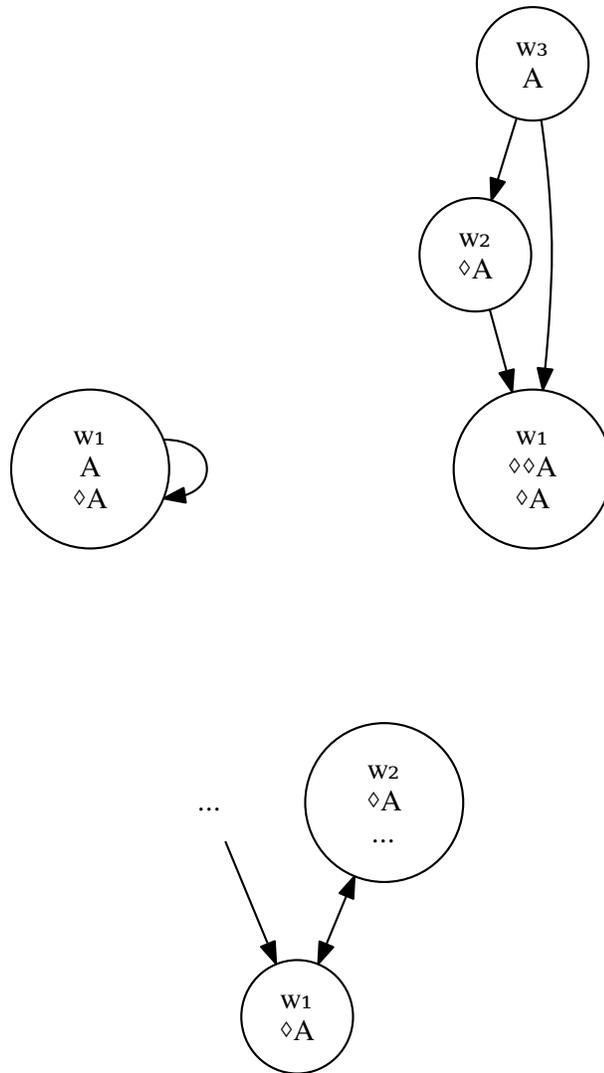


Figura 2.1: La relación de accesibilidad en M (arriba, izquierda), S4 (arriba, derecha) y B (abajo)

- d) $\phi(\Diamond A, H) = 1$ ssi $\phi(A, H') = 1$ para un $H' \in K$ tal que HRH' . De otro modo, $\phi(\Diamond A, H) = 0$.
 O: A es posible en H ssi A es verdadero en algún mundo posible relativamente (accesible) a H .

Siguiendo estas cláusulas se pueden hacer asignaciones de valores de verdad para todas las expresiones del lenguaje modal descrito anteriormente.

Un aspecto importante de la teoría es que la asignación de valores de verdad siempre se realiza en relación con un mundo posible. Entre todos los mundos posibles, el mundo real (actual) tiene un lugar particular, porque son los juicios de posibilidad verdaderos relativamente a este mundo los que interesa fundamentar epistémicamente más a menudo. En el contexto de la semántica de los mundos posibles, y el aparato conceptual que habilita, esto es equivalente a determinar cuál es la base epistémica de juicios sobre las propiedades de mundos posibles.

La idea de que hay series estratificadas de conceptos de posibilidad (por ejemplo, que hay una posibilidad física que es un subconjunto de la posibilidad metafísica, que es un subconjunto de la posibilidad lógica) y que hay posibilidades subjetivas (por ejemplo, lo posible de acuerdo a lo que un sujeto sabe) pueden integrarse con este modelo semántico mediante la definición del conjunto K en las estructuras de modelo. Por ejemplo, para construir una semántica de lo “nomológicamente posible” (esto es, lo posible según las leyes naturales), se construye un modelo en que K sea el conjunto de los mundos posibles en que las leyes naturales son verdaderas; para construir una semántica de lo “epistémicamente posible” (lo posible de acuerdo a lo que se sabe), un modelo en que K es el conjunto de los mundos posibles que pueden ser descritos solamente mediante proposiciones que los sujetos de tales o cuales características conocen. La estratificación de conceptos modales depende de las relaciones entre los conjuntos K de las estructuras de modelo (si son subconjuntos, si se intersectan, etc.).⁵

Sobre este marco conceptual, hay varias cuestiones de interpretación que vale la pena examinar, pero no antes de revisar cómo Kripke lo aplica en su discusión de varios problemas filosóficos.

2.2. LA TESIS DE LA RIGIDEZ DE LOS NOMBRES

2.2.1. EL PROBLEMA DEL SIGNIFICADO DE LOS NOMBRES

Un hecho elemental sobre el significado de las palabras es que algunas de ellas mantienen su significado en todas las situaciones en que se usan, mientras que otras varían en cuanto a su significado según el contexto. Los problemas en la filosofía del lenguaje, desde este punto de vista, derivan principalmente de este hecho.

La determinación del contexto de significado es un problema básico, debido al rol que cumple el contexto en la determinación del significado. Las opciones disponibles para resolverlo varían según

⁵Kratzer (2002) utiliza una semántica de este tipo para discutir las modalidades en el idioma alemán. Ella utiliza una terminología diferente, sin embargo: un concepto modal se define por un “trasfondo conversacional”, que cumple el rol de K , a grandes rasgos, y una “relación modal”, que cumpliría el rol de R (en este caso, determina la fuerza ilocutiva).

cómo se conciba la relación entre el contexto y el significado, y cubren un amplio rango. Por un lado se tiene la postura según la cual las palabras aisladas tienen significado independientemente del contexto (podríamos llamar a esta postura “monadismo”). Por otro, aquella según la cual la unidad del significado es el lenguaje completo a que pertenece la palabra (holismo). La primera postura fue abandonada en gran parte gracias a los argumentos fregeanos a favor de la composicionalidad (el significado de una expresión es una función de los miembros que la componen); en realidad ellos no son sino una clarificación de la intuición de sentido común de que no es necesario que una expresión sea significativa: por ejemplo, la palabra “pero” por sí sola no es significativa, y lo mismo sucede con muchas otras expresiones (quizás la única excepción son ciertas interjecciones, que podrían constituir su propio contexto—por ejemplo, alguien dice “ay!” cuando se electrocuta).⁶ Un buen diccionario dice qué significa una palabra mostrando qué puede querer decir alguien que la emplee (en algunos casos, incluso mostrando un ejemplo del uso de la palabra).

¿Cómo se aplican estas observaciones a la cuestión del significado de los nombres? En primer lugar, si la unidad de significado se desplaza desde la simple palabra a unidades más complejas como enunciados, el análisis del significado de las partes deberá hacer referencia a su contexto necesariamente. Un nombre es una voz vacía si no se entiende en un contexto. Nombrar algo o alguien es establecer o indicar la referencia de un término expresado en palabras. La relación de referencia conecta las palabras con objetos en el mundo; se dice que “las palabras escritas en el papel no tienen sentido” y con la expresión “las palabras escritas en el papel” se refiere a las palabras escritas en el papel. ¿Tiene la expresión necesariamente esa relación? ¿Se refiere con esas palabras a la misma cosa? Ciertamente, no. De hecho, aquello de lo que se habla ni siquiera necesita existir; y en efecto, en este caso no es claro en absoluto de que se hable de ciertas palabras existentes realmente (de hecho, es altamente dudoso).

Una solución posible a la dificultad es determinar que la unidad de significado es el enunciado y analizar el enunciado de modo que la expresión supuestamente referencial sirva solamente como indicador de una cierta estructura analítica. Esta es la estrategia de Russell (1905). Según él, un enunciado como

a) El rey de Francia es calvo

debe analizarse como

b) Existe un único x tal que x es rey de Francia y x es calvo.

⁶El contexto del significado de un término no necesita limitarse a lo que dice o haga su emisor. Considérese el diálogo: “– No hay piezas disponibles, pero...” – “¿pero...?” – “...¿Quizás pueda preguntarle a alguien si no le molestaría compartir su habitación?” (Nótese que el hecho de que el segundo “pero” esté incrustado en una pregunta modifica el análisis de su significado, que intuitivamente no es el mismo que el del primer “pero”, donde cumple el rol de una función lógica (usualmente, “pero” es una conjunción lógica, equivalente a “y”)).

donde x es una variable que puede tomar como valor un individuo. Una expresión como “El rey de Francia” se llama a menudo una descripción definida. Respecto a los nombres, se puede desarrollar una teoría (que vamos a llamar la teoría descriptivista) que propone que la referencia de los nombres se determina mediante descripciones definidas, y que su significado es el significado de las descripciones definidas que determinan esa referencia (hay que notar que estos aspectos de la teoría son separables).

Por ejemplo, según esta postura el significado de “Blaise Pascal” es el mismo que el de la descripción “el filósofo natural francés que defendió la posibilidad del vacío”. Si esta es efectivamente una consecuencia de la teoría, es evidente que hay que resolver varias dificultades. En primer lugar, no es claro que la descripción “el filósofo natural francés que defendió la posibilidad del vacío” aplique a un único individuo. Es necesario determinar que hay un contexto en el que se puede asumir que esa condición se satisface; una opción es decir que la descripción es incompleta (por ejemplo, que habría que añadirse que Blaise Pascal publicó su defensa de la posibilidad del vacío el año 1647 en un libro llamado *Experiences nouvelles touchant le vide*⁷), otra, que hay que incluir en el análisis las creencias del hablante sobre el objeto referido (así, que “Blaise Pascal” significa “la persona de quien creo que es verdad únicamente que es el filósofo natural que defendió la posibilidad del vacío”). Esta última línea de análisis conduce a la propuesta de determinar el significado de un nombre como la suma de las descripciones que se pueden dar del objeto nombrado (por ejemplo, el significado de “Blaise Pascal” sería la suma de descripciones “el filósofo natural”, “quien escribió *Experiences nouvelles touchant le vide*”, “quien defendió la posibilidad del vacío”, etc.).

Kripke (1980) describe una teoría de este tipo mediante las siguientes 6 tesis:

- 1) A cada nombre o expresión designativa ‘X’ le corresponde un racimo de propiedades, a saber, la familia de propiedades ϕ tal que A (el hablante) cree ‘ ϕX ’.
- 2) A cree que una de las propiedades, o algunas conjuntamente, seleccionan a un individuo de manera única.
- 3) Si la mayoría de las ϕ s, o una mayoría sopesada de ellas, es satisfecha por un objeto único y , y es el referente de ‘X’.
- 4) Si la selección de ϕ s no entrega ningún objeto, ‘X’ no refiere.
- 5) El enunciado ‘Si X existe, entonces X tiene la mayoría de las ϕ s’ es sabido *a priori* por el hablante.
- 6) El enunciado ‘Si X existe, entonces X tiene la mayoría de las ϕ s’ expresa una verdad necesaria (en el idiolecto del hablante).

¿Cómo funcionaría una teoría como esta? Imagínese que alguien pregunta si Gabriel Gonzalez Videla (GGV, para abreviar) es la persona que llamamos “Gabriel Gonzalez Videla” (“GGV”).

⁷Pero este análisis arriesga la circularidad: ¿cómo determinamos el significado de “*Experiences nouvelles touchant le vide*”?

Supongamos que la persona que llama “GGV” a GGV cree que “GGV fue un presidente de Chile”, “GGV fue senador”, “GGV nació en La Serena”. Si estas creencias son verdaderas sobre GGV, y si seleccionan a un único individuo⁸, se sigue que el referente de “GGV” es GGV.

2.2.2. EL ARGUMENTO MODAL CONTRA LA TEORÍA DESCRIPTIVISTA Y LA RIGIDEZ DE LOS NOMBRES

La observación elemental de Kripke, sobre la que se funda su crítica a la teoría descriptivista (tanto la versión de Russell como la de Strawson) del significado de los nombres es que la relación de designación debe acomodar casos contrafácticos, y en particular, debe soportar una serie de intuiciones sobre qué debería ser el caso en los escenarios que describen.

Lo que esto significa puede verse claramente mediante ejemplos como el siguiente. Supóngase que alguien identifica a Walter Gropius como el fundador de la Bauhaus. Ahora bien, considérese un mundo posible en que la Bauhaus fue fundada por un tal László Moholy-Nagy, y en que Walter Gropius fue un carpintero en Dessau.⁹ Es evidente que en ambos casos se puede decir que se habla del mismo individuo cuando se habla de Walter Gropius. Walter Gropius pudo no ser el fundador de la Bauhaus; luego, no pertenece al significado de “Walter Gropius” el identificar al fundador de la Bauhaus.

El punto crucial (o al menos, el que ahora interesa) del argumento de Kripke en contra de la teoría descriptivista es el rechazo de la tesis 6 antes mencionada:

- El enunciado ‘Si X existe, X tiene la mayoría de las propiedades que el hablante A cree que X satisface’ expresa una verdad necesaria (en el idiolecto del hablante)

donde X es un objeto y “X” es el nombre del objeto. Que, como en el ejemplo, sea posible que un individuo no satisfaga una propiedad que sí satisface en el mundo real es un contraejemplo directo de esta tesis. Si el enunciado fuera una verdad necesaria, X debería satisfacer las mismas propiedades en todos los mundos posibles.

¿Es esto así? Hay una dificultad respecto a esta lectura porque no se ha establecido cuál es la interpretación modal correcta de la tesis incrustada en 6:

0: Si X existe, X tiene la mayoría de las propiedades que el hablante A cree

que X satisface

Dado que se evalúa la necesidad de la verdad expresada en este enunciado, puede asumirse que se cuantifica sobre una variable w que tiene como rango la totalidad de mundos posibles, con un alcance de todo el enunciado.

⁸Las tres creencias que se listaron son un conjunto mínimo para seleccionar a un presidente de Chile, pues ningún otro presidente de Chile a la fecha satisface las tres conjuntamente. El único presidente de Chile nacido en La Serena, aparte de GGV, fue Bartolomé Blanche, pero él no fue senador (y sólo fue presidente provisionalmente, en 1932).

⁹Walter Gropius fue el fundador de la Bauhaus, célebre escuela de artes de corriente modernista. László Moholy-Nagy fue profesor de la escuela desde 1923. La Bauhaus operó en Dessau entre los años 1925 y 1932.

El primer punto de contención es cómo hay que interpretar “Si X existe”. Por un lado, podría querer significar

- i) Si X existe en a (donde a es el mundo real)... ,
- pero por otro lado, podría querer decir
- ii) Para todo mundo posible w , si X existe en w ...

El segundo punto de contención es el análisis de “X tiene la mayoría de las propiedades ϕ ”. Nuevamente, hay dos opciones:

- iii) X tiene la mayoría de las propiedades ϕ en a
- iv) X tiene la mayoría de las propiedades ϕ en w

Estos dos primeros puntos se coordinan, pues es natural pensar que el mundo posible que se evalúa en el segundo caso ha de ser el del primer caso; es decir, o bien i) + iii) o ii) + iv) son el análisis correcto. Si se adopta la primera opción, deberá explicarse cómo un hecho sobre a es verdadero sobre a en todos los mundos posibles (esto sólo podría ser posible si la relación de accesibilidad es de identidad, es decir, reflexiva, simétrica y transitiva, pues el mundo real debería ser accesible a todos los mundos posibles).

Un tercer punto de contención es el análisis de “el hablante A cree que X satisface las propiedades ϕ ”. Esto puede analizarse como

- a) En a , el hablante A cree que X satisface ϕ en a
- b) En a , el hablante A cree que X satisface ϕ en w
- c) En w , el hablante A cree que X satisface ϕ en a
- d) En w , el hablante A cree que X satisface ϕ en w

De estas opciones, sólo parecen viables a) y b), porque se asume que el hablante existe en el mundo real, en que se evalúa la necesidad del enunciado 0. Del mismo modo, a) parece ser el análisis correcto dada la asunción creíble de que las creencias de A se fijan respecto a las propiedades de X en a . Nótese que X podría ser meramente posible, y aún así tener un nombre, lo que da pie a otra manera de atacar esta formulación de la teoría descriptivista; para el argumento modal de Kripke esto no es importante porque incluso si eso se resolviera, podría ponerse en duda la necesidad del enunciado 0. Para resolver la dificultad introducimos r como el mundo posible sobre el que A establece el conjunto de propiedades que cree que X tiene que satisfacer en caso de que exista en un mundo posible (r podría ser a). Por ejemplo, piénsese que yo podría tener un hijo.

La tesis que Kripke ha de negar es, entonces,

- 0₁**: Para todo mundo posible w , si X existe en w , X tiene en w la mayoría de las propiedades que el hablante A cree en a que X satisface en r .

o lo que es lo mismo, ha de sostener que

- 1:** Existe un mundo posible w , en que X existe y donde X no tiene la mayoría de las propiedades que el hablante A cree en a que X satisface en r

Ejemplos como el que se presentó antes sobre Walter Gropius, o los que Kripke sugiere en NN, de que Aristóteles podría no haber sido el maestro de Alejandro Magno, o de que Hitler podría no haber sido el hombre que haya hecho matar al mayor número de judíos en la historia, son, efectivamente, apoyo para la tesis 1 (son situaciones que describen parte de mundos posibles en que se realiza la condición que 1 prescribe) y contraejemplos para 0_1 . Ahora bien, la plausibilidad de estos escenarios contrafácticos son intuitiva (en el sentido de que existe la disposición para creer que son en efecto posibles), pero aún podría dudarse de ellos. Este punto será discutido mas adelante.

Ahora bien, hay que apreciar un detalle sobre el argumento. El tipo de posibilidad y necesidad que pone en juego es metafísico, o al menos, no epistémico. Si la posibilidad fuera epistémica (por ejemplo, si se considerara lo posible como lo compatible con las creencias del hablante), el esquema 0 podría ser satisfecho, e incluso necesariamente (por ejemplo, si el hablante no pudiera llegar a creer que X es de otra manera que como lo concibe). En el modelo que presenta Kripke, “el lucero de la mañana” es necesariamente idéntico al “lucero de la noche” en un sentido metafísico, pero no en un sentido epistémico: es un hecho contingente que los hablantes crean en esa identidad.

¿Qué condiciones deben satisfacer los nombres para que su referencia se sostenga a través de mundos posibles?

La propuesta de Kripke es que es constitutivo (esto es, que no podría ser esencial) de los nombres ser lo que él llama *designadores rígidos*, esto es, deben designar al mismo objeto en todos los mundos posibles en que exista el objeto. Una manera alternativa de construir la noción de rigidez es decir que esta consiste en poder designar al mismo objeto en todos los mundos posibles en que sean accesibles mundos en que el objeto existe; de esta manera de explicita que se puede nombrar a un objeto en un mundo posible en que este no exista.¹⁰ Kripke observa

Si dices, “supón que Hitler no haya nacido nunca”, “Hitler” aún refiere rígidamente a algo que no existiría en la situación contrafáctica descrita" (p. 78, trad. mía)

Esta idea recoge la intuición de que podemos utilizar el mismo nombre para referirnos a un mismo individuo en distintos mundos posibles incluso si no comparte las mismas propiedades. Si decimos que “Aristóteles podría no haber sido el maestro de Alejandro Magno”, hablamos del mismo “Aristóteles” que fue el maestro de Alejandro Magno.

¹⁰ Aquí puede apreciarse un problema sobre el rol de la transitividad en la relación de accesibilidad que requiere el tipo de posibilidad que Kripke toma en cuenta al presentar la tesis de la rigidez. ¿Tiene sentido decir que se puede nombrar a un objeto X que no es posible en el mundo real, pero que es posible en un mundo posible accesible al mundo real? Por otra parte, ¿acaso no pueden “nombrarse” objetos imposibles (objetos que ni siquiera satisfacerían la condición recién propuesta)?

El concepto de rigidez permite distinguir de manera más clara a los nombres de las descripciones definidas. “El presidente de Chile entre el 18 de septiembre de 1886 y el 29 de agosto de 1891” no designa rígidamente a un único individuo: alguien más que José Manuel Balmaceda (el referente de la descripción en el mundo real) pudo haber sido el referente de la descripción.

2.3. LA RIGIDEZ DE LOS TÉRMINOS DE CLASE NATURAL

La tesis de la rigidez de los nombres tiene como consecuencia que los enunciados de identidad verdaderos entre nombres son necesarios. Es decir, juicios como

1) Aristóteles es Aristóteles

son necesarios, lo que es trivial, pero también juicios como

2) El lucero de la mañana es el lucero de la noche

han de serlo, lo que no es trivial. En realidad, lo que hace menos intuitiva la necesidad de los juicios de este segundo tipo es la interferencia de varias intuiciones epistémicas que se pueden tener sobre el caso. Es claro que el hecho que describe 2) es algo que puede haber sido descubierto, por lo tanto, es contingente que se tenga la creencia en 2).

Conceptualmente, esto da pie a la idea de que se puede conocer verdades necesarias *a posteriori*. La discusión sobre este asunto no es de interés aquí.¹¹

Kripke utiliza un argumento similar al que presenta para defender la rigidez de los nombres es para defender la idea de que los términos de clase natural son rígidos también. El objetivo de Kripke en este caso es defender la necesidad de identidades como

3) El agua es H₂O

4) El calor es la energía cinética molecular

5) El oro es el elemento que tiene el número atómico 79

Nuevamente, estas son verdades (si son verdades) que alguien puede descubrir, pero que (Kripke argumenta) deben ser necesarias. Nótese que para que estas identidades se mantengan los dos términos involucrados en cada caso deben ser rígidos.

El argumento de Kripke (como está planteado en NN, pág. 123 y ss.) depende de que se distinga entre propiedades esenciales y superficiales. Una propiedad esencial no puede no ser satisfecha por un objeto. Aristóteles pudo no haber sido el maestro de Alejandro (serlo es una propiedad superficial), pero debe tener alguna propiedad que tenga en todos los mundos posibles en que exista. La importancia de la propiedad de ser el maestro de Alejandro Magno, o de ser el creador de la

¹¹Por otro lado, Kripke propone que hay que considerar cierta clase especial de juicios como contingentes *a priori*. El caso más famoso es el del metro de París, sobre el cual Kripke adopta una postura que se contrapone firmemente a la de Wittgenstein (según quien el metro de París es un objeto del que *no se puede decir correctamente* que mide un metro).

teoría hilemórfica, para la fijación de la referencia del nombre (los hablantes pueden aprender a usar el nombre para referir a Aristóteles mediante la descripción) no es determinante en la cuestión de la esencia del objeto. Si se hace esta distinción, se puede apreciar qué es lo que Kripke quiere proponer cuando afirma que, por ejemplo, si se descubriera una sustancia que tuviera todas las propiedades con que se ha fijado inicialmente la referencia del término “agua” (que es incolora, inodora, que aparece en estado líquido, sólido y gaseoso en la naturaleza, etc.), pero que no tuviera la composición química H_2O , no sería agua. Los casos que Kripke examina (los juicios “los gatos son animales”, “el agua es H_2O ”, “la luz es una corriente de fotones”, “el calor es movimiento molecular”) siguen todos este mismo patrón. La estrategia de Kripke es mostrar que en los casos contrafácticos en que es el caso que los miembros de una clase de cosa que comparte las propiedades superficiales de las clases en cuestión las intuiciones de los lectores se alinean con la idea de que los términos de clase natural son rígidos. Así, en el caso de la luz

Originalmente identificamos la luz mediante las impresiones visuales características que puede producir en nosotros, que nos permiten ver... Imaginen una situación que los seres humanos fueran ciegos... ¿Sería una situación en que la luz no existiría? Me parece que no. Sería una situación en que nuestros ojos no serían sensibles a la luz... Sería una situación en que existiría luz, pero no podríamos verla. (pp. 129-130)

En el caso del calor:

El término “calor” no *significa* “aquello que le da a la gente esas sensaciones”. Pues en primer lugar, la gente podría no haber sido sensible al calor, y el calor aún habría existido. En segundo lugar, supongamos que los rayos luminosos, a causa de alguna diferencia en sus terminaciones nerviosas, *les diera* esas sensaciones. No sería el calor, sino la luz la que les daría a esa gente la sensación que llamamos la sensación del calor.

2.4. LA SEPARABILIDAD DE LA MENTE Y EL CUERPO, NUEVAMENTE

Las observaciones anteriores le permiten a Kripke desarrollar un ataque a la teoría de la identidad mente y cuerpo. Este ataque tiene dos aspectos. En primer lugar, ataca la idea de que la identidad entre la mente y el cuerpo es una cuestión meramente contingente. En segundo lugar, defiende la separabilidad de la mente y el cuerpo.

La idea de que la identidad entre la mente y el cuerpo (o más precisamente, entre estados mentales y estados cerebrales) es una consecuencia de que no se admita la posibilidad de lo necesario *a posteriori*. Quienes apoyan la teoría de la identidad habrían concebido juicios como “el dolor que siento es un estado de tipo C de mi cerebro” de manera análoga a juicios como “Aristóteles es el maestro de Alejandro Magno”. Como se ha visto, este tipo de enunciados es contingente en virtud de que uno de los designadores (“el maestro de Alejandro Magno”, en este caso) que lo constituyen no es rígido, mientras que para ser necesaria la identidad ambos deberían serlo. Este análisis

simplemente no es aplicable a las identidades que defiende la teoría de la identidad: supóngase que “Descartes” y “B” son los nombres de una determinada persona y su cuerpo. Si Descartes y B fueran idénticos (este es el enunciado que quiere defender el teórico de la identidad), esta identidad debería ser necesaria (cf. *NN*, pág. 145). Supóngase ahora que “A” nombra una sensación y “B” un estado cerebral. Si se quiere afirmar su identidad, esa identidad ha de ser necesaria (ibíd. pág. 146). Es más: si ser una sensación fuera una propiedad contingente de A, así como ser el maestro de Alejandro Magno es una propiedad contingente de Aristóteles, se cae en el absurdo de que una sensación podría no ser una sensación; lo mismo sucede con B y la propiedad de ser un estado cerebral. La misma observación se puede hacer respecto a la cuestión de si *tipos* mentales y físicos como el dolor y la activación de fibras-C pueden ponerse en una relación de identidad. Si es así, esta debería ser contingente, dado el supuesto de que los términos “dolor” y “activación de fibras-C” son rígidos.

Si esto es correcto, las intuiciones cartesianas de que:

- a) la mente puede existir sin el cuerpo
- b) el cuerpo puede existir sin la mente
- c) la presencia correlativa de algo con propiedades mentales es meramente contingente a los estados cerebrales
- d) la presencia correlativa de algo con propiedades físicas es meramente contingente a los estados mentales

deben ser rechazadas como ilusorias por el materialista.

¿Puede el materialista aplicar la estrategia que se puede usar para mostrar que, por ejemplo, la idea de que es posible que la luz no sea una corriente de fotones es ilusoria (la estrategia para mostrar que la contingencia de los casos examinados de lo necesario *a posteriori* son ilusorios)? En esos casos lo que llevaba al error era que la referencia de los términos se fijaba mediante descripciones de hechos que eran meramente contingentes sobre la referencia de los términos. Por ejemplo, la referencia del término luz podría haberse fijado mediante la descripción “aquello que produce ciertas sensaciones visuales”. La propiedad de producir ciertas sensaciones es contingente, de modo que se puede concebir una situación contrafáctica en que haya criaturas que, estando en la misma situación epistémica en que nos encontramos al determinar la referencia del término “luz” mediante la descripción “aquello que produce ciertas sensaciones visuales” sin embargo se referirían (incluso rígidamente) a algo que no es la luz (Kripke plantea como ejemplo que podría ser que el calor produzca en ellos las mismas sensaciones que tenemos gracias a la luz). Aplicando esta estrategia al juicio de que “el dolor es la activación de fibras-C”, por ejemplo, se supondría que el enunciado “aquello que se siente de tal o cual manera (dolorosamente) es la activación de fibras-C” expresa una verdad contingente, y que esto es lo que causa la ilusión de la contingencia del juicio de identidad.

Asúmase que la referencia del término dolor se fija mediante la sensación de dolor; si uno se limita a esa situación epistémica, es correcto decir que el dolor es “aquello que se siente de tal y tal manera” (es probable que esta descripción sea deíctica, y tome la forma “aquello que se siente así”). Kripke parece pensar que quien siga la estrategia antes esbozada debe asumir

i) Es posible que la estimulación de fibras-C exista sin que se sienta dolor

(esto es equivalente a decir que es contingente que la estimulación de fibras-C se presenta correlativamente a sensaciones de dolor). Pero parece intuitivo pensar que

ii) Si se siente dolor, existe dolor,

pues “estar en la situación epistémica de alguien que sentiría dolor es tener un dolor” (p. 152). De este modo, de a) se puede concluir que

iii) Es posible que la estimulación de fibras-C exista sin que exista dolor

lo que contradice limpiamente la tesis de la identidad. Como consecuencia, no se puede apelar a la estrategia descrita para calificar de ilusorio el juicio de la contingencia de la identidad entre el dolor y las fibras-C. Y si esto es así, se tiene un contraejemplo para la tesis de identidad general, que afirmaría que los tipos mentales son idénticos a tipos físicos.

Una manera alternativa de presentar el problema para quien defienda la tesis de identidad implica la noción de rigidez. La diferencia entre el caso de la identidad entre el calor y el movimiento molecular, y el dolor y la estimulación de fibras-C puede formularse como una diferencia en el tipo de propiedad que fija la referencia del término. En el caso del calor, este se puede identificar mediante la propiedad *contingente* de ser lo que produce la sensación de calor. En el caso del dolor, la propiedad con la que se identifica la referencia es el *ser un dolor* mismo, que es esencial al dolor. No hay manera de determinar la referencia a menos que sea mediante esta propiedad esencial.

Los argumentos kripkeanos reflejan la intuición de que no se puede comprender el significado de las expresiones que hablan del dolor a menos que se haya tenido un dolor; en esto puede radicar su atractivo. Pero de modo más significativo (en el sentido de desarrollar una propuesta sustantiva), estos argumentos expresan una creencia en que, en el caso de la relación entre el dolor y la activación de fibras-C (como ejemplo), es contingente, en cuanto la activación de fibras-C podría sentirse de otra manera. Pero, ¿no es esto dar por sentado la conclusión? Los argumentos contra la estrategia que mostraría la ilusión de la contingencia no muestran sino que la estrategia no es viable. ¿No es posible otra estrategia? ¿No muestra precisamente el punto de que la referencia del término dolor se fija mediante una propiedad esencial y no contingente que el caso no puede tratarse de la misma manera? Kripke tendría que mostrar que incluso en un caso en que la referencia de los designadores

en una identidad se fijara mediante una propiedad esencial, podría resultar que la relación fuera contingente.¹²

Kripke ilustra el punto de la contingencia entre la activación de las fibras-C y el dolor utilizando un experimento mental. Propone dos escenarios. En el primero, examina qué tendría que hacer Dios para hacer que se satisfaga la identidad entre el calor y el movimiento molecular. En este caso, Dios sólo tiene que hacer que las moléculas se muevan. Sin embargo, Dios tendría que hacer algo más para que el calor se sintiera como se lo siente: tendría que crear cierto tipo particular de seres conscientes (*sentient* en el original). Para estos seres, la verdad de que el calor es el movimiento molecular constituiría conocimiento *a posteriori*, a pesar de que es una verdad necesaria.

En el segundo caso, examina qué debería hacer Dios para que se sostenga la identidad entre la estimulación de fibras-C y el dolor. Parecería que Dios sólo tiene que crear seres que dispongan de fibras-C que *puedan* ser estimuladas de cierta manera. No parece relevante, según Kripke, que estos seres sean conscientes o no. Pero

It would seem... that to make the C-fiber stimulation correspond to pain, or be felt as pain, God must do something in addition to the mere creation of the C-fiber stimulation; He must let the creatures feel the C-fiber stimulation as *pain*, and not as a tickle, or as a warmth, or as nothing, as apparently would also be within His powers. If these things are in fact within His powers, the relation between the pain God creates and the stimulation of C-fibers cannot be identity. (pp. 153–154).

El punto problemático del experimento mental es que no es claro, de ninguna manera, que, si la identidad se satisface, esté “dentro de los poderes” de *nadie* hacer que la estimulación de fibras-C se sienta como un cosquilleo, calor, o como nada. El teórico de la identidad puede aceptar que la conclusión se sigue de las premisas, pero no asentir necesariamente a ellas. Sin embargo, Kripke propone que la contingencia es un elemento “obvio” (sus palabras, cf. pág. 154) de la correspondencia entre un estado cerebral y un estado mental; del mismo modo podría afirmar que es obvio que esas son realmente posibilidades. Por otro lado, quizás la disposición kripkeana a aceptar como posibles esos casos se deba a que la tesis de identidad entre fibras-C y el dolor es materialmente falsa; esto no obsta que pueda sostenerse otra tesis de identidad entre un estado físico (entendido en un sentido amplio) y un estado mental. Por consiguiente, el alcance del argumento de Kripke podría ser más limitado de lo que parece ser a primera vista; súmese a esto los problemas que se mencionaron antes. *Naming and Necessity* termina de hecho con la aceptación de ciertos límites del argumento.¹³

¹²Quizás por este problema la postura de Kripke no revierte en el dualismo sustancial cartesiano, que, como se examinó en el capítulo anterior, se puede derivar de la separabilidad entre mente y cuerpo (sin considerar los problemas que ellos presenta).

¹³Allí dice:

De cualquier modo, Kripke requiere de sustento epistémico para defender las posibilidades que propone aquí. ¿Dispone de él?

2.5. LA ACTITUD KRIPKEANA SOBRE LOS JUICIOS DE POSIBILIDAD

Como se ha visto, los argumentos de Kripke dependen en gran parte del avance de una serie de posibilidades.¹⁴ Las observaciones que acaban de hacerse sobre la manera en que se presentan en el caso del experimento mental de Dios y el dolor reflejan cierta actitud respecto a la cuestión del juicio de posibilidad. Por lo general, pareciera que Kripke considera que los juicios de este tipo fundamentalmente no tienen mayores problemas.

Para ser justos, la mayoría de los juicios que Kripke defiende no *parecen* ser problemáticos. Por ejemplo, el juicio de que

1) Nixon pudo no haber sido el presidente de Estados Unidos

no se cuestionaría a menos que se asuma una forma radical de determinismo (que los hechos no habrían podido ser distintos ni siquiera en otros mundos posibles, ya sea sobre este caso o de forma más general). Es probable, como el mismo Kripke observa (Kripke (1980), pág. 41), que muchas personas estén dispuestas a asentir a este enunciado:

... it is very far from being true that this idea [that a property can meaningfully be held to be essential or accidental to an object independently of its description] is a notion which has no intuitive content, which means nothing to the ordinary man. Suppose that someone said, pointing to Nixon, 'That's the guy who might have lost'. Someone else says 'Oh no, if you describe him as "Nixon", then he might have lost; but of course, describing him as the winner, then it is not true that he might have lost'. Now which one is being the philosopher, here, the unintuitive man? It seems to me obviously to be the second... When you ask whether it is necessary or contingent that Nixon won the

Materialism, I think, must hold that a physical description of the world is a *complete* description of it, that any mental facts are 'ontologically dependent' on physical facts in the straightforward sense of following from them by necessity. No identity theorist seems to me to have made a convincing argument against the intuitive view that this is not the case. (pág. 155)

El problema podría tomar un aspecto completamente distinto si se abandona la idea de que sólo puede darse *una* descripción completa del mundo. El problema metafísico que da pie a la disputa entre materialismo y anti-materialismo no sería una cuestión sustantiva, en el sentido de que la diferencia entre adoptar una y otra postura podría ser una cuestión meramente práctica, y no tocar en absoluto la cuestión de "cómo es verdaderamente el mundo" (que quizás no tenga solución; piénsese en la distinción kantiana entre númenos y fenómenos). Una forma de pragmatismo podría constituir una manera de resistir esta idea, pero parece que no carecería de problemas.

¹⁴Para una reseña independiente de la trama de *Naming and Necessity*, véase Putting (2009), págs. 31 y ss. Allí se defiende la tesis de que los pasos de Kripke no están fundamentados en argumentos, sino en intuiciones.

election, you are asking the intuitive question whether in some counterfactual situation, *this man* would in fact have lost the election.

Respecto a la garantía que se tiene para sostener la posibilidad de que Nixon podría no haber ganado las elecciones, el énfasis se pone en cierta intuición de que esa situación contrafáctica (llamaremos a esta situación “A”) es posible o que, al menos, es evidencia de que es posible. Aquí hay dos maneras de interpretar la cuestión. Por un lado, podría estar afirmándose que es intuitivo que A. Sin embargo, podría estar afirmándose en cambio que es intuitivo algo, “B”, que implica A o que es implicado por A. El candidato para B en este caso es que una propiedad pueda ser esencial o accidental a un objeto de manera independiente a su descripción; Kripke parece proponer que en este caso B se sigue de A.

La diferencia entre estas interpretaciones es importante, porque de ellas puede derivarse el modelo que siga la justificación de los juicios de posibilidad. Kripke podría estar asumiendo que el conocimiento de las posibilidades que propone es dado de forma inmediata como intuiciones, o bien que es derivado de alguna clase de tesis intuitivas (o al menos, de creencias justificadas), o bien, que ambas opciones son posibles. Si se abandona el requerimiento de que B (o una tesis que cumpla el mismo rol en el esquema) sea intuitivo, la segunda opción podría ayudar a Kripke con la crítica de que la intuición no es una fuente confiable de conocimiento o evidencia y que por lo tanto sus juicios de posibilidad son dudosos, ya que en este modelo no sería necesario que los juicios de posibilidad condujeran a intuiciones. Incluso sin abandonar el requerimiento de la intuitividad, es plausible que este modelo sea más estricto respecto a qué juicios de posibilidad puedan afirmarse; al menos, en el modelo que se deriva de la primera opción son aceptables juicios de posibilidad que expresen intuiciones que no tengan relación lógica con otros juicios, lo que no ocurre en el segundo modelo.

No parece haber evidencia decisiva en el texto de *Naming and Necessity* para determinar cuál de estos modelos sería el preferido por Kripke. Sin embargo, es claro que Kripke aceptaría que la intuición puede ser una fuente evidencia para los juicios de posibilidad. De hecho, dice:

Of course, some philosophers think that something’s having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don’t know, in a way, what more conclusive evidence once can have about anything, ultimately speaking. (pág. 42)

Esto indica que cuando propone, por ejemplo, que es posible que la estimulación de fibras-C pudiera sentirse como un cosquilleo, y dice que esto muestra la correspondencia entre un estado cerebral y un estado mental tienen un *obvio* elemento de contingencia, está siguiendo sus intuiciones modales, y *apelando* a las intuiciones modales de quienes reciben sus ideas. Asimismo, cuando en el *Prefacio* explica cuál es su postura, dice:

My main remark, then, is that we have direct intuition of the rigidity of names, exhibited in our understanding of the truth conditions of particular sentences. (pág. 14)

El paradigma de los enunciados de los que habla aquí es:

Lo que expresa el enunciado “A Aristóteles le gustaban los perros” pudo haber sido el caso. (cf. págs. 6, 13)

esto es, se trata de juicios de posibilidad.¹⁵

Sea como sea, ¿son confiables estas intuiciones? Hace falta una defensa de la confiabilidad de la intuición como fundamento para la evidencia.

La manera en que Kripke construye la noción de mundo posible es iluminadora para la comprensión de su actitud respecto a los juicios de posibilidad. Para clarificar la manera en que debería usarse el término “mundo posible”, utiliza una analogía. Supóngase, nos dice, que se lanzan dos dados de 6 caras. Para cada dado hay seis posibles resultados; para ambos juntos, treinta y seis. Ahora bien, nos dice Kripke, cada uno de los estados posibles de los dados es efectivamente un mundo posible “en miniatura”. Ahora bien, si se quisiera, se podría hacer una figura de cada uno de ellos. Si la noción de mundo posible se traslada en *escala* y se aplica a posibilidades como aquella en que la estimulación de fibras-C es acompañada de un cosquilleo y no de una sensación de dolor, es difícil ver cómo podría tenerse una figura de ese mundo posible; es difícil ver, si se lo compara con el caso de los dados, cómo podrían tenerse intuiciones sobre esta posibilidad. Kripke está pensando en esto tal vez cuando dice:

... the philosopher of ‘possible worlds’ must take care that his technical apparatus not push him to ask questions whose meaningfulness is not supported by our intuitions of possibility that gave the apparatus its point. (pág. 18)

Es decir, es plausible que haya cuestiones que vayan más allá de lo que se puede intuir como posible. Igualmente, el criterio final de si un mundo posible es aceptable son intuiciones. Un mundo posible, dice (pág. 44) “es dado por las condiciones descriptivas que asociamos a él”. La intuitividad de la existencia de un mundo posible en que ocurra “tal y tal cosa” depende de que la descripción de “tal y tal cosa” sea aceptable. Es significativo que Kripke pregunte en el caso de Nixon por qué no podría ser parte de la descripción de un mundo posible que contuviera a Nixon y que en ese mundo Nixon no fuera elegido como presidente. Lo que hace al plantear *esta* pregunta es volcar el peso de la prueba a quien dude de la posibilidad. “Podría haber una cuestión sobre si este mundo *es* posible”, acepta, “pero aquí parece, *prima facie*, que es claramente posible”. Ahora bien, ¿por qué es así? Aparentemente Kripke no está muy interesado en esa pregunta.

¹⁵En vista a los problemas que se han apuntado antes, ¿no es esto decir que ciertos juicios de posibilidad (los que expresan posibilidad *de re*, en particular) se justifican *mediante* nuestras intuiciones sobre la rigidez de los nombres?

Entendimiento y juicios de posibilidad

Ahora que se ha visto cómo opera el problema de la epistemología de los juicios de posibilidad en varios contextos filosóficos (se han mostrado las conexiones sistemáticas que las tesis modales pueden tener en una filosofía), voy a tratar de esbozar una manera de resolver las dificultades; en realidad, una manera global de ver el asunto que me parece interesante y que, creo, aporta positivamente a la discusión del problema de los juicios de posibilidad.

3.1. EL ASPECTO METAFILOSÓFICO

El objetivo de esta sección es esbozar con cierto detalle una postura metafilosófica que a) me parece interesante, aunque problemática y b) pone particularmente de relieve el problema de los juicios de posibilidad en filosofía.

En primer lugar, es necesario clarificar en qué sentido se habla aquí de “posturas metafilosóficas”, y qué consideraciones deben hacerse al hablar de ellas.

3.1.1. DISTINCIONES FUNDAMENTALES.

La metafilosofía cubre bajo sí todo estudio teórico que tiene a la *actividad* filosófica como objeto; así, por ejemplo, las cuestiones de demarcación sobre qué es la filosofía, en qué consiste el conocimiento filosófico, si lo hay, y cuáles son los métodos filosóficos, entre otras.

Existe la tendencia a denominar a esta clase de investigaciones con la etiqueta de “filosofía de la filosofía”, aunque no es necesario que ellas tengan una orientación filosófica. Por un lado, la filosofía puede estudiarse desde la perspectiva de varias ciencias naturales y sociales; en este sentido es claro (me parece, al menos) que la sociología, la psicología, la lingüística, la antropología y la historia pueden contribuir investigaciones metafilosóficas valiosas.¹ De hecho, deberían incentivarse investigaciones de este tipo, porque aunque un filósofo puede intentar explorar estas perspectivas, es posible que carezca del entrenamiento adecuado para ello, o que crea que ciertas ideas dentro de la disciplina que trata de explorar son correctas cuando en realidad no son consideradas de esa manera

¹Alvarez (2007) (U. de Melbourne) es un ejemplo que me gustaría destacar, a pesar de que no tiene resultados notables.

por los “expertos” dentro de esa disciplina; del mismo modo, deberían incentivarse investigaciones multidisciplinarias con el fin de enriquecer los enfoques sobre los problemas. Por otro lado, hay que considerar que son concebibles teorías metafilosóficas que nieguen la posibilidad del conocimiento filosófico. Para evitar acusarlas de constituir un contraejemplo a sí mismas (en el sentido de

que se seguiría de ellas mismas que no podría saberse si son verdaderas o no), podría considerarse que tales teorías no son filosóficas (dicho sea de paso, la cuestión de cómo resolver este tipo de problemas es un ejemplo interesante de problema metafilosófico).

No creo que sea prudente establecer límites demasiado rígidos sobre el campo *temático* de la metafilosofía; esto es, no han de restringirse las preguntas que puedan formularse como tema de las investigaciones metafilosóficas.

No obstante lo anterior, puede plantearse que la metafilosofía tiene como su centro focal dos preguntas:

- 1) ¿qué es la filosofía?
- 2) ¿qué debe ser la filosofía?

En lo que sigue voy a hablar de metafilosofía *descriptiva* cuando me refiera a una investigación de la primera cuestión o una teoría que responda a ella, y de metafilosofía *prescriptiva* cuando me refiera a una investigación de la segunda, o a una teoría que la responda.

A menudo se ve que las cuestiones descriptiva y prescriptiva se confunden; muchos filósofos construyen o simplemente adoptan una metafilosofía prescriptiva que luego aplican como respuesta a la cuestión descriptiva. Un ejemplo de esto puede encontrarse en el *Tractatus* de Wittgenstein (1974), cuando dice que

Philosophy aims at the logical clarification of thoughts.

Philosophy is not a body of doctrine, but an activity.

A philosophical work consists essentially of elucidations. Philosophy does not result in ‘philosophical propositions’, but rather in the clarification of propositions. Without philosophy thoughts are, as it were, cloudy and indistinct: its task is to make them clear and to give them sharp boundaries. (4.112)

Que la filosofía sea una actividad es una observación enteramente descriptiva (si es correcta o incorrecta no es importante aquí),² pero que el objeto de la filosofía sea la “aclaración lógica del pensamiento” depende de la tesis prescriptiva de que “la filosofía debe esclarecer y delimitar con precisión los pensamientos”, y por lo tanto no puede considerarse como un enunciado puramente

²Que “una” filosofía sea una actividad es un poco más difícil de evaluar. Por un lado, puede ser correcto decir que lo que hace quien filosofa es un cierto tipo de actividad, y que en ese sentido esa filosofía (la actividad filosófica de ese sujeto) es una actividad. Sin embargo, se puede querer hablar de “una” filosofía para referirse al producto de esa actividad, y en ese caso es problemático decir que es una actividad, en oposición a una teoría, si quien filosofa cree que lo que hace es teorizar. Aquí los problemas recaen en cómo ha de entenderse la noción de teoría.

descriptivo desde un punto de vista metafilosófico, aunque puede ser correcto en un sentido descriptivo decir que el resultado de la actividad de un filósofo en particular consiste en aclaraciones lógicas. La relativización de la postura descriptiva a la prescriptiva establece aquí un criterio demasiado estricto sobre, por ejemplo, en lo que consistiría una obra filosófica, que se enuncia como si fuera una tesis descriptiva cuando en realidad expresa una noción prescriptiva.

Este último punto está estrechamente relacionado con la cuestión de cómo se hacen juicios sobre el valor de una obra filosófica. Hay que distinguir el problema de si algo es filosofía de la pregunta de cuál es su valor filosófico. De otra manera, la teoría metafilosófica descriptiva resultará en falsos negativos. Así, por ejemplo, la obra de Hegel no satisface el criterio de Wittgenstein porque en gran parte no puede describirse como la mera exposición de elucidaciones lógicas (en parte porque la amplitud de los temas de los que habla en sus obras no permiten sostener esa idea), aunque quizás no haya que ir tan lejos como decir que no es filosofía—quizás lo correcto sea decir que hay parte de ellas que son mala filosofía (así como se puede decir que la metafilosofía de Wittgenstein está cruzada de confusiones). Sobre Hegel vale la pena mencionar las opiniones de Schopenhauer (1998), como ejemplo de la manera en que se hacen juicios sobre estas dos cuestiones. Nótese que también aquí se confunden:

I am convinced that Hegel not only lacks all philosophical merit, but has had on philosophy... and extremely pernicious, really stupefying, one might say pestilential, influence. (pág. 14)

... if... I were to say that the so-called philosophy of this fellow Hegel is a colossal piece of mystification which will yet provide posterity with an inexhaustible theme for laughter at our times, that it is a pseudo-philosophy paralyzing all mental powers, stifling all real thinking, and, by the most outrageous misuse of language, putting in its place the hollowest, most senseless, thoughtless, and, as is confirmed by its success, most stupefying verbiage, I should be quite right. (pág. 15)

Para considerar correctamente casos como este, hay que observar que los juicios sobre la calidad de una obra filosófica responden a criterios tanto internos como externos a las distintas metafilosofías que se aplican en cada caso. Esto lleva a problemas de demarcación (evidentemente, la cuestión de qué es la filosofía es un problema de demarcación; los aspectos descriptivo y prescriptivo son las superficies de tensión, por así decirlo, de este problema).

Una filosofía va emparejada siempre a una metafilosofía específica que la sostiene; en consecuencia, una filosofía puede estar en oposición a otra no sólo por cuestiones de contenido, sino también a causa de las diferencias en las metafilosofías que les corresponden. Por ejemplo, la consideración de que las contradicciones son señal de una falla en el razonamiento puede ser o no parte de una metafilosofía. De serlo, puede establecerse un criterio interno para la evaluación de la calidad de una obra filosófica: ninguna obra aceptable debe caer en contradicciones. ¿Es ese criterio aplica-

ble a filosofías que no asumen este criterio metafilosófico? Aquí se llega a un serio problema, pues quienes no acepten este criterio rechazarán su aplicación.³ Creo que hay criterios que no pueden relativizarse según la metafilosofía que se adopte; es decir, que hay que aceptar que haya metafilosofías incorrectas, así como también hay filosofías incorrectas. Es posible que el criterio que rechaza las contradicciones en una obra filosófica sea, en una formulación más refinada quizás que la que se ha esbozado, un *sine qua non* metafilosófico, por ejemplo (aunque no quiero defenderlo aquí). Sin embargo, si bien este criterio (o bien, esta clase supuesta de criterios) es de naturaleza prescriptiva, no se sigue que todas las tesis prescriptivas de una metafilosofía respondan a cuestiones en que se puede decidir que hay una opción correcta. Por ejemplo, y esta es una tesis que defenderé luego, no es una cuestión objetiva que una filosofía deba producir o consistir en conocimiento sobre el mundo real, aunque ese es un objetivo permisible para una filosofía.

En vista de estos problemas, creo que:

1) Sobre la cuestión descriptiva, hay que mantener un criterio liberal pero al mismo tiempo realista (en el sentido de que hay una única manera objetiva en que una metafilosofía descriptiva puede ser correcta) sobre qué sea la filosofía.

La combinación de estos requisitos sobre la cuestión descriptiva implica, me parece, la necesidad de un criterio “minimalista” en el sentido de “limitado en sus compromisos”, aunque no necesariamente en sus consecuencias. Por ejemplo, la tesis de que la filosofía es un tipo de actividad no tiene compromisos restrictivos, pero es rica en consecuencias que pueden tener cierto rendimiento metafilosófico (por ejemplo, que se aprecie que su realización es motivada, requiere esfuerzo, etc.). Es claro que si esta es una condición necesaria de satisfacción sobre qué sea la filosofía, son precisas tesis auxiliares que fijen precisamente qué tipo(s) de actividad sería(n) la(s) filosófica(s).

2) Sobre las cuestiones prescriptivas, hay un problema de demarcación entre cuestiones sobre las que hay que ser realista (por ejemplo, sobre ciertas condiciones de satisfacción mínimas sobre

³Como ilustración de este punto, obsérvese la siguiente cita, extraída de Adorno (2001, 81–82):

You may have had the experience of finding yourself in a discussion in which instead of focusing on the subject in question, people have tried to pin you down by arguing that you have contradicted yourself. . . Similarly, it will not have escaped your attention that there is a certain kind of inferior philosophical criticism that makes a virtue of attacking a writer like Nietzsche. . . on the grounds that they have been guilty of some so-called logical contradiction or other. This makes such critics feel good even though the reader does not have the sense that writers like Nietzsche are at all diminished in consequence. . . Such a justification must go in the direction of saying that our aim is *not* to juggle concepts, arranging and rearranging them as neatly as possible like a stamp collection, but to deploy concepts in order to bring the subject, whatever it may be, to life. The unending problem arises because what that subject and that life are is not something we have in our pockets. Nevertheless, philosophy directs its effort precisely towards the recuperation of what has been lost through this conceptual cleansing operation, this so-called contradiction-free, bland presentation of philosophical problems.

la expresión de las ideas filosóficas, como la cuestión de la claridad, la no-contradicción, etc.) y otras sobre las que hay que ser anti-realista (por ejemplo, cuál debe ser el fin de la actividad filosófica, y en general qué direcciones deben tomarse en ella, incluyendo cuestiones temáticas y metodológicas—por ejemplo, no parece incorrecto que la filosofía de las matemáticas sea tratada ni desde la perspectiva analítica ni desde una perspectiva fenomenológica).

Esta distinción puede explicarse observando que en el primer caso se trata de cuestiones sobre normas que no se aplican solamente en un contexto filosófico, mientras que en el segundo caso sí (por ejemplo, la búsqueda de claridad es una norma general que se aplica en todo tipo de contextos, con algunas excepciones particulares en contextos en que las normas pragmáticas requieren que se rompa el principio de claridad; por ejemplo, podría argumentarse que esto es así en el caso en que dos personas hablan de algo que no quieren que una tercera persona sepa frente a esa persona, aunque en ese caso el principio de claridad se mantiene entre quienes hablan y sólo se rompe respecto a la tercera persona). Aún si la línea se traza de esta manera parece posible que cuestiones que sólo se apliquen en contextos filosóficos puedan decidirse objetivamente. Voy a dejar este problema como una pregunta abierta.

En resumen, una postura metafilosófica consiste de dos elementos, el descriptivo y el prescriptivo. Las observaciones que quiero hacer ahora son fundamentalmente descriptivas. Se espera mostrar que a pesar del minimalismo de la postura que emerge de estas observaciones, resalta el rol del juicio de posibilidad en la actividad filosófica.

3.1.2. UNA POSTURA METAFILOSÓFICA

La postura descriptiva que quiero presentar tiene sus bases en tres elementos:

- 1) La filosofía *no* tiene como fin la obtención de conocimiento sobre el mundo real,
- 2) El producto de la actividad filosófica es la obtención de entendimiento,
- 3) La actividad filosófica consiste en la “exploración de espacios (o marcos) conceptuales”,

1) *La filosofía no tiene como fin la obtención de conocimiento sobre el mundo real.* Esta tesis debe entenderse con cuidado, porque puede analizarse en varios niveles, y cruza cuestiones tanto descriptivas como prescriptivas.

En el nivel más superficial, la tesis tiene que entenderse en un sentido enteramente descriptivo. A este nivel no se quiere decir con ella que una filosofía no debe intentar obtener conocimiento sobre el mundo real, sino que no es necesario que lo intente, en el sentido (trivial, me parece) de que una filosofía que no lo haga no deja por ello de ser filosofía.

En un segundo nivel es posible hacer una distinción entre el “fin” de la filosofía en el sentido de “propósito” y “fin” en el sentido de “producto”. Esta distinción se sustenta en la idea de que la filosofía es un tipo de actividad. Me parece que no se puede decir que el fin de la filosofía sea la obtención de conocimiento sobre el mundo real en ninguno de estos sentidos, aunque las cuestiones son separables (el producto efectivo de una actividad no es necesariamente idéntico a lo que se

intenta obtener con él, ni lo que lo motiva). Por un lado, porque (como queda a la vista en la cita de Wittgenstein que se examinaba antes, pero también si se recuerda a los filósofos pirrónicos, por ejemplo) hay filósofos que explícitamente rechazan la idea de que la filosofía tenga o siquiera pueda producir conocimiento. Por otro, porque si se requiere que el conocimiento consista en creencias *verdaderas*, es claro que mucho de lo que se ha propuesto como “conocimiento filosófico” históricamente no puede serlo, y que por lo tanto el producto de toda aquella actividad no pudo haber sido conocimiento (la pregunta es, por supuesto, cuál fue el producto de toda esa actividad, a la que volveré luego.)

Aún hechas estas precisiones, hay un tercer nivel en que se puede analizar la cuestión. La pregunta ahora es si la tesis debe tomarse o no como compatible con que el juicio de que una obra filosófica que no intente obtener conocimiento sobre el mundo real tiene menor valor filosófico que una que sí lo intente. La postura que se adopte en este punto resultará en metafilosofías prescriptivas diferentes: que el valor supuesto de una filosofía sea mayor o menor a causa de un factor x es una razón (o al menos parece serlo) para juzgar que la actividad filosófica debería incorporar o no ese factor. Sobre este punto creo que una obra filosófica que no pretenda entregar conocimiento sobre el mundo real no tiene *por ello* menos valor que una que sí lo pretenda: es posible que sea “mejor” que una que tenga esa pretensión sin ningún fundamento para esperar entregarlo. Esto no es sino decir que, sea lo que sea en que pueda consistir el valor de una obra filosófica, no es necesariamente relevante para ese valor si su autor o autores pretenden que ofrezca conocimiento sobre el mundo real.

¿Es esta una cuestión que pueda decidirse objetivamente, sin embargo? ¿Es esta una cuestión prescriptiva del primer o el segundo tipo—es decir, depende de cuestiones normativas extra-filosóficas que pueden ser decididas objetivamente (así como puede decidirse que un argumento es correcto, o que un texto es claro) o es una cuestión intra-filosófica, relativa, dependiente del contexto? Un argumento que puede presentarse para defender lo primero es el siguiente: el juicio de que una filosofía que no intente obtener conocimiento del mundo real tiene menor valor que una que lo haga tiene que estar errado *porque* los objetivos de una filosofía en particular *no* son objetivamente cuestionables. ¿Pero por qué habría que asumir eso? La disposición a creerlo puede cimentarse en un razonamiento análogo al que se hace cuando se juzga que la profesión de alguien no cambia su valor como persona. No estoy seguro de que esta sea una razón sólida para sostener esta creencia. La observación de que es posible que una obra que no pretenda entregar conocimiento sobre el mundo real sea mejor que una que sí lo pretenda sin éxito puede brindar un argumento más apropiado, pues para explicar esta intuición quizás se pueda apelar a cuestiones extra-filosóficas que puedan decidirse objetivamente (como el que no cumplir con las expectativas que se han aceptado sobre uno es un tipo de fallo).

Difiero fundamentalmente, por lo tanto, de autores como Unger (2012), quienes creen que es un defecto de gran parte de la filosofía contemporánea el que no ofrezca tesis sustantivas sobre la

realidad concreta.⁴

Ciertamente, el producto de la actividad filosófica *puede* ser conocimiento sobre el mundo real (simplemente creo, por las razones que he dado, que no *tiene* que serlo). Hay al menos un tipo de conocimiento sobre el mundo real que puede ser el producto de la actividad filosófica: conocimiento acerca de las creencias de los sujetos pensantes que habitan en él. Ahora bien, ¿cuáles de esos sujetos pensantes? Cuando un filósofo habla acerca de las intuiciones de sentido común, ¿está proyectando sus propias creencias o describe las creencias de sujetos como él? Las investigaciones propuestas por los adherentes a la filosofía experimental sugieren que incluso sobre cuestiones que a primera vista parecen universales puede haber variaciones.⁵ Esto es, incluso en casos en que es plausible describir el producto de la actividad filosófica como conocimiento sobre algún aspecto del mundo real, es posible que se interprete incorrectamente en qué consiste ese conocimiento. De cualquier modo, es implausible que pueda decirse de todas las investigaciones filosóficas que ofrecen conocimiento sobre algún aspecto del mundo real de este tipo.

¿Puede la filosofía ofrecer otro tipo de conocimiento sobre el mundo real? Me parece que algunas de las preguntas filosóficas sobre el mundo real (por ejemplo, la pregunta de qué justifica nuestros juicios de posibilidad, que es una pregunta sobre la justificación que ellos tienen en el mundo real) deben tener una respuesta definitiva, y apuntan a cuestiones objetivas sobre el mundo real. Lo que no es seguro es si alguna de las soluciones ofrecidas a los problemas es esa respuesta.

2) *El producto de la actividad filosófica es la obtención de entendimiento.* Si la obtención de conocimiento sobre el mundo real no es necesariamente ni el propósito ni el producto de la actividad filosófica, y si es problemático decir que la filosofía tiene un único propósito, ¿cuál es el producto de la actividad filosófica?

¿Basta con eliminar la restricción de que el producto de la actividad filosófica tiene que ser conocimiento *sobre el mundo real*? ¿No puede ser el caso que la filosofía produzca conocimiento sobre realidades abstractas, así como lo hacen las matemáticas? Esta alternativa no es satisfactoria porque, como se dijo antes, la filosofía parece poder producir conocimiento sobre el mundo real. Tampoco parece razonable proponer que el producto de la filosofía sea o bien conocimiento sobre el mundo real o conocimiento sobre objetos abstractos: si fuera así, ¿qué distinguiría el concepto de filosofía del de ciencia en general? Otro problema, quizás más importante, es que es plausible que haya filosofías que no pretendan producir conocimiento en absoluto; el ejemplo de los filósofos pirrónicos es el más claro.

⁴Estoy de acuerdo con Unger, no obstante, en su diagnóstico de que gran parte de las tesis que se ofrecen en metafísica y filosofía de lenguaje no son realmente tesis substantivas sobre la realidad concreta. Por ejemplo, como él observa (1, págs. 9–13), la disputa de si aparte de los objetos materiales hay *arreglos* de esos objetos no las ofrece.

⁵La más notable de estas investigaciones ha sido la de Machery (2004), que sugiere que las intuiciones de sujetos asiáticos no soportan la teoría causal de la referencia. El estudio de Machery *et al* ha sido cuestionado por, entre otros, Devitt (2011) y Ichikawa (2011). Machery (próximo) responde a los críticos.

¿Si el producto de la actividad filosófica no es necesariamente conocimiento, qué es? Lo que se obtiene mediante la actividad filosófica es entendimiento. Más adelante voy a caracterizar más completamente a qué me refiero con esto, pero es necesario hacer algunas precisiones mínimas sobre ello.

La distinción entre conocimiento y entendimiento es intuitiva. Se puede conocer algo sin entenderlo, aunque no es posible entender algo sin conocerlo.⁶ Una persona que se enfrenta a una cultura distinta puede llegar a saber o estar al tanto que los miembros de esa cultura realizan ciertas acciones sin necesariamente entenderlas. Por ejemplo, que las mujeres mayores de 10 años se manchan la palma de las manos con tiza todos los quintos días del mes al amanecer (el ejemplo es imaginario). Saber eso no implica que entienda lo que hacen; aún puede preguntar por qué.

La relación entre el conocimiento y el entendimiento no es simple. Considérese otro ejemplo: pareciera que alguien puede saber que una palabra puede usarse en un determinado contexto sin entenderla. Por ejemplo, los niños a menudo dicen groserías en contextos *apropiados* sin entenderlas. ¿No entienden las groserías o simplemente no saben todo lo que hay que saber sobre ellas? ¿No podría decirse que entender algo implica saber “lo que hay que saber” sobre lo que se entiende, donde “lo que hay que saber” debe construirse como una noción prescriptiva sobre la corrección del uso de los conceptos en ciertos contextos? Este conocimiento podría ser implícito. Volveré al problema de cómo debe caracterizarse el entendimiento más adelante.⁷

La actividad filosófica produce esencialmente entendimiento, si no conocimiento. Una teoría filosófica pretende unificar una serie de fenómenos y conceptos en un esquema de significado cuyo valor va más allá del que se puede esperar del mero conocimiento de sus partes (aunque ha de ser una función de este).

Una ventaja que tiene la caracterización del producto de la actividad filosófica como entendimiento, en vez de conocimiento, es que permite acomodar casos en que no es claro que las teorías filosóficas sean correctas: se puede entender algo que no es verdadero. Tómese el caso de la monadología de Leibniz; supóngase que es una teoría falsa sobre la naturaleza de la realidad. ¿Qué nos puede entregar? En primer lugar, se podría extraer de ella conocimiento sobre una manera de ver cómo podría explicarse la interacción entre lo físico y mental, la naturaleza de la causalidad, etc.

⁶¿No constituye, por ejemplo, el problema de la conciencia un caso en que se entiende algo, pero no se sabe qué es? Me parece que aquí podría haber una confusión: podría ser que se sabe a grandes rasgos qué es, pero que no se lo entienda ni se sepa que no se lo entienda (en cuyo caso podría haber una ilusión de entendimiento) o también que ni se sepa qué es ni se lo entienda (en cuyo caso también habría una ilusión de entendimiento). Hay otras opciones: puede ser que para entender algo no sea necesario saber qué es, pero sí que sea necesario que se sepan otras cosas (por ejemplo, cuál es su estructura, y cómo es que existe). Para precisar la intuición de la separabilidad de conocimiento y entendimiento es necesario introducir algunas distinciones adicionales—así como está planteada aquí, es ambigua. En la siguiente sección se examinará con más detalle el análisis de la noción de entendimiento y su relación con el conocimiento.

⁷Creo que puede haber una confusión entre ‘entender que ...’ (entender que algo es así o asa, o que podría serlo) y ‘entender “...”’ (entender un concepto o un término). *Prima facie* puede ser que respondan al mismo proceso intelectual, pero no me parece claro cómo debería analizarse la relación entre enunciados de estos tipos.

(Desde aquí se puede avistar el rol que cumplen los juicios de posibilidad en filosofía. ¿Es la descripción del mundo que presenta la monadología falsa? Como se dijo antes, la suponemos falsa. ¿Es posible? (esto es, ¿describe un mundo posible?), y ¿en qué sentido? ¿Se gana conocimiento modal entendiendo la monadología? ¿De qué tipo? La filosofía podría producir sobre todo conocimiento de posibilidades, y en efecto podría *consistir* en explorar distintos “espacios” de posibilidades; sobre esto volveré luego.) Por otro lado, en contraste a otras filosofías, podría permitir comprender los problemas relacionados con las cuestiones que intenta explicar (por ejemplo, si se compara la monadología leibniziana con el monismo de Spinoza y el dualismo cartesiano se pueden apreciar varios problemas sobre la noción de sustancia, y qué debería esperarse de una teoría sobre ella). A este tipo de entendimiento se le pueden dar nuevos propósitos sobre distintas cuestiones—así como pueden ofrecerse hipótesis sobre problemas empleando razonamientos analógicos.

Esta caracterización preliminar del entendimiento filosófico debe complementarse con una descripción de los mecanismos epistémicos que son su fuente. La propuesta que se hará luego sobre el juicio de posibilidad intentará proporcionar parte de esa descripción.

3) *La actividad filosófica consiste en la exploración de espacios conceptuales.* ¿Cómo se obtiene, entonces, entendimiento mediante la actividad filosófica? Tómese la pregunta desde el ángulo de cómo opera la filosofía que la distingue de las matemáticas, las ciencias naturales o las ciencias sociales. La pregunta va al caso porque la ampliación del entendimiento no es una característica distintiva de la actividad filosófica. De hecho, puede verse que en todas esas disciplinas ocurre que se ofrecen explicaciones que añaden “sentido”, si se quiere, al conocimiento que se recava. Por supuesto, existe la tentación de decir que esas explicaciones son el producto de actividad filosófica, pero esto es poco caritativo con el verdadero alcance de esas disciplinas: ellas producen entendimiento por sí solas. Lo que sí puede especularse es que lo que permite que en ellas se produzca entendimiento es similar a lo que lo permite en filosofía; el entendimiento puede proyectarse como una facultad general que puede aplicarse en distintos dominios.

Aún así, es plausible que la filosofía tenga una relación más estrecha con el entendimiento que otras disciplinas. Si el producto esencial de la actividad filosófica es entendimiento, como se ha propuesto, esto ha de deberse a alguna razón intrínseca.

Una explicación plausible de esto es que el objeto de investigación de la filosofía tiene un vínculo estructural con el entendimiento (o bien *es* entendimiento, o su análisis depende de la aplicación del concepto de entendimiento). Por ejemplo, podría ser que lo que un filósofo estudie, diseñe y proponga sean, dicho informalmente, formas de entender las cuestiones que le preocupan (la génesis del concepto de naturaleza, el análisis de la noción de justicia, etc.).

¿Qué es una “forma de entender” algo? La expresión puede usarse para referirse a o bien el proceso mediante el cual algo se entiende (cómo fue el caso que se entendió que...) o bien el contenido

del entendimiento eventual⁸ de un sujeto sobre algo (lo que entiende sobre...). Aquí interesa este segundo sentido; adoptar el primero significaría poner demasiado énfasis en las cuestiones epistemológicas: si bien el análisis de las “formas de entender” en ese sentido es parte de la actividad filosófica, no puede decirse que esté presente en ella toda.

¿Qué contiene (en qué consiste) una “forma de entender” algo? Me parece que debe incorporar la malla de creencias que el sujeto tiene sobre ese “algo” y junto a ellas una estructura que las reúna (estas creencias no deben estar simplemente juntas, sino que debe haber algo que las reúna). Este segundo elemento es lo que (me parece) permite que se diga que se entiende algo y no meramente que se lo conoce o reconoce. Para identificarlo en cada caso, sin embargo, es necesario examinar las relaciones sistemáticas entre las creencias del sujeto—mi propuesta consiste en que gran parte de la actividad filosófica consiste precisamente en eso. La suma de un conjunto de creencias y un elemento que las reúna puede denominarse “espacio conceptual”.

La expresión “espacio conceptual” es útil porque permite tener una imagen intuitiva de la clase de operaciones que se pueden realizar sobre las estructuras conceptuales de que se habla. Así, por ejemplo, Ryle (2009) dice

It is, however, one thing to know how to apply such concepts, quite another to know how to correlate them with one another and with concepts of other sorts. Many people can talk sense with concepts but cannot talk sense about them; they know by practice how to operate with concepts, anyhow inside familiar fields, but they cannot state the logical regulations governing their use. They are like people who know their way about their own parish, but cannot construct or read a map of it, much less a map of the region or continent in which their parish lies.

For certain purposes it is necessary to determine the logical crossbearings of the concepts which we know quite well how to apply. The attempt to perform this operation upon the concepts of the powers, operations and states of minds has always been a big part of the task of philosophers. Theories of knowledge, logic, ethics, political theory and aesthetics are the products of their inquiries in this field. Some of these inquiries have made considerable regional progress, but it is part of the thesis of this book that during the three centuries of the epoch of natural science the logical categories in terms of which the concepts of mental powers and operations have been co-ordinated have been wrongly selected. Descartes left as one of his main philosophical legacies a myth which continues to distort the continental geography of the subject.

A myth is, of course, not a fairy story. It is the presentation of facts belonging to one category in the idioms appropriate to another. To explode a myth is accordingly not to deny the facts but to re-allocate them. . .

⁸Un mismo sujeto puede entender eventualmente de varias maneras una misma cosa (por ejemplo, si tiene explicaciones paralelas para ella sobre las que no se ha decidido, o cambia de parecer sobre un asunto).

To determine the logical geography of concepts is to reveal the logic of the propositions in which they are wielded, that is to say, to show with what other propositions they are consistent and inconsistent, what propositions follow from them and from what propositions they follow. The logical type or category to which a concept belongs is the set of ways in which it is logically legitimate to operate with it. (págs. lx–lxi)

El modelo metafilosófico de Ryle es consistente con el que se presenta aquí.⁹

La construcción de teorías filosóficas puede formularse en este modelo como la aplicación de entendimiento obtenido mediante el análisis de algún concepto en la construcción de mallas de conceptos. El entendimiento en este sentido funciona como un principio generativo para las teorías filosóficas. Se toma una idea y se la hace fértil (por ejemplo, la idea de que el entendimiento es central para el análisis de la actividad filosófica resulta en el modelo metafilosófico que se propone aquí).

La imagen que se ha dado de la filosofía en las páginas anteriores no es sino una figura panorámica, con los problemas que eso implica (es en partes ambigua, en algunas partes hacen falta argumentos más sólidos, etc.). Si es una imagen correcta, sin embargo, su propuesta puede entregar precisamente lo que uno puede esperar de una filosofía: entendimiento (en algún grado) sobre una cuestión problemática.

3.2. EL ENTENDIMIENTO Y EL JUICIO DE POSIBILIDAD

En esta sección exploraré las relaciones entre el entendimiento y los juicios de posibilidad. Como se indicó en la introducción, esta es una de las alternativas abiertas sobre el problema de la epistemología de los juicios de posibilidad: la justificación de estos juicios puede explicarse mediante nuestra comprensión conceptual.

3.2.1. CARACTERIZANDO EL ENTENDIMIENTO

Las teorías clásicas sobre el entendimiento están ligadas estrechamente al concepto de explicación. Esto es natural, porque las relaciones entre ambos conceptos son sistemáticas. Como Lipton (2004) sugiere, el entendimiento parece ser el bien intrínseco de las explicaciones: se dice que una explicación es intrínsecamente valiosa porque provee entendimiento.¹⁰ Cuando alguien busca una explicación de un evento o proceso busca algo más que el mero conocimiento *de que* sucedió o se llevó a cabo (es altamente plausible que quien pregunte por qué sucedió el suceso ya sepa que su-

⁹También puede mencionarse en este punto a John Wisdom, quien propone una visión similar en “Philosophical Perplexity” (1936). En ese ensayo, propone explícitamente que el propósito de la actividad filosófica es “iluminar” los problemas conceptuales.

¹⁰Es decir, buscamos explicaciones porque nos proporcionan de entendimiento. Del mismo modo puede suceder en otros campos de la actividad intelectual. Así, por ejemplo, Thurston (1994) propone que lo que los matemáticos hacen es “buscar maneras en que la gente pueda entender y pensar sobre las matemáticas”.

cedió), busca comprender *por qué* sucedió. A su vez, cuando se entiende por qué ocurrió un evento se puede dar una explicación de él. En consecuencia, puede decirse que al buscar una explicación general para un fenómeno (por ejemplo, porqué hay actividad volcánica), se intenta obtener entendimiento que permita entregar otras explicaciones (por ejemplo, una explicación de la erupción del volcán Descabezado Grande en 1932). Una buena explicación puede habilitar a alguien a dar nuevas explicaciones.

Este ejemplo permite ilustrar la diferencia intuitiva entre saber por qué y entender por qué. Saber por qué no implica que se tenga la habilidad de dar explicaciones a fenómenos particulares distintos al que se conoce pero relacionados con él, mientras que entender por qué sí. Incluso si se puede dar una explicación para un caso particular (por ejemplo, si se sabe por qué ocurrió la erupción del volcán Descabezado Grande en 1932), si no se lo puede conectar con una explicación más general y posibles explicaciones de casos particulares no puede decirse que se entiende por qué ocurrió el evento en cuestión.

Baumberger (2011) da otro ejemplo de esta distinción, que puede ser más sencillo de apreciar. Se puede saber que la temperatura global ha aumentado sin entender por qué. Ahora bien, si un científico explica que ha incrementado a causa de la concentración de los gases de invernadero, y si esa explicación es confiable, se puede decir que se sabe por qué ha aumentado la temperatura global. Pero si no se comprende cómo la concentración de gases de invernadero puede causar el calentamiento global, no se puede decir que se entiende por qué la temperatura ha subido.

¿Puede ser que, a pesar de estas intuiciones, el entendimiento sea una forma de conocimiento? Para examinar el problema tienen que hacerse distinciones más finas sobre el concepto.

Las distinciones básicas que hay que hacer sobre la noción se fundamentan en la diversidad de usos de la expresión “entender” y sus derivadas en el lenguaje ordinario. Considérense los siguientes casos (considero que cada uno de ellos representa un *tipo* de enunciado):

- 1) Entiendo lo que me dices.
- 2) Entiendo que me iban a entregar un paquete.
- 3) Linus entiende el programa de computadora.
- 4) Gracias a la explicación del profesor, entiendo por qué la teoría newtoniana es una mejora sobre la de Galileo.

¿Cómo *analizar* enunciados de estos tipos? El examen de los tipos de enunciados que representan estos ejemplos permite apreciar varios aspectos de la noción de entendimiento.

1) *Entiendo lo que me dices*. Entender lo que dice alguien es parte crucial de la vida cotidiana. Implica, claro está, más que simplemente saber qué es lo que están diciendo. Por ejemplo, si escucho a alguien decir “Al gaan ek ook in ’n dal van doodskaduwee, ek sal geen onheil vrees nie; want U is met my” puedo decir que sé lo que dijo (así como quien lee este texto puede decir que sabe lo que dice), pero es plausible que no entienda lo que dijo (podría ser que incluso no pueda identificar el

idioma en que habló).¹¹ Hay, claro está, una cierta ambigüedad en esta observación. A menudo en español “lo que dijo” se toma en el sentido de “lo que quiso decir”, y es claro que si se lo toma en este sentido habría que decir que no se sabe “lo que se dijo”.¹² Sin embargo, parece que el sentido original es primario. *Entender* lo que se dice tiene o puede tener un aspecto normativo, en el que el hablante (quien dice lo que se entiende) y el receptor (quien entiende lo que se dice) quedan posiblemente comprometidos prácticamente; por ejemplo, en el diálogo

A: El plan es... ¿Entendido?

B: Sí.

B asiente al plan respondiendo que “sí”, que es equivalente en este caso a la afirmación “entendido”. En este caso “entender” implica “asentir al plan”, que es distinto a “saber cuál es el plan”. Por supuesto, en este caso, el asentimiento puede estar implícito a las circunstancias del diálogo.

2) *Entiendo que me iban a entregar un paquete*. La construcción “entender que...” se usa generalmente en contextos en que la expresión conlleva una implicatura pragmática; de forma general, es una forma de expresar las creencias y su grado, y en algunos casos de llevar al receptor a realizar alguna acción. Por ejemplo, si alguien llega a un edificio y le dice al conserje “entiendo que me iban a entregar un paquete”, el conserje entiende que la persona le sugiere que haga algo (contactarse con alguien o buscar el paquete mencionado, verificar la identidad de la persona, etc.) con el fin de que le entreguen el paquete. Si alguien dice, por ejemplo, “tengo entendido que el presidente dijo...” es plausible que busque confirmación de su creencia (por ejemplo, si se dice en tono de interrogación). En este último caso es plausible también que no se requiera una acción de parte del receptor, y que el enunciado tenga que interpretarse solamente como una expresión de creencia. Expresiones de creencia de este tipo pueden incluso interpretarse como una expresión de conocimiento (por ejemplo, el enunciado “María entiende que su clase empieza en una hora” podría expresar simplemente que María sabe que su clase empieza en una hora).

3) *Linus entiende el programa de computadora*. Podríamos querer conectar este enunciado a

7) Linus sabe cómo opera el programa de computadora.

8) Linus sabe operar el programa de computadora.

Es necesario tener un criterio para distinguir si 3) significa 7) u 8) en algún contexto determinado. Si bien inicialmente podríamos estar dispuestos a pensar que 7) implica 8), no ocurre lo mismo al revés (para la mayoría de los usuarios de programas de computadora ellos son, efectivamente, cajas negras).

¹¹ Es la versión en afrikaans del Salmo 23:4: “Aun si voy por valles tenebrosos, no temo peligro alguno porque tú estás a mi lado”.

¹² Una complicación adicional es la distinción entre “lo que [se] quiso decir” y “lo que quiere decir lo que se dijo”. Intuitivamente, uno puede saber lo segundo sin saber lo primero. Por ejemplo, cuando alguien usa una expresión en un doble sentido y quien escucha interpreta la expresión literalmente.

Grimm (2010) examina la relación entre enunciados como ‘Linus understands the computer program’ (i) y enunciados como ‘Linus knows the program’ (ii) en términos de la noción de *saber-cómo* (*know-how*). En inglés los enunciados del segundo tipo son ambiguos y pueden tomarse en tres sentidos. (i) puede entenderse como ‘Linus tiene familiaridad con el programa de computadora’, como ‘Linus tiene conocimiento sobre el programa de computadora’ y como ‘Linus sabe cómo opera el programa’. En español, ‘Linus conoce el programa’ (iii) sólo exhibe el primer sentido en forma primaria (es decir, que si alguien dice (iii) en el segundo sentido, se tomaría su enunciado como una forma divergente del uso habitual), mientras que en inglés ‘Linus knows the program’ es genuinamente ambiguo.¹³ Grimm sugiere que los enunciados como (i) pueden analizarse según el patrón del tercer sentido de ‘Linus knows the program’.

El enunciado “Linus sabe cómo operar el programa de computadora” presenta una ambigüedad adicional. Uno puede saber cómo *se* opera el programa sin saber operarlo uno mismo (en el sentido, por ejemplo, en que alguien puede saber que se usa una interfaz de línea de comandos, sin saber cuáles son los comandos que hay que emplear en esa interfaz). La diferencia entre el saber-cómo de tercera persona del primer caso y el de primera persona en el segundo es significativa. Se puede construir la noción de saber-cómo de tercera persona como una forma de saber cosas sobre cómo operar el programa. La noción de primera persona no excluye la de tercera persona.

Esto es, me parece, *parte* de la distinción entre 7) y 8). 7) implica saber cómo de tercera persona y 8) implica saber-cómo de primera persona. ¿Puede reducirse el concepto de entendimiento de 7) al saber-cómo de tercera persona, y el de 8) al saber-cómo de primera persona?

Se ha caracterizado al entendimiento como una disposición de un sujeto hacia algo que va más allá de la que constituye el conocimiento. Una vez que se entiende algo, se lo puede entender *mejor* (el entendimiento es gradual), pero no hay algo así como un super-entendimiento que implique entendimiento, que sea distinto a él en tipo, y que mejore la situación epistémica del sujeto que lo tenga (así como el entendimiento puede implicar conocimiento, es distinto a este en tipo, y no sólo en grado, y mejora la situación epistémica de quien entiende). ¿Son el saber-cómo de primera y tercera persona el nivel más alto, por así decirlo, de las disposiciones que alguien puede tener respecto al programa de computadora?

Si alguien sabe cómo opera un programa de computadora, en el sentido de saber cosas sobre cómo opera, es claro que aún podría estar en una disposición epistémica superior. Por ejemplo,

¹³Quizás en inglés el segundo sentido tampoco sea primario. Si alguien quisiera expresar ese contenido (‘Linus tiene conocimiento sobre el programa’), diría que ‘Linus knows [a lot/a little/somewhat] about the program’ o usaría la expresión explícita ‘Linux has knowledge about the program’. En el primer caso es necesario el complemento de grado (‘a lot’, ‘a little’, ‘somewhat’, etc): la forma ‘Linux knows about the program’ expresa más bien que Linus tiene conocimiento de que el programa existe, o que le ha sido mencionado, lo que no es tampoco equivalente al primer sentido descrito en que podría tomarse ‘Linus knows the program’ (es posible pensar que el enunciado ‘Linus knows about the program, but he doesn’t really know it’ puede ser correcto, incluso si se toma el ‘know’ en ‘he doesn’t really know it’ en el primer sentido).

podría no poder explicar la operación del programa. Me parece que en un sentido de “entender”, si no se puede hacer eso, no se puede decir que alguien “entiende” algo. El entendimiento no puede reducirse en este caso al saber-cómo de tercera persona. ¿Sucede lo mismo en el caso del saber-cómo de primera persona? En este caso podría suceder también que quien sepa cómo operar el programa no pueda explicar por qué lo opera así como lo hace. Puede “saber” operar el programa y aún así sólo poder describir lo que hace con él e indicar que lo que hace funciona, esto es, cumple el objetivo de operar el programa, satisface sus metas. Pero esto último *cuenta* como una explicación suficiente de por qué hace lo que hace, son razones por las que lo hace. Para explicarse, no es necesario que sepa por qué opera el programa cómo opera (eso es algo que podría no saber). Poder explicar por qué el programa opera como opera no es constitutivo del saber-cómo de primera persona. Quizás esto significa que el saber-cómo de primera persona puede ser una forma de entendimiento. En todo caso, es plausible decir que si esto es así 7) y 8) expresan distintos tipos de entendimiento, si es que expresan entendimiento.

Relacionado con este punto, existe la intuición de ciertas cosas no pueden entenderse a menos que se adopte la primera persona. Por ejemplo, se ha propuesto que la experiencia no puede entenderse sin una explicación de la experiencia cualitativa. El argumento clásico para esta postura es el experimento mental de Mary, propuesto por Frank Jackson (1982). En una variante del experimento mental, Mary es una científico brillante que ha pasado toda su vida en una habitación en la que no se ven colores (no hay cosas en la habitación que ella pueda ver coloreadas), en la que ha obtenido sin embargo toda la información física sobre lo que pasa cuando se percibe y se juzga que tal o cual cosa tiene tal o cual color.¹⁴ En un momento, Mary abandona la pieza y ve ella misma colores. Jackson pregunta: “¿Aprenderá algo o no? Parece obvio que aprenderá algo sobre el mundo y nuestra experiencia visual de él. Pero entonces es ineludible que su conocimiento previo era incompleto.” Se puede describir la situación que propone el experimento mental de esta manera: inicialmente, ella tenía saber-cómo de tercera persona de la percepción de colores sin tener saber-cómo de primera persona de ella, que sólo obtiene una vez que percibe ella misma colores; la intuición de la conclusión es que originalmente Mary “no entendía realmente” la percepción de colores. ¿Qué hay de cierto en el contenido de esta intuición? *No* que en su situación inicial no tiene entendimiento de la percepción de colores, sino que su entendimiento es mayor (más completo) una vez que lo percibe. Me parece que una teoría sobre el entendimiento tiene que explicar este caso.¹⁵

¹⁴En la versión original de Jackson, Mary es forzada a investigar el mundo desde una habitación en blanco y negro mediante un televisor en blanco y negro. La variante que se ofrece intenta evitar la objeción de que en esas circunstancias Mary aún podría tener experiencia de los colores.

¹⁵Esta formulación de la situación del experimento mental evita algunas críticas al experimento mental por desplazar la cuestión del entendimiento que Mary ganaría a la del entendimiento que Mary ganaría. El punto es que la teoría fisicalista no permite obtener entendimiento sobre la percepción, y precisamente eso, como se ha propuesto, debe ser el objetivo de una teoría sobre ella. Así, por ejemplo, la objeción de que Mary no gana conocimiento nuevo sino tan sólo habilidades no socava el peso del argumento; en efecto, me parece, mejora el caso de que el punto sobre el entendimiento

4) *Gracias a la explicación del profesor, entiendo por qué la teoría newtoniana es una mejora sobre la de Galileo.* Respecto al contenido del enunciado, las caracterizaciones que se han dado de la noción de “entender por qué” deberían ser suficientes. Sin embargo, en vistas a las observaciones que se han hecho sobre juicios como “Linus entiende el programa de computadora”, es apropiado explicitar los contrastes entre ellos y juicios de este tipo.

En general, un juicio de la forma

O) *S* entiende por qué *p*

y uno de la forma

Ñ) *S* entiende *p*

van a expresar, en el caso ser una instancia del esquema P), que *S* tiene *entendimiento de por qué* y, en el caso de ser una instancia de O), que *S* tiene *entendimiento objetual*. La primera diferencia es que el entendimiento de por qué tiene como objeto un determinado contenido proposicional, mientras que esta restricción no es necesaria sobre el entendimiento objetual. A su vez, tanto el entendimiento objetual como el entendimiento de por qué se distinguen del *entendimiento de que*, que puede ejemplificarse con el enunciado 2), que se examinaba más arriba. Como el contenido del entendimiento de que es proposicional, se puede querer definir al objeto del entendimiento objetual como algo específicamente no-proposicional.

Considerar algunos ejemplos puede dar una perspectiva más sustantiva sobre la diferencia entre el entendimiento de por qué y el entendimiento objetual. Baumberger (2011) da como ejemplo de la distinción que entender el calentamiento global entalla más que entender por qué ocurre: implica entender qué efectos tendrá, qué relación tiene con la actividad humana y cuánto es probable que suba la temperatura en el futuro; involucra ser capaz de comprender y dar explicaciones para un cierto rango de hechos, trazar conclusiones a partir de la variedad de información disponible y responder preguntas sobre qué pasaría en varios escenarios contrafácticos. En comparación al entendimiento de por qué, el entendimiento objetual involucra apresar más relaciones, tanto explicativas como formadoras de coherencia, dentro de un cuerpo comprensivo de información (véase también Elgin (2007)). O, como propone Grimm (2010), el objeto del entendimiento es una representación abstracta de un sistema, que codifica las maneras en que los distintos aspectos del sistema se relacionan unos con otros. Este último además conecta este tipo de estructuras al *saber-cómo*: el objeto del saber-cómo son estructuras, nos dice, que pueden ser analizadas para ver cómo se relacionan sus partes. En la sección siguiente se verán algunas teorías sobre el entendimiento que pueden acomodar estas ideas.

es correcto. Lewis (1990) dice: “... knowing what an experience is like just is the possession of these abilities to remember, imagine, and recognize. ... It isn't knowing-that. It's knowing-how” (pág. 516).

Otras distinciones que es preciso hacer tienen que ver con la *normatividad* del entendimiento. El problema se presenta en dos niveles: primero, si el entendimiento tiene que responder a los hechos, y segundo, cómo tiene que responder a los hechos si es que tiene que hacerlo.

El siguiente escenario ilustra el primer problema. Una persona que cree en la magia (la llamaremos L) desea que alguien (le llamaremos V) muera horriblemente, y realiza un ritual para, en su opinión, inducir la muerte de esa persona. Ahora bien, resulta que esa persona en efecto muere horriblemente en un accidente automovilístico. L cree que el ritual afectó la serie de acontecimientos que llevaron a la muerte de V, y por lo tanto, que los hechos respondieron a sus deseos. ¿Tiene L entendimiento de la muerte de V?

Existe la fuerte disposición a pensar que no; desear que las cosas sean así o así no tiene influencia causal, ni siquiera mediante un ritual. Sin embargo, podría querer decirse que, porque logra dar sentido al hecho de la muerte de V y a sus propias acciones, L en realidad tiene una especie de entendimiento de la muerte de V. Ese entendimiento, sin embargo, no responde a los hechos. No logra, por otra parte, constituir entendimiento en el sentido en que L no tiene entendimiento de la muerte de V.

La distinción que existe aquí es la de las formas *subjetivas* y *objetivas* del entendimiento. A grandes rasgos, una persona entiende algo subjetivamente si logra hacerse una imagen del mundo que lo explique y que tenga coherencia con el resto de *sus* creencias sobre el mundo. La explicación de L de la muerte de V es coherente con su visión del mundo. En este sentido no importa si las creencias del sujeto sobre el mundo sean correctas o incorrectas. Pero se diría que alguien tiene entendimiento objetivo de algo si, además de hacerse una imagen coherente del mundo que explique lo que se entiende, las creencias en las que consiste esa imagen son correctas.

El entendimiento subjetivo no puede desecharse simplemente por no requerir que las creencias que implica no tengan que ser correctas. Por un lado, es plausible que el entendimiento que un sujeto tenga sea fundamentalmente subjetivo sin ser además objetivo, y que aún así, le permita adecuarse a su entorno (claro está, puede adoptarse un criterio que sugiera que las creencias que le permitan a un sujeto adaptarse a su entorno son correctas, pero hay un sentido obvio en que esto se contradice con otras intuiciones sobre la noción de verdad, entendida como la corrección de un enunciado, como por ejemplo, que ella tiene que ser estable, y que un enunciado verdadero puede o bien no afectar la adaptación de quien lo sostenga a su entorno o que la reduzca; las dificultades que pueden aparecer en este punto nos alejan de lo importante aquí). Por otro lado, el entendimiento subjetivo juega un rol importante en entender las creencias de los demás; entender a alguien que tiene entendimiento objetivo sobre un fenómeno puede incluso incrementar el entendimiento objetivo que se tenga sobre el fenómeno (por ejemplo, si se entiende que tal o cual estrategia para abordarlo, que esa persona adoptó, es incorrecta). Esta forma de entendimiento parece ser común en disciplinas académicas como la filosofía. Otra razón para no desechar el entendimiento subjetivo es que una teoría que parece correcta puede resultar incorrecta, y que por lo tanto el entendimiento que se crea objetivo

se reduzca a entendimiento objetivo. La relación entre una teoría incorrecta y una correcta puede ser “estructural” (en el sentido de que puede consistir en una cierta similaridad), de modo que sólo se pueda llegar a la correcta mediante la incorrecta (aunque no es así en todos los casos). Elgin (2007) da un ejemplo:

A second grader’s understanding of human evolution might include as a central strand the proposition that human beings descended from apes. A more sophisticated understanding has it that human beings and the other great apes descended from a common hominid ancestor (who was not, strictly speaking, an ape). The child’s opinion displays some grasp of evolution. It is clearly cognitively better than the belief that humans did not evolve. But it is not strictly true. . . Perhaps it is not unreasonable to conclude that she does not understand. But the pattern exhibited in this case is endemic to scientific education. We typically begin with rough characterizations that properly orient us toward the phenomena, and then refine the characterizations as our understanding of the science advances. Think of the trajectory from naïve folk physics through Newtonian mechanics to relativity and quantum mechanics.

Esto indica que la separación entre el entendimiento objetivo y el subjetivo no es una cuestión de límites fijos, sino que es más o menos borrosa. Si alguien quiere decir que entiende algo *correctamente* estará dispuesto también, me parece, a asumir que esa corrección está fundada en los hechos, aunque el objeto del entendimiento no sea precisamente verdadero. No sólo se entiende lo que es el caso, sino también lo que no lo es.

Hay un último punto que vale la pena mencionar, y que guarda relación con lo que acaba de decirse: el problema de la suerte epistémica. Supóngase que se lee un libro sobre la caída del Imperio Romano escrito por una persona que no hizo ninguna clase de investigación historiográfica, sino que sólo imaginó una serie de sucesos y sus relaciones. A pesar de esto, la mayoría de los enunciados del libro son literalmente verdaderos (accidentalmente, los hechos imaginados por el autor ocurrieron en realidad).¹⁶ ¿Puede decirse que se gana conocimiento sobre la caída del Imperio Romano leyendo ese libro? Pareciera que no. ¿Puede decirse que se obtiene entendimiento de la caída del Imperio Romano leyéndolo? Quizás sí; si bien el conocimiento no admite la suerte epistémica, el entendimiento sí parece admitirla. No es suficiente para explicar esto apelar a la distinción entre entendimiento objetivo y subjetivo, porque en este caso los enunciados (y las creencias) en cuestión son verdaderos, pero hay al menos un vínculo importante entre casos como este y esa distinción: es claramente posible que alguien tenga entendimiento subjetivo a partir de creencias meramente afortunadas. No está tan claro, sin embargo, que pueda tener entendimiento objetivo de esta manera.

Puede distinguirse entre casos de suerte epistémica del tipo de Gettier y otros en que la suerte epistémica es “ambiental”. En el primer caso, se llega a creencias verdaderas sin que la justificación

¹⁶Este ejemplo está inspirado en los que propone Kvanvig (2003, 197).

subjetiva de la creencia sea correcta objetivamente. Por ejemplo: al mirar un campo se llega a la creencia de que hay una oveja, porque se ve algo que tiene la figura de una. En realidad, se trata de un animal distinto, pero además sí hay una oveja en el campo, sólo que no puede verse. La creencia de que hay una oveja es verdadera, pero su justificación no es objetiva (que en este caso tendría que ser ver a la oveja oculta). En el caso de la suerte epistémica ambiental, hay justificación objetiva, pero las circunstancias en que se obtiene la creencia socavan esta justificación. Por ejemplo, alguien puede tener justificación para creer que hay un taller de bicicletas cerca, pero en circunstancias en que también hay muchos talleres de bicicletas falsos (meras fachadas de construcciones) cerca, lo que hace parecer que se llegó a la creencia de que hay un taller de bicicletas como una mera cuestión de suerte. Se ha propuesto que el entendimiento resiste o bien uno u otro tipo de suerte epistémica, o ambos, pero aquí basta observar que de cualquiera de estas maneras puede inferirse que el entendimiento no puede reducirse a alguna forma de conocimiento. Pritchard (2009) propone que el entendimiento objetivo sólo puede resistir casos de suerte epistémica ambiental. Él propone como contraejemplo a la tesis de la resistencia del entendimiento en casos de suerte epistémica del tipo de Gettier que no se puede obtener entendimiento mediante las explicaciones de alguien que realmente no entiende lo explicado (alguien que da una explicación correcta de un fenómeno por mera casualidad); por ejemplo, una persona vestida de bombero, pero que no es un bombero, que da una explicación correcta de por qué se quemó una casa sin haber hecho ninguna investigación sobre las causas del incendio, simplemente diciendo lo primero que se le viene a la mente. El entendimiento sí puede resistir el caso análogo en que un verdadero bombero le da una explicación correcta a alguien, a pesar de que esa persona pudo haberle pedido la explicación a otras personas vestidas como bomberos.

En resumen, una caracterización correcta del entendimiento debe asumir:

- 1) El entendimiento es el bien intrínseco de las explicaciones,
- 2) El entendimiento debe distinguirse del conocimiento. Entender por qué p es distinto a saber que p y a saber por qué p . Entender un concepto es distinto a saber que se usa en ciertas condiciones.
- 3) Hay que distinguir entre el *entendimiento de por qué...* y el *entendimiento objetual*.
- 4) El entendimiento objetual se sustenta en el *saber-cómo*.
- 5) Hay que distinguir entre el concepto de saber-cómo de primera persona y el de tercera persona.
- 6) El concepto de entendimiento es gradual.
- 7) El entendimiento no es fáctico. Entender x no significa que x sea el caso. El entendimiento correcto, sin embargo, se cimienta en los hechos.
- 8) El entendimiento resiste algunas formas de suerte epistémica.

3.2.2. EXPLICACIÓN Y ENTENDIMIENTO

Como se ha apuntado antes, el entendimiento parece jugar un papel importante en la explicación (el objetivo de las explicaciones sería, en el modelo propuesto, proporcionar entendimiento sobre

los *explananda*). Como consecuencia, las teorías de la explicación (o más precisamente, sobre lo que realiza una explicación) pueden proponerse como teorías sobre el entendimiento.

Siguiendo a Lipton (2004), se pueden ofrecer tres pruebas cruciales para determinar la aceptabilidad de una teoría de la explicación. Una teoría de la explicación correcta debe

- 1) distinguir entre el conocimiento y el entendimiento que se obtiene mediante la explicación (ya que este punto fue examinado en detalle en la sección anterior no me voy a detener en él),
- 2) permitir que las explicaciones no estén necesariamente explicadas ellas mismas (esto es, impedir que la cadena de explicaciones caiga en un regreso al infinito),
- 3) dar cuenta de explicaciones que se evidencien a sí mismas.

Respecto a lo que concierne al entendimiento, 2) implica que es correcto decir que se entiende algo aunque no se pueda dar respuesta a las preguntas de por qué que puedan aparecer sobre el asunto. 2) afirma que es permisible dar una explicación de algo sin que *cada* elemento de la explicación tenga que ser explicado. Por ejemplo, un historiador imaginario que se pregunte por qué los registros de las cosechas en Europa durante el siglo XIV indican una disminución en la producción podría intentar explicar el fenómeno apelando a los efectos de la Peste Negra en la población rural: la disminución en la población rural activa implicó que importantes zonas agrícolas no fueran explotadas o fueran explotadas con menor intensidad. Esta es una explicación suficiente de la disminución de las cosechas, y permite, en efecto entender los registros que observa el historiador. Sin embargo, uno aún podría preguntar por qué ocurrió la Peste Negra. Para que se diga que se entiende el fenómeno gracias a esa explicación no hace falta que se tenga una respuesta a esa pregunta. Incluso, sería intuitivamente incorrecto decir que la disminución de las cosechas se explica por la rápida propagación de la plaga de la bacteria *Yersinia pestis* (una posible explicación de la Peste Negra). Esa explicación no ofrecería entendimiento sobre el fenómeno en cuestión. Por otro lado, la posibilidad de que pueda entrarse en una cadena de preguntas de por qué puede ser vista como una característica benigna de la explicación porque en cada eslabón se puede obtener entendimiento; es posible incluso que sólo se pueda entender algo correctamente (que es distinto a simplemente entender algo, como se vio cuando se hizo la distinción entre el entendimiento subjetivo y objetivo) si se pueden dar explicaciones para varias cadenas de preguntas que surgen desde la consideración de lo que se busca entender.

La relación de 3) con el entendimiento es más indirecta. Una explicación que se evidencia a sí misma es una explicación en que lo que se explica sirve como evidencia para su explicación. Por ejemplo, las marcas en la nieve son evidencia de lo que las explica: que alguien o algo (dependiendo de las características de las marcas) pasó por allí. El corrimiento al rojo de las estrellas es evidencia de que se alejan, lo que a su vez explica el fenómeno del corrimiento al rojo. Una teoría de la explicación que no permita explicaciones de este tipo o que interprete la circularidad que presentan como viciosa debe estar equivocada, porque es claro al observar estos ejemplos que

son explicaciones correctas. Lipton propone 3) como requisito para una teoría de la explicación en parte también porque él mismo (2001) cree que hay que dar cuenta de las explicaciones de este tipo como prácticas inferenciales abductivas:

In these cases, it is not simply that the phenomena to be explained provide reasons for inferring the explanations; we infer the explanations precisely because they would, if true, explain the phenomena. . . The sleuth infers that the butler did it, since this is the best explanation of the evidence before him. . . our inferential practices are governed by explanatory considerations. Given our data and our background beliefs, we infer what would, if true, provide the best of the competing explanations we can generate of those data. (págs. 57–58)

La noción de “la mejor de las explicaciones potenciales en competencia”, sugiere (pág. 61 y ss.), debe construirse como la de lo que entregue el mayor grado de entendimiento potencial. Según él, las consideraciones sobre la plausibilidad de una explicación están guiadas por las consideraciones explicativas que vayan al caso: la explicación que proporcione el entendimiento “más profundo” es la explicación que más probablemente es verdadera. Siguiendo la misma estrategia que prescribe, dice:

Such an account suggests a really lovely explanation of our inferential practice itself, one that links the search for truth and the search for understanding in a fundamental way. (pág. 63)

Siguiendo esta línea, Khalifa (2011) ha propuesto que “entender por qué p ” implica tener la creencia en que p y en que otra proposición, q , es la mejor explicación de p . Es claro que esta caracterización toma el partido contrario al que la que se ha sugerido aquí, pues según esta última no es necesario que se crea lo que se entiende, ni siquiera cuando se trata de entendimiento del por qué de algo. Por ahora no es necesario que me detenga en este punto.

Hechos explícitos estos criterios y las razones para adoptarlos, se puede proceder a examinar los distintos tipos de teorías de la explicación. De estos, hay cinco tipos que pueden considerarse como los principales: la teoría de la familiaridad, la racional, la de la necesidad, la causal y la de la unificación.¹⁷ Hay que tener en cuenta que estas teorías fueron propuestas originalmente como

¹⁷Sigo en esta clasificación a Lipton (2004). Esta taxonomía de teorías es, en cierto sentido, idiosincrática. Por ejemplo, no es obvio a primera vista cómo situar teorías sobre la explicación como el modelo nomológico-deductivo de Hempel y Oppenheim y la que Friedman (1974) llama “de las modas intelectuales”. Esto es más bien producto de que esas teorías no presentan explícitamente tesis que se acomoden a los esquemas con que se puede caracterizar cada tipo que de un problema con la categorización. La mayoría de las reseñas del avance de las investigaciones en el campo de la explicación científica—por ejemplo, Salmon (1989)—sigue una exposición histórica, partiendo por el modelo nomológico-deductivo, para luego pasar al causal, la disputa sobre el realismo y luego introducir la teoría de la unificación.

explicaciones de la explicación y el entendimiento *científicos*, pero pueden generalizarse a otros dominios.

i) *Teoría de la familiaridad*. Según estas teorías, explicar algo consiste en reducirlo a elementos que son familiares (ya conocidos o ya comprendidos):

Familiaridad Un sujeto S entiende [por qué] p si [y sólo si] S puede reducir p a q , donde q es familiar a S o S entiende q .

Lo que no se entiende es lo que no es familiar y precisamente porque no es familiar es que se le busca una explicación.

Por ejemplo,¹⁸ podría sugerirse que la teoría cinética de los gases gana su poder explicativo mediante las comparaciones que pueden hacerse con ella entre los fenómenos que quieren explicarse, como la ley de Boyle, que afirma que a una temperatura constante la presión y el volumen de un gas son inversamente proporcionales, y otros que son más familiares, como el movimiento de las bolas de billar, que mientras más se acercan unas a otras (este sería el análogo a la acción de la presión en las moléculas de un gas) ocupan un menor volumen.

Lipton (2004) observa sobre la teoría de la familiaridad que su atractivo consiste en

... the natural way it accounts for the fact that it is often surprise that prompts a request for explanation. It is often when things do not turn out as we expected that we want to know why. Moreover, even when we ask why about what is already in some sense familiar, the prompt for the question often involves 'defamiliarisation': we are brought to see the everyday situation as somehow strange or surprising... Most people do not wonder why the same side of the moon always faces the earth, perhaps because they erroneously suppose that this is simply a consequence of the moon not spinning. Once they are shown that, not only does the phenomenon require that the moon spin, but that its period be precisely the same as the apparently unrelated period of the moon's orbit around the earth, the phenomenon becomes surprising and prompts a why-question.

(pág. 6)

La teoría de la familiaridad presenta varias dificultades. En primer lugar, muchas explicaciones científicas van en la dirección opuesta: se explica algo familiar mediante algo que no es familiar (por ejemplo, se puede explicar el enlace entre el hidrógeno y el oxígeno al componer el agua mediante el comportamiento de onda de los electrones en las moléculas de agua). El problema aquí es conceptual: entender y tener familiaridad con algo son separables. Por ejemplo, tengo familiaridad con el funcionamiento de mi calefactor, pero no puedo decir que lo entiendo. Si bien esto hace que la teoría de la familiaridad satisfaga el primero de los criterios que se propusieron para una teoría de la explicación correcta (a saber, que se pueda distinguir entre conocer algo y entenderlo), lo hace

¹⁸Cf. Friedman (1974, 9).

de manera tal que de la teoría se siguen consecuencias incorrectas, como que explicaciones intuitivamente correctas como la que se acaba de mencionar, en que algo familiar se explica mediante algo que no es familiar no podrían considerarse propiamente como explicaciones (!). Por lo demás, esto significa que la teoría de la familiaridad no puede dar cuenta de la regresión de explicaciones: si hay que explicar o entender algo con base en algo que ya es familiar o comprensible, se puede caer en una regresión viciosa pues decir que una explicación es correcta requeriría de que toda la cadena de explicaciones que puedan surgir esté dada (para decir que se entiende algo, habría que entender lo que lo explica, y a su vez habría que entender lo que explica lo que lo explica, etc.)

La teoría de la familiaridad tampoco parece poder dar cuenta de explicaciones que se evidencian a sí mismas. Algo que requiera de una explicación por no ser familiar y que sea en sí mismo evidencia de algo familiar parece absurdo. Un evento nunca antes visto y que requiera explicación no parece poder ser evidencia de la presencia de algo familiar: si lo fuera, o bien no podría ser caracterizado como algo “nunca antes visto” o bien no podría requerir de una explicación (¿cómo dar cuenta entonces de la ilusión de que requiera de una explicación?).

ii) *Teoría racional*. De acuerdo con esta teoría, entender por qué ocurre algo es tener buenas razones para creer que ocurre:

Racionalidad Un sujeto S entiende [por qué] p si [y sólo si] S tiene buenas razones para creer que p .

La teoría racional es en realidad una teoría sobre el significado de las preguntas de por qué (si es una teoría revisionista depende de si se incluye la cláusula “y sólo si” en el análisis). Según ella, cuando se pregunta ‘¿por qué p ?’ hay que entender la pregunta en el sentido de ‘¿por qué debería creer que p ?’.

Hay que recordar que C. Hempel (1965) llama a este tipo de preguntas “preguntas de por qué que buscan razones” y las distingue de las que buscan explicaciones. Si se tiene una respuesta a una pregunta de por qué que busque una explicación se pueden derivar respuestas a preguntas que busquen razones. Por ejemplo, si sé por qué me enfermé (por ejemplo, que estuve en contacto con alguien enfermo y no tomé las precauciones debidas), puedo saber qué razones justificarían una creencia en que esté enfermo. La teoría de la familiaridad implica que además, de una respuesta a la segunda pregunta puede derivarse una respuesta a la primera.

No es claro que esto sea correcto. En primer lugar, porque se puede saber que algo es, fue o será el caso (lo que implica tener razones para tener esa creencia) sin entender por qué. Alguien puede saber que siempre que se observe la Luna, está presentará el mismo lado que ha presentado antes sin entender por qué: pero según la teoría de la racionalidad saber ese hecho sobre la Luna implica entenderlo. Así, la teoría no puede satisfacer el primer requisito.

Tampoco puede satisfacer el segundo requisito. Puede suponerse que una explicación no provee de una razón para creer en algo a menos que haya una razón para creer en esa explicación. Por

ejemplo, creo que me enfermé porque creo que no me cuidé. Para que esta explicación sea correcta, según la teoría racional, tengo que tener razones para la creencia de que no me cuidé, es decir, tiene que estar ella misma explicada. Pero como se vio antes, esto no es necesario.

Por último, la teoría racional tampoco puede acomodar explicaciones que se evidencien a sí mismas. Si una explicación es una razón para creer por qué ocurrió algo, decir que el que haya ocurrido es una razón para creer en la explicación es viciosamente circular.

iii) *Teoría de la necesidad*. Según teorías de este tipo, entender algo implica ver la necesidad de que ocurra:

Necesidad Un sujeto S entiende [por qué] p si [y sólo si] S sabe que p es necesario (o sabe $\Box p$) en un sentido adecuado.

Una explicación implica (en un sentido amplio) la necesidad de lo que explica: muestra que tenía que suceder o ser como se lo describe. El atractivo de la teoría está en que recoge la intuición de que entender algo va más allá del mero conocerlo: tener el juicio de que algo tiene que ser así o asá no se sigue de que *sea* así o asá (de este modo, la teoría acomoda el primer criterio).

Es posible que se explique (por ejemplo) la alta temperatura que se registró el día... en ... como un suceso físicamente necesario de acuerdo con las condiciones dadas y a las leyes naturales. Algo físicamente necesario puede definirse como algo que es verdadero en todos los mundos posibles en que se sostiene el conjunto de leyes naturales que rigen los fenómenos naturales en el mundo real. En este sentido, que la temperatura haya sido de... grados el día... en ... se sigue con necesidad de las leyes naturales y las condiciones iniciales dadas. El aspecto metafísico del modelo nomológico-deductivo de Hempel y Oppenheim (1948) sigue fundamentalmente este patrón de explicación (aunque este modelo puede generalizarse para cubrir casos de explicación probabilística también). En este modelo una explicación adecuada debe cumplir al menos estas condiciones lógicas (pág. 137):

- 1) El explanandum debe ser una consecuencia lógica del *explanans*.
- 2) El explanans debe contener leyes generales, y esas debe ser requeridas para derivar el explanandum.

La noción de consecuencia lógica implica que la relación entre el *explanandum* y el *explanans* debe mantenerse en todo mundo posible en que sean el caso: es decir, que son necesarias. Que las explicaciones incorporen leyes generales explica por qué entender por qué sucede algo permite generar nuevas explicaciones en casos similares: estas explicaciones se derivan de las leyes generales que incorpora la explicación desde la que se generan. Entender por qué sucede algo es saber que es esperable dadas las leyes generales.

En otra variedad de esta teoría, Glymour (1980) propone que el entendimiento puede consistir en la eliminación de contingencias mediante la correlación de verdades necesarias, con el fin de obtener otras verdades necesarias (que serían las explicaciones). Por ejemplo, sugiere que al considerar

la explicación cinética de las leyes de los gases ideales (la ley de Boyle y la ley de Charles), y mediante ciertas identificaciones entre las propiedades del caso, estas leyes se convierten en una verdad matemática (pág. 34).

La teoría de la necesidad acomoda las explicaciones que se evidencien a sí mismas. Si el explanans muestra que el explanandum es necesario, no es inconcebible que el explanandum pueda ofrecer evidencia para el explanandum. Por ejemplo, que alguien pasara por aquí hace necesario (en un sentido limitado) que haya ciertas marcas en el suelo, y esas marcas son indicio de que haya pasado alguien por aquí.

No se tiene el mismo éxito al intentar acomodar el regreso de explicaciones con esta teoría: para mostrar que un suceso es necesario, la explicación también tendría que ser necesaria, pero esto no es plausible sobre todos los casos. Por ejemplo, puedo querer decir que me resfrié porque no me cuidé; en este caso no se puede decir que sea necesario que no me haya cuidado.

A primera vista, el aspecto modal de la teoría sugiere que podría interesar adoptarla para los propósitos de esta investigación: el vínculo que establece entre los juicios modales y el entendimiento es fundamental. Sin embargo, la teoría de la necesidad hace depender el entendimiento de la epistemología modal y por lo tanto, no sirve para iluminar los problemas epistemológicos que se han visto sobre los juicios modales; en realidad los presupone. Incluso si pudiera resolverse el problema de la teoría de la necesidad respecto al regreso en la cadena de explicaciones (y es plausible que eso no pueda hacerse), ella es estratégicamente indeseable.

iv) *Teoría causal*. Según la teoría causal, entender implica tener conocimiento causal sobre lo que se intenta entender, conocimiento sobre cuáles son sus causas:

Causalidad Un sujeto S entiende [por qué] p si [y sólo si] S sabe cuáles son las causas de p (o bien, conoce q , donde q son las causas de p)

Esta concepción del entendimiento ligado a las explicaciones tiene bastante apoyo en la observación de las prácticas explicativas cotidianas, donde a menudo se expresan juicios como “me caí porque no me amarré bien los zapatos”, o “me caí porque me empujaron”. En estos casos podemos decir que el que me empujaron o el que no me amarrara los zapatos son en cierto sentido causas de lo que estos juicios explican (a saber, que me caí). En general, entender que c fue la causa de un evento e es tener una explicación de e .

Un locus clásico de la teoría causal es Salmon (1984). Explicar una ocurrencia es mostrar cómo calza en la red causal del mundo (pág. 276). En vez de concebir la causalidad como una relación entre eventos, Salmon sugiere que es una propiedad de *procesos*. Un proceso es algo que mantiene una cierta estructura que lo constituye *a través del tiempo* (por ejemplo, la escritura de esta tesis es un proceso; en contraste, su término es un evento). Un proceso es causal si es capaz de transmitir una modificación de la estructura, lo que a menudo se llama una “marca”. Por ejemplo, el calor en la superficie de una taza es una marca en el proceso en que se enfría el té en la taza. Una marca

se manifiesta en todas las etapas del proceso desde el momento en que se introduce mediante una interacción con una característica del proceso (cf. pág. 148). Una interacción entre procesos causales puede, asimismo, dejar una marca en la estructura de ambos procesos (el ejemplo clásico es el de lo que ocurre en un choque de vehículos). Empleando esta terminología, Salmon propone que explicar un evento consiste en dar una descripción de los procesos e interacciones causales que conducen al evento (o al menos una parte de ellos) y de los procesos e interacciones que constituyen el evento en sí (el suceso se define como una etapa en el proceso causal).

La teoría causal satisface los tres criterios propuestos. En primer lugar, es claro que se puede conocer algo sin saber sus causas. En segundo lugar, no es necesario remontarse a las causas de las causas para explicar causalmente un suceso. Consideraciones sobre la relevancia causal deberían permitir que se tengan explicaciones finitas y se evite la posibilidad de la regresión viciosa. Imagínese que ocurre un incendio. Puede descubrirse que el evento que gatilló el evento fue un cortocircuito en el interior de una muralla de madera. La explicación del incendio apelará a los factores que “hicieron una diferencia” entre que ocurriera el evento (que los cables hayan estado descubiertos, que hubiera material inflamable en el lugar donde inicio el fuego, que se lanzara chispa, etc.) y no, en vez por otros que, estando en conexiones causales con estos factores, no hicieron una diferencia (por ejemplo, que hayan traído las planchas de madera al lugar del incendio cuando se construía el edificio, etc.) En tercer lugar, no es inconcebible que la explicación de un suceso sea su causa y que a la vez el suceso proporcione evidencia para la explicación. Por ejemplo, si se apela a la concepción de Salmon de la causalidad, el proceso causal que explica el suceso debe registrar una marca en la estructura del suceso. Es claro que una marca de este tipo puede servir como indicio de sus causas.

A pesar del éxito aparente de la teoría, presenta varios problemas. Por un lado, no existe consenso sobre cuál es el análisis correcto de la noción de causalidad. La teoría de Salmon (por ejemplo), parece a la vez excluir procesos causales (el requisito de que un proceso debe permanecer uniforme durante un periodo de tiempo es problemático respecto a sucesos de cortísima duración, abundantes en la escala subatómica) y tolerar ciertos pseudo-procesos; no es un candidato indiscutible sobre el problema de la causalidad. Esta dificultad sobre la teoría causal del entendimiento no es insuperable en principio, y no socava la idea central de este tipo de teorías, pero cada variante de ella debe enfrentar los problemas de sus compromisos metafísicos y epistemológicos con una teoría de la causalidad. La teoría causal de la explicación debe proveer también un criterio para determinar qué causas son explicativas, pues es claro que no todas lo son (recuérdese el ejemplo que se dio al hablar de la cuestión del regreso de las explicaciones). Una alternativa es, como se ha apuntado, apelar a la noción de “relevancia causal”, pero tampoco es claro cómo haya que construir esa noción o incluso si es coherente.

Un problema más grave es que no todas las explicaciones parecen tener un aspecto causal; por ejemplo, es dudoso que las explicaciones matemáticas y el entendimiento matemático puedan

construirse en términos causales.¹⁹ Lipton, que es partidario de la teoría causal, argumenta que esto indica que ella no puede considerarse como una teoría sobre la explicación en general, y que podría ser que tal teoría no puede darse (cf. pág. 10.) Glymour (1980) hace la misma observación. Otra posibilidad es que exista una perspectiva más general que a su vez explique las intuiciones de la teoría causal.

v) *Teoría de la unificación*. Este tipo de teoría observa que entender algo consiste en colocarlo en una estructura en la que es coherente. El objetivo del entendimiento sería, en este caso, *integrar* las explicaciones de los fenómenos. La intuición fundamental de esta idea puede formularse de esta manera:

Unificación Un sujeto S entiende [por qué] p su [y sólo si] S sabe cómo p se unifica con una estructura q (o bien, si S integra p en una estructura q).

A menudo se asume que la estructura en la que se integra lo que se entiende es una concepción de cómo es la realidad. Cuando se hace esta suposición se puede hablar de una teoría realista del entendimiento como unificación; en el caso contrario se hablará de una teoría antirealista.

Así como la teoría causal, una teoría de este tipo puede satisfacer los tres criterios propuestos. Primero, porque podemos saber qué algo es el caso sin que sepamos cómo se integra con otras cosas (permite distinguir entre saber que... y entender). Sé, por ejemplo, que la luz se curva como efecto de la gravedad, pero no conozco la totalidad de que fenómenos con que eso está conectado. Segundo, puede acomodar explicaciones que se evidencian a sí mismas. Lipton (2004) ejemplifica esto mediante una analogía: un trozo de un patrón (imagínese un azulejo en un mosaico) puede proveer evidencia sobre el patrón como un todo, mientras que el patrón completo ubica el trozo en un contexto que lo unifica y en el que tiene sentido (pág. 7). Lipton expresa cierta duda sobre si la teoría de la unificación puede acomodar el regreso de las explicaciones: una explicación que unifique parece tener que estar ella misma unificada. Pero quizás haya que decir tan sólo que entender un fenómeno es incrustarlo en un patrón más amplio, aunque no se tenga a su vez un patrón más amplio en que incrustar la explicación. Friedman (1974, 18–19) observa que, ante un crítico que cuestione cómo es posible entender algo mediante una explicación que no ha sido explicada ella misma, no basta con responder que los fenómenos se explican “uno a la vez”, hay que responder que al explicar no se reemplaza un fenómeno inexplicado por otro que tampoco está explicado, sino que se lo reemplaza por un fenómeno más comprensivo.

Es conveniente examinar en detalle algunas de las propuestas específicas que se han hecho sobre cómo enlazar los conceptos de explicación, entendimiento y unificación.

¹⁹Woodward (1989) observa que incluso en física no todas las explicaciones son causales, y menciona explicaciones que se fundamentan en consideraciones sobre la estructura del espacio-tiempo. Por ejemplo, la contracción de las longitudes y la dilación del tiempo en un marco inercial relativista son “efectos” geométricos, y no causales.

El desarrollo original de una teoría del entendimiento basada en el concepto de unificación se debe a Friedman (1974). Allí (pág. 14 y ss.), él toma como ejemplo la teoría cinética de los gases para explicar las motivaciones del enfoque unificacionista. ¿Qué explica la teoría? Una multitud de fenómenos, como por ejemplo el que los gases sigan las leyes de Boyle y de Charles. ¿Cómo los explica? Haciendo referencia al comportamiento cinético de las moléculas: un conjunto de moléculas que obedezcan ciertas leyes mecánicas obedecerán también las leyes de Boyle y de Charles. El número de leyes se reduce, y esto es significativo porque con ello la teoría cinética permite *integrar* el comportamiento de los gases con otros fenómenos, como el movimiento de los planetas y el hecho de que los objetos que caen obedezcan las leyes de Galileo. Friedman afirma que la esencia de la explicación científica es que “incrementa nuestro entendimiento del mundo al reducir el número de fenómenos independientes que tenemos que aceptar como últimos o dados” (pág. 15). La comprensibilidad del mundo se explica mediante la posibilidad de unificar los fenómenos.

Para formular esta idea Friedman desarrolla el concepto de *enunciados aceptables independientemente*. Friedman supone que los fenómenos son representados por enunciados legaliformes, y por lo tanto que hablar del número de fenómenos independientes es equivalente a hablar del número de enunciados legaliformes independientes. En este sentido, una explicación modifica el conjunto de enunciados que una comunidad de investigadores acepta (K): reduce el número de enunciados independientes. El problema es determinar bajo qué condiciones puede decirse que un enunciado legaliforme permite una reducción de este tipo. Supóngase que S es el conjunto de las leyes mecánicas, y que K contiene el conjunto L que incluye la ley de Boyle, la ley de Charles, la ley de Galileo sobre la caída libre y las leyes de Kepler. La intuición de Friedman es que S permite reducir el número de enunciados independientes en K porque se puede reemplazar L por S (se asume que la cardinalidad de L es mayor que la de S). Pero, ¿no puede decirse que esta intuición no es correcta, porque no hay nada que nos impida decir que, L en realidad deba entenderse como la conjunción de esas leyes (es decir, como un único enunciado independiente, en cuyo caso no se reduce el número de enunciados independientes en K)? Un enunciado cualquiera es lógicamente equivalente a una multitud de conjuntos de enunciados de forma trivial: si P es una consecuencia lógica de S , S es equivalente a $P, P \rightarrow S$. Friedman propone salvar esta dificultad introduciendo el concepto de enunciados *aceptables independientemente*. El conjunto $P, P \rightarrow S$ no es aceptable independientemente de S : por ejemplo, podría ser que el único fundamento para aceptar que $P \rightarrow S$ es que sea una consecuencia de S . (pág. 16). Que un enunciado p sea aceptable independientemente a otro enunciado q significa que hay fundamentos suficientes para aceptar p que no son fundamentos suficientes para aceptar q (por ejemplo, la ley de Boyle es aceptable independientemente a la ley de Galileo).

Friedman llama *partición de un enunciado* S a un conjunto Γ de enunciados tal que 1) sus miembros son aceptables independientemente a S y 2) que es equivalente a S . Así por ejemplo, el conjunto de la ley de Boyle, la ley de Galileo y las leyes de Kepler (para abreviar, el conjunto Ω) es

una partición de la conjunción de esas leyes. Un enunciado que no tiene una partición es *K-atómico*. Una *partición-K* de un conjunto de enunciados δ es un conjunto de enunciados *K-atómicos* que es lógicamente equivalente a δ . La *cardinalidad-K* de un conjunto de enunciados δ es la cantidad de elementos en la partición-K de δ (así, la cardinalidad-K del conjunto Ω es al menos 3). Finalmente, se puede decir que un enunciado S reduce un conjunto de enunciados δ si la cardinalidad-K de la unión de δ y S es menor que la cardinalidad-K de δ (así, si S es la conjunción de los miembros de Ω , no se puede decir que S reduzca a Ω , porque tienen la misma cardinalidad-K).

Para definir la noción de explicación Friedman propone que el conjunto de enunciados independientes que se quiere reducir al explicar algo es el conjunto de las consecuencias independientemente aceptables de la explicación. De este modo, las leyes de Newton son un buen candidato para explicar la ley de Boyle porque reducen el conjunto de sus consecuencias independientemente aceptables, que contiene la ley de Boyle, la ley de Galileo, etc. Esto sugiere que la definición del concepto de explicación podría darse según el esquema

D S explica F si y sólo si F es una consecuencia independientemente aceptable de S y S reduce el conjunto de consecuencias independientemente aceptables de S .

Friedman observa que según esta definición es que según ella la conjunción de una explicación S_1 para un enunciado F_i y una ley independientemente aceptable S_2 no puede explicar F_i , puesto que no reduce el conjunto de consecuencias independientemente aceptables del conjunto S_1, S_2 . Pero, por ejemplo, la conjunción de las leyes de Newton y, por ejemplo, las leyes del electromagnetismo, puede servir como una explicación de la ley de Boyle, aunque las leyes del electromagnetismo sean aceptables independientemente de la ley de Boyle. Para resolver esta dificultad, Friedman propone debilitar D1 reformulándola como

D* S explica F si y sólo si existen una partición Γ de S y un enunciado S_γ que pertenece a Γ tal, que F es una consecuencia independientemente aceptable de S_γ y S_γ reduce el conjunto de consecuencias independientemente aceptables de S_γ .

Friedman concluye diciendo que esta teoría de la explicación muestra que ella es un asunto global, no local:

We don't simply replace one phenomenon with another. We replace one phenomenon with a more comprehensive phenomenon, and thereby effect a reduction in the total number of accepted phenomena. We thus genuinely increase our understanding of the world. (pág. 19)

Philip Kitcher (1976) critica el análisis de Friedman en varios puntos y muestra que el análisis de Friedman no se sostiene. Kitcher observa que D implica que sólo enunciados *K-atómicos* pueden ser explicaciones (los enunciados que no son *K-atómicos* no pueden reducir su conjunto de consecuencias aceptables independientemente), pero se pueden dar contraejemplos que muestren

que enunciados que no son K-atómicos pueden explicar. Esto puede ocurrir de dos formas. Primero, considérese la explicación de la ley de la expansión adiabática de un gas ideal (la temperatura de un gas ideal que se expande en el vacío no cambia, y el producto de la presión inicial y el volumen inicial es el mismo que el producto de la presión final y el volumen final), que se da usualmente mediante la conjunción de las leyes de Boyle y Charles y la primera ley de la termodinámica. Estas leyes son aceptables independientemente, y por lo tanto su conjunción no es K-atómica. Sin embargo, la derivación de la ley de la expansión adiabática es una explicación aceptable. En segundo lugar, también ocurre a menudo que la explicación del comportamiento de un sistema complejo apela a leyes que provienen de distintas teorías, que son aceptables independientemente. Por ejemplo, para explicar completamente por qué los ojos humanos son sensibles sólo a ciertos rangos de frecuencias luminosas hay que apelar a leyes biológicas, físicas y ópticas. D^* tiene, de acuerdo a Kitcher, la consecuencia de que ningún enunciados K-atómico puede servir como explicación, lo que no parece ser correcto ni responder a las intuiciones de Friedman. Como consecuencia, Kitcher sugiere que Friedman podría querer reformular su definición como

D^{**} S explica F si y sólo si o bien 1) F es una consecuencia independientemente aceptable de S y S reduce el conjunto de consecuencias independientemente aceptables de S , o 2) existen una partición Γ de S y un enunciado S_γ que pertenece a Γ tal, que F es una consecuencia independientemente aceptable de S_γ y S_γ reduce el conjunto de consecuencias independientemente aceptables de S_γ .

Lamentablemente, ni D^* ni D^{**} pueden acomodar correctamente los contraejemplos de Kitcher (cf. pág. 211–212).

Kitcher (1981, 1989) propone una teoría sobre el entendimiento basada en la noción de unificación diferente a la de Friedman y que evita sus problemas. Según Kitcher (1989), aunque la conexión intuitiva que hace Friedman entre los conceptos de explicación y unificación es correcta, su enfoque para describir esta relación está errado; según él, “entender los fenómenos no es meramente una cuestión de reducir las ‘incomprensibilidades fundamentales’ sino de ver conexiones, patrones comunes, en lo que inicialmente parecían ser situaciones diferentes” (pág. 432). En (1976) ya había dicho que

What is much more striking than the relation between [el número de leyes que explican y el número de leyes que *se* explican] is the fact that Newton’s laws of motion are used again and again and that they are always supplemented by laws of the same types, to wit, laws specifying force distributions, mass distributions, initial velocity distributions, etc. Hence the unification achieved by Newtonian theory seems to consist not in the replacement of a large number of independent laws by a smaller number, but in the repeated use of a small number of types of law which relate a large class of apparently diverse phenomena to a few fundamental magnitudes and properties. (pág. 212)

¿Cómo desarrolla Kitcher esta idea?

Como antecedente, hay que mencionar que Kitcher concibe las explicaciones como un cierto tipo de acto que pone a un sujeto en una actitud específica hacia un argumento, y no solamente como un tipo de argumento. Es por esto que a menudo habla de “acts of explanation” (actos de explicación). En lo que sigue voy a usar la terminología de Kitcher, y seguir el esbozo de la teoría que presenta en Kitcher (1981).

La caracterización inicial del problema del análisis de la noción de explicación de Kitcher (1981, 509–510) es que el problema es equivalente a la cuestión de qué rasgos tiene que tener un argumento si ha de servir de base para un acto de entendimiento. Para ilustrar la relación que tienen los argumentos y las explicaciones describe un escenario plausible: un fusilero le pregunta a Galileo por qué su arma alcanza su rango máximo cuando el cañón se eleva a 45° del suelo, si este es plano. Como respuesta, Galileo reformula el problema idealizando las condiciones del escenario: asume un proyectil ideal, una velocidad fija, un plano perfectamente horizontal; las desviaciones en el mundo real de este escenario idealizado no son significativas, dice. Luego, emplea un argumento cinemático que concluye que en esas condiciones, el proyectil alcanza su rango máximo cuando el ángulo de elevación es de 45° . Le explica al fusilero los conceptos que no le son familiares (por ejemplo, el de ‘aceleración uniforme’), justifica los principios que emplea y simplifica la explicación omitiendo algunos pasos en el cálculo. Ambos quedan satisfechos. ¿Por qué? ¿Qué enseñanza se puede extraer de ejemplos como el de Galileo y el fusilero respecto al problema del análisis de la noción de explicación? ¿Qué puede hacer entender el ejemplo?

Kitcher propone que hay que describir el caso como si el argumento que Galileo emplea perteneciera a un arsenal de argumentos que tiene disponibles para propósitos explicativos (a este “arsenal” lo llama la *reserva explicativa*). La ciencia no ofrecería argumentos individuales para fundamentar los actos de explicación, sino que ofrecería una reserva de argumentos explicativos que se pueden utilizar cuando es necesario. La reserva explicativa que tenga disponible alguien depende de los enunciados que acepte; es una función del conjunto de ellos. ¿Qué rasgos tiene que tener un argumento que sirva de base para un acto de explicación? Tiene que ser a parte de la reserva explicativa de los enunciados que acepte quien lleve a cabo el acto. ¿En qué condiciones lo es?

Supóngase que K es el conjunto de los enunciados que acepta una comunidad de sujetos S . Entonces la reserva explicativa de K , $E(K)$ es el conjunto de argumentos aceptables por los miembros de S como la base de actos explicativos. La propuesta de Kitcher es que $E(K)$ es el conjunto de argumentos que mejor unifica K . ¿Qué significa esto?

Sobre un mismo conjunto de enunciados K podría ser el caso que varios conjuntos de argumentos puedan servir de base para derivar enunciados dentro de K . Para entender cómo podría pasar esto, considérese la situación análoga de las proposiciones formales: un sistema lógico como el cálculo proposicional clásico puede derivarse a partir de varios conjuntos de axiomas. Kitcher llama a un conjunto de argumentos que deriva a miembros de K de otros miembros de K una *siste-*

matización de K . $E(K)$ es, entonces, la mejor sistematización de K . Los argumentos que componen una sistematización de K deben ser *aceptables respecto a K* , es decir, argumentos válidos cuyas premisas pertenecen todas a K . Kitcher admite que esto es una idealización (porque en una explicación podrían usarse premisas que no se consideren como parte de K), y propone que quizás el conjunto de argumentos aceptables respecto a K deba complementarse con argumentos 1) cuyas premisas sean aproximadamente verdaderos desde la perspectiva de K , 2) que se aproximen estructuralmente a miembros de $E(K)$ (es decir, similares a los argumentos que mejor unifican K), y 3) más simples que los que les corresponden en $E(K)$.

Cuando se consideran las posibles sistematizaciones de K debe prestarse atención a los *patrones argumentales* que se emplean en cada sistematización. Recuérdate la cita de Kitcher (1976): cuando se observan casos como el de la teoría newtoniana, se puede apreciar que la unificación es producto “del uso repetido de leyes que relacionan grandes clases de fenómenos aparentemente diversos mediante unas propiedades fundamentales”. La unificación se logra usando argumentos similares en la derivación de muchos enunciados aceptados (Kitcher 1981, 519). Un ejemplo interesante que menciona Kitcher es el de la presentación original de la teoría de la evolución mediante selección natural de Darwin. En el *Origen de las Especies*, sugiere Kitcher, el argumento no procede mediante la caracterización detallada de los mecanismos que produjeron tal o cual rasgo en los individuos de una especie, sino mediante la presentación de “esbozos de explicación”, que exhiben un patrón argumental que puede instanciarse en principio por una derivación rigurosa de las características de los miembros de una especie particular. Así es como Darwin emplea experimentos mentales (cf. Darwin (1876, 86 y ss.)²⁰ e incluso de un diagrama (pág. 90) para “ayudar a comprender el *rather perplexing subject*” que es la acción de la selección natural en la “divergencia de carácter” (ibíd.)

²⁰Como referencia, este es el primero de estos experimentos mentales (destaco en cursiva el experimento mental propiamente dicho):

... the more diversified the descendants from any one species become in structure, constitution, and habits, by so much will they be better enabled to seize on many and widely diversified places in the polity of nature, and so be enabled to increase in numbers.

We can clearly discern this in the case of animals with simple habits. *Take the case of a carnivorous quadruped, of which the number that can be supported in any country has long ago arrived at its full average. If its natural power of increase be allowed to act, it can succeed in increasing (the country not undergoing any change in conditions) only by its varying descendants seizing on places at present occupied by other animals: some of them, for instance, being enabled to feed on new kinds of prey, either dead or alive; some inhabiting new stations, climbing trees, frequenting water, and some perhaps becoming less carnivorous. The more diversified in habits and structure the descendants of our carnivorous animals become, the more places they will be enabled to occupy.* What applies to one animal will apply throughout all time to all animals—that is, if they vary—for otherwise natural selection can effect nothing. (pág. 87)

Según Kitcher, el objetivo de Darwin en el *Origen* es entregar una teoría cuyo poder explicativo consiste en su poder unificador.

Aparte de la evidencia histórica a favor de la idea de que la unificación se realiza mediante la aplicación de patrones argumentales, Kitcher (1981, 515–516) observa que alguien que asume un argumento como explicativo tiene que asumir como explicativos también a otros argumentos que son similares a aquél, y que siguen el mismo patrón argumental. De otro modo, se diría que sus prácticas explicativas son anormales.

Para dar un modelo formal de la unificación en el sentido que acaba de delinearse, Kitcher desarrolla un arsenal de conceptos. Primero, necesita precisar su noción de patrón argumental. Para ello comienza definiendo un *enunciado esquemático* como una expresión que resulta del reemplazo de expresiones no-lógicas de un enunciado por variables (por ejemplo, del enunciado “los tigres son mamíferos” resultan los enunciados esquemáticos “los tigres son f ”, “los f son mamíferos” y “los f son ϕ ”). Que no se reemplacen necesariamente todas las expresiones no-lógicas tiene el resultado que un enunciado esquemático en este sentido impone ciertas restricciones respecto a su realizabilidad; para codificar esto se introduce el concepto de un conjunto de *instrucciones de llenado* para un enunciado esquemático, que es el concepto de un conjunto de direcciones para reemplazar las variables por ciertos valores (por ejemplo, que para “los f son mamíferos”, “ f ” puede reemplazarse por el nombre de una especie de animal con tales o cuales características). Un *argumento esquemático* es una secuencia de enunciados esquemáticos. Una *clasificación* de un argumento esquemático es un conjunto de enunciados que describe los rasgos inferenciales del argumento esquemático: cuáles son las premisas, qué reglas de inferencia deben usarse, etc. Un *patrón argumental general* es un triple ordenado que consiste de un argumento esquemático, un conjunto de conjuntos de instrucciones de llenado para los enunciados esquemáticos que componen el argumento esquemático y una clasificación del argumento esquemático.

Los argumentos que siguen un patrón argumental pueden ser similares o bien por su estructura lógica o por el vocabulario no-lógico que emplean. Kitcher observa que los argumentos que se usan en ciencia presentan compromisos respecto a la importancia de ambos elementos; en particular, dice, a los científicos les interesan patrones argumentales *estrictos* (*stringent*), es decir, patrones que contienen determinadas expresiones no-lógicas y que son similares estructuralmente. Mientras más restricciones impone un patrón argumental, más estricto es (el concepto de estrictez implica un orden porque codifica la similaridad de los argumentos que siguen un patrón).

Con esta noción de patrón argumental, Kitcher procede a definir un *conjunto generador* de un conjunto de argumentos Σ como un conjunto de patrones Π tal, que cada miembro de Σ es una instancia de Π . Si se tiene un conjunto de enunciados K , se dirá que un conjunto generador Π de Σ (en este caso Σ es el conjunto de argumentos en K) es completo respecto de K si y sólo si cada argumento que es aceptable respecto de K es una instancia de Π y pertenece a Σ . Para determinar la reserva explicativa de K , se consideran, como se apuntó antes, las sistematizaciones de K . De ellas,

se consideran a su vez los conjuntos generadores que son completos respecto de K . El conjunto generador que tenga el mayor poder unificador es la *base* de la sistematización. *La sistematización de K cuya base tenga el mayor poder explicativo es la reserva explicativa de K .*

¿Cómo se determina el poder unificador de la base de una sistematización? Kitcher (1989, 434) propone que deben considerarse al menos dos factores: que se derive la mayor cantidad de conclusiones a partir de un menor número de patrones, y que estos sean estrictos (en el sentido delineado antes). Así, el poder unificador de un conjunto generador de una sistematización varía directamente con el tamaño de las conclusiones de ella, directamente con la estrictez de los patrones del conjunto e inversamente con la cantidad de patrones (pág. 435).

Queda por verse cómo una teoría como esta puede acomodar las formas de entendimiento que se caracterizaron en la sección anterior (en particular, cómo puede darse un análisis de las nociones de entendimiento objetual y entendimiento subjetivo en el marco de una teoría como las de Kitcher). El modelo de Kitcher está enfocado a dar una descripción y explicación de las prácticas explicativas científicas, lo que deja abierta la cuestión si puede emplearse para construir conceptos como el de entendimiento filosófico, que como se apuntaba en la primera parte de este capítulo, son importantes para resolver el problema de la corrección del modelo metafilosófico que se ha propuesto. En la sección siguiente voy a tratar de explorar las posibilidades de emplear un modelo unificacionista del entendimiento para iluminar el problema del juicio de posibilidad; allí deberá verse cómo pueden resolverse estos puntos. Las críticas al modelo que sean relevantes también serán examinadas allí.

3.2.3. LA CONSTITUCIÓN DEL JUICIO DE POSIBILIDAD EN EL ENTENDIMIENTO

La propuesta que se quiere hacer es que el juicio de posibilidad depende del entendimiento, en el sentido de que es necesario para un juicio de posibilidad de que “es posible p ” *entender* algo que guarda una determinada relación con p . El problema es determinar cómo caracterizar esta supuesta relación y qué podría guardar esa relación con p en forma general.

Primero, sin embargo, es necesario *motivar* esta postura. ¿Por qué tendría que decirse que el juicio de posibilidad tiene algo que ver con el entendimiento? ¿Y por qué asumir *además* que el entendimiento puede fundamentar el juicio de posibilidad?

Supóngase que dos personas piensan que es necesario mover un cierto objeto, demasiado pesado como para que lo hagan sin herramientas ni por separado ni juntos. Disponen de un martillo y de una soga. Uno de ellos tiene la creencia correcta de que pueden mover el objeto empleando la soga; el otro no tiene esa creencia. ¿Quién de ellos tiene el mayor grado de entendimiento, si sólo se los puede distinguir por este hecho? Creo que habría que decir que el primero tiene mayor entendimiento. En general, poder emitir juicios de posibilidad parece ser indicio de la posesión de entendimiento. ¿Qué tipo de entendimiento, y de qué? ¿Es esta una conexión constante? Para responder esta pregunta no es suficiente proporcionar más ejemplos en que se mantenga esta relación entre el juicio de posibilidad y el entendimiento. Es necesario tener una teoría que explique esta

relación. Esta teoría puede desarrollarse en dos direcciones: explicando el entendimiento como resultado de los procesos cognitivos que justifican el juicio de posibilidad, o explicando la justificación de los juicios de posibilidad como resultado del proceso mediante el que se gana entendimiento. Esta segunda opción puede motivarse por las siguientes consideraciones.

Entender objetualmente una estructura implica poder juzgar cuáles son sus estados posibles. Por ejemplo, alguien que entienda el ajedrez (que es un objeto que tiene reglas por estructura) puede juzgar que en una partida, es posible que un peón puede capturar al alfil enemigo si se dan ciertas condiciones; en efecto, podría querer decirse que entender el juego es necesario para poder construir estrategias y tácticas eficientes porque implica poder considerar escenarios posibles, que a su vez son constitutivos de esas estrategias y tácticas (“constitutivos” en el sentido en que, por ejemplo, poder adoptar una táctica consiste en poder reconocer ciertas condiciones y resolver un plan de acción *posible*). Así, juzgar que tal o cual evento o estado es posible podría ser resultado de entender una estructura de la que es parte (la malla causal del mundo, por ejemplo, que tiene a las leyes naturales como su estructura).

Del mismo modo, entender por qué ocurrió algo permite *construir* escenarios similares en que ocurra lo mismo. Si entiendo que la estrategia napoleónica de atacar Rusia en invierno falló porque los generales a cargo de la operación no consideraron apropiadamente las dificultades que trae consigo el clima de la región, puedo juzgar que en condiciones similares, una estrategia similar podría fallar por esas mismas razones. Si no pudieran hacerse juicios como estos, no diríamos que se entiende por qué falló la estrategia. Pero al mismo tiempo, no parece que la posibilidad de hacer estos juicios implique que se entienda el fallo de la estrategia. Esto no significa que el entendimiento no pueda ser la fuente de estos juicios: podrían depender de un cierto entendimiento objetual (como en el caso anterior). Pero aún así, podría ser que juicios particulares sobre esas posibilidades dependan de que se tuviera cierto entendimiento de por qué (en este caso, entendimiento de por qué falló la estrategia napoleónica).

3.2.3.1. *Un antecedente: El confiabilismo modal de Bealer*

El *confiabilismo modal* de Bealer (2002) puede considerarse como un antecedente de una teoría epistemológica sobre el juicio de posibilidad que emplea el concepto de entendimiento. Hay que hacer la observación de que Bealer habla de entendimiento (*understanding*) en al menos dos sentidos: primero, en el sentido en que *el* entendimiento es una suerte de facultad cognitiva general en que se generan juicios sobre la base de la aplicación de conceptos, siguiendo la tradición clásica de, por ejemplo, Locke, Hume y Kant, y en segundo lugar, como una forma de entendimiento objetual objetivo en el sentido descrito en la sección anterior—por ejemplo, cuando habla de “entender un concepto”. A partir de las observaciones de Bealer (1987) podría esbozarse una manera de conectar estos sentidos de “entendimiento”: el entendimiento es la facultad general de entender objetualmente y de manera objetiva la aplicabilidad de conceptos.

Según Bealer, esta facultad rastrea la verdad, de modo que se sigue lo que llama la *tesis del racionalismo moderado*: necesariamente, la mayoría de las creencias no-inferenciales de una persona sobre la aplicabilidad de un concepto a casos hipotéticos elementales son verdaderos.²¹ Bealer restringe la tesis a las creencias no-inferenciales porque busca defender las intuiciones sobre casos hipotéticos como una genuina fuente de evidencia (ibíd, págs. 320–321). Las creencias que defiende la tesis del racionalismo moderado, de acuerdo a él, tienen que sostenerse sin importar cómo sea el mundo o las circunstancias de quien las tenga: lo único que importa es que se deben a la posesión de determinados conceptos:

These are the beliefs that any normal person would have if the person truly has the concept. If a person seems to us not to make mostly true judgments about the applicability of the concept to elementary hypothetical cases, that would indicate, not that the person is mistaken, but rather that the person does not really have the concept to begin with. (pág. 320)

Bealer (2002) llama a esta variedad de posesión de conceptos “entendimiento determinado de un concepto” (*determinately understanding of a concept*). Para que un sujeto entienda un concepto determinadamente debe

- a) poseer actitudes proposicionales naturales hacia proposiciones que tienen el concepto como contenido constitutivo, y
- b) poseerlas sin entender erróneamente (esas actitudes proposicionales deben ser correctas; por ejemplo los juicios deben ser verdaderos) ni de forma incompleta (se tiene actitudes proposicionales naturales sólo dentro de ciertos rangos dentro del campo de aplicabilidad del concepto), o meramente en virtud de las prácticas atributivas de terceras personas (se siguen las atribuciones de otras personas), o de una forma similar.

Para clarificar su concepto de entendimiento determinado, emplea la siguiente ilustración:

Suppose that in her journal a sincere, wholly normal, attentive woman introduces thorough use (not stipulation) a new term ‘multigon’. She applies the term to various closed plane figures having several sides (pentagons, octagons, chiliagons, etc.). Suppose her term expresses some definite concept — the concept of being a multigon — and

²¹En Bealer (1996), propone

For suitably good cognitive conditions, it is necessary that, if while in such conditions a subject goes through the whole procedure of *a priori justification, then most of the propositions derivable from the resulting comprehensive theoretical systematization of the subject’s intuitions would have to be true.

(pág. 130)

that she determinately understands this concept. By chance, she has neither applied her term ‘multigon’ to triangles and rectangles nor withheld it from them; the question has just not come up. Eventually, however, she considers it. Her cognitive conditions (intelligence, etc.) are good, and she determinately understands these concepts. Suppose that the property of being a multigon is either the property of being a closed, straight-sided plane figure, or being a closed, straight-sided plane figure with five or more sides. Then, intuitively, when the woman entertains the question, she would have an intuition that it is possible for a triangle or a rectangle to be a multigon if and only if being a multigon = being a closed, straight-sided plane figure. Alternatively, she would have an intuition that it is not possible for a triangle or a rectangle to be a multigon if and only if being a multigon = being a closed, straight-sided plane figure with five or more sides. That is, the woman would have truth-tracking intuitions. If she did not, the right thing to say would be that either the woman does not really understand one or more of the concepts involved, or her cognitive conditions are not really ideal. (pág. 103)

El pasaje sugiere que el entendimiento determinado tiene un enlace fuerte con la posibilidad metafísica: la mujer *sabe* que un triángulo puede ser un multígono dadas ciertas condiciones. Bealer acepta que juicios intuitivos como este podría fallar, pero en principio podrían corregirse si las condiciones cognitivas del sujeto mejoraran, o si aumentara su repertorio conceptual. El entendimiento determinado, nos dice, puede ser explicado en términos de la posibilidad metafísica de intuiciones que rastreen la verdad: es la base categórica de esa posibilidad (pág. 104).

El análisis más preciso que da Bealer de la noción de entendimiento determinado es el siguiente:

entendimiento determinado = el modo m del entendimiento tal que, necesariamente, para todo x y para todo p entendido a la manera de m por x ,

- i) p es verdadero si es posible que x determine con estabilidad *a priori* que p es verdadero.
- ii) p es verdadero sólo si es posible que x determine con estabilidad *a priori* que p tiene una contraparte verdadera (aquí p es una identidad entre propiedades).
- iii) p es verdadero sólo si es posible que x crea a la manera de m que p es verdadero (aquí p es algo que x puede crear).

La primera cláusula quiere decir que una vez que un sujeto determina que p es verdadero, siempre se juzgará igual mientras el sujeto siga entendiendo determinadamente que p (a esto refiere la expresión “estable *a priori*”). Los juicios empíricos no son estables *a priori*; por ejemplo, no lo es el juicio “me siento feliz”: es verdadero a veces y otras veces es falso. En cambio, enunciados como “ $1 + 1 = 2$ ” o “yo soy yo” son estables *a priori*. Para que un enunciado sea estable *a priori* es necesario que el sujeto que lo determine satisfaga cierto grado de capacidad cognitiva, y posea un repertorio conceptual de cierta amplitud; si el sujeto tiene sus capacidades disminuidas, puede no poder entender de modo determinado el enunciado aunque este sea en principio estable.

La segunda cláusula quiere decir que para que un sujeto entienda determinadamente que p es necesario que el sujeto pueda determinar que en un mundo posible hay una contraparte de p que es verdadera. El objetivo de esta cláusula es establecer que si p se entiende de modo determinado, no se lo puede entender solamente *a posteriori*. Si alguien satisface la 1a y 2a cláusulas, se puede decir que tiene *dominio categorial* de un concepto.

La tercera condición, según Bealer, significa que en la ausencia de la intuición, la determinación del entendimiento recurre a la malla de creencias que tenga el sujeto. Bealer asume que x sólo puede creer p si el resto de sus creencias se lo permite, y que poseer un concepto con entendimiento significa poder conectarlo con otros conceptos. Por ejemplo, incluso si alguien tuviera dominio categorial de los conceptos de “haya” y “olmo” podría no ser capaz de determinar si ciertos árboles son hayas o elmos (en este caso no se diría que entiende esos conceptos determinadamente). Pero si adquiriera nuevas creencias podría lograr determinarlo, y obtener entendimiento estable sobre ello.

¿Cómo puede emplearse el concepto de entendimiento determinado para explicar la justificación del juicio de posibilidad? La verdad de los juicios de posibilidad correctos depende de que se los entienda determinadamente. Como se ha mencionado, Bealer considera que la mayoría de las intuiciones que se tengan sobre la aplicabilidad de un concepto, mientras se entiendan esos conceptos de manera determinada, son confiables. El contenido de esas intuiciones *es* el contenido de juicios modales: un árbol de tales o cuales características podría ser un elmo (tenemos la intuición de que es así, y es una intuición sobre la aplicabilidad del concepto de elmo), un científico que no ha visto por sí mismo el color rojo no puede entender la experiencia cualitativa (una intuición sobre la aplicabilidad de los conceptos de experiencia cualitativa y entendimiento), etc. ¿Qué hace que *esas* intuiciones/juicios de posibilidad sean confiables, si lo son? ¿Por qué ciertas intuiciones y no otras? Bealer propone que depende del tipo de conceptos que estén en juego: en particular, los conceptos de categoría (predicación, número, identidad, propiedad, relación, cualidad, cantidad, cosa, cosa compuesta, cosa funcional, etc.) y los conceptos de contenido (sobre conceptos fenomenales y de actitudes psicológicas) sirven como soporte a las intuiciones porque son intrínsecamente *estables semánticamente*: una expresión de un concepto de este tipo significa lo mismo para cualquier sujeto cuya situación epistémica es cualitativamente idéntica a la nuestra (cf. Bealer (2000)). La expresión “dolor” es semánticamente estable porque, si el término se usa para referir a una cualidad específica, alguien que esté en esa misma situación epistémica tendría que usar el término para referirse a la misma cualidad. Al contrario, los términos de clase natural son inestables semánticamente. Esto significa que puede tenerse conocimiento de necesidades sobre esos conceptos *a posteriori*.

Bealer (2002) da varios ejemplos de cómo las intuiciones sobre la aplicabilidad de estos conceptos da pie a juicios de necesidad *a priori* y *a posteriori*; en realidad, intenta explicar cómo se puede pasar de cierta información empírica a conclusiones modales sobre necesidades (cf. pág. 107). En estos casos asume que los juicios de posibilidad son básicos. Supóngase, por ejemplo, que se entiende el concepto de categoría de cosa compuesta, de modo que se puede asentir al siguiente

principio: si una muestra de una cosa compuesta tiene la composición c , todas las muestras de esa misma cosa tendrán la composición c . Si es así, si se imagina una situación en que se aplica el principio, las intuiciones que se tengan sobre ella serán confiables (por ejemplo, que una muestra de agua en un mundo posible cualquiera debe tener la composición H_2O , si entiendo que el agua es una cosa compuesta y sé que en el mundo real el agua tiene la composición H_2O). El peso evidencial de la intuición no se sustenta en la estabilidad del término “agua”, sino en la estabilidad del principio categorial que se emplea. Del mismo modo, puede ocurrir que no se pueda tener intuiciones sobre ciertos escenarios hipotéticos porque no hay conceptos que permitan entender determinadamente esos escenarios (como ejemplo, Bealer propone que no se puede tener intuiciones según las cuales una muestra de agua en un mundo posible no tenga la composición H_2O .)

En general, un juicio de posibilidad equivocado aparece cuando no hay entendimiento de los conceptos que se aplican en el escenario que propone. Sin embargo, Bealer sugiere que esto no es suficiente para explicar la totalidad del rango de errores en el juicio modal. ¿Cómo se puede explicar que sujetos de los que puede decirse que entienden ciertos conceptos tengan intuiciones contrarias sobre su aplicación (por ejemplo, que dos sujetos que entienden el concepto de “agua” tengan intuiciones opuestas sobre si una muestra de agua podría tener una composición distinta a H_2O)? Bealer responde que estos errores se deben a fallos locales en el entendimiento, en que simplemente no se sabe aplicar correctamente un concepto en un caso determinado. Este tipo de error puede corregirse, sugiere, mediante la consideración de otros casos en que se aplica el concepto y la sistematización de las intuiciones que se obtengan mediante ese examen.

¿Se puede aceptar la tesis del confiabilismo modal de Bealer? Es plausible que nuestros juicios de posibilidad sean el producto de nuestra capacidad de entender conceptos, pero no es plausible que entenderlos baste para justificar que esos juicios de posibilidad expresen posibilidades “reales” (que, por ejemplo, en el mundo real pueda ocurrir tal o cual cosa). Me parece plausible que, por ejemplo, alguien pueda entender el concepto de mónada sin que pueda juzgar que es posible que existan entidades que caigan bajo esa categoría. El concepto de mónada es, sin embargo, semánticamente estable y, de acuerdo al argumento de Bealer, las intuiciones que podemos tener de él son confiables (esto parece ser al menos parcialmente correcto). En este sentido, la mayoría de los supuestos juicios de posibilidad no son confiables: no se puede garantizar que expresen juicios sobre posibilidades del mundo real. Podemos tener entendimiento subjetivo que nos haga creer que tal o cual cosa es posible realmente, pero este entendimiento subjetivo debe complementarse por algo más para volverse objetivo: debe entenderse el “alcance” del juicio, esto es, sobre qué mundos posibles es verdadero y cuál es la relación de esos mundos posibles y el mundo real. Así pues, meramente tener intuiciones sobre casos hipotéticos no justifica los juicios de posibilidad que les corresponden.

Supóngase que, en efecto, Bealer está en lo correcto. ¿Es su teoría lo suficientemente general como para explicar la justificación de todos los juicios de posibilidad? Podría ser que no. Además de las creencias no-inferenciales sobre la aplicabilidad de conceptos en situaciones hipotéticas, existen

también otras bases para el juicio de posibilidad; por ejemplo, en el caso en que se infiere que algo es posible. Ahora bien, ¿es necesario que la teoría de Bealer explique la justificación de todos los juicios de posibilidad? Nuevamente, podría ser que no. Como él mismo admite, su objetivo es defender a las intuiciones como una fuente válida de evidencia y conocimiento modal; pero podría haber otras fuentes para este. Así pues, sin abandonar los aspectos aceptables de la teoría de Bealer, se podría apelar a ella como una explicación de una de esas fuentes del conocimiento modal (a saber, como una explicación de las intuiciones modales). Quizás el poder explicativo de la teoría de Bealer pueda explicarse él mismo mediante una teoría más general sobre la justificación del juicio de posibilidad.

3.2.3.2. Unificación y juicios de posibilidad

¿Qué es lo que hay que entender para poder afirmar que algo es posible? Bealer sugiere que hay que entender los conceptos que se poseen. Sin embargo, esto no nos permite distinguir cómo rastreen nuestros juicios la verdad; en particular, no permite distinguir el alcance modal de los juicios de posibilidad. Lo que hace falta en el análisis es, me parece, distinguir entre formas subjetivas y objetivas de entendimiento, e incorporar la idea de que el entendimiento puede ser gradual. Como resultado, debería obtenerse una teoría sobre el juicio de posibilidad que acomode la distinción entre posibilidades subjetivas y objetivas, y la idea de que pueden especificarse ciertas jerarquías entre varios conceptos de posibilidad (por ejemplo, que hay un tipo de posibilidad metafísica que es un subconjunto de un tipo de posibilidad lógica). Lo que se busca es un modelo general de la conexión entre el entendimiento y el juicio de posibilidad.

Lo que quiero presentar ahora es un modelo de cómo se justifican los juicios de posibilidad. Las ideas centrales de este modelo son:

- 1) los juicios de posibilidad objetivos se justifican por el entendimiento objetivo de los escenarios que describen los contenidos de esos juicios,
- 2) el entendimiento consiste en un proceso de unificación,
- 3) la creación de escenarios hipotéticos no implica la posibilidad objetiva de esos escenarios (esto es consecuencia de 1 y 2: no es necesario que los escenarios hipotéticos que se construyan estén unificados).

Un juicio de posibilidad *objetivo* en el sentido que se emplea el término aquí es un juicio de posibilidad que es verdadero en el mundo real. En oposición a esto, un juicio de posibilidad subjetivo es verdadero en el contexto de un determinado conjunto de creencias de quien juzga. Un juicio de posibilidad subjetivo puede ser también objetivo; por ejemplo, el juicio de que “tal o cual teoría podría ser válida” es un juicio subjetivo (expresa la compatibilidad de la validez de la teoría y las creencias aceptables del hablante) que podría ser objetivo si, de hecho, la teoría puede ser válida sobre el mundo real.

Cuando alguien imagina un escenario hipotético, ese sujeto se pone en una posición epistémica en que podría hacer un juicio de posibilidad sobre ese escenario (a saber, que *el escenario* es posible subjetiva u objetivamente). De este modo, se puede distinguir entre considerar un caso hipotético y considerar una posibilidad. Un juicio de posibilidad consiste en la evaluación de la posibilidad de un caso hipotético. Las condiciones en que se construye el escenario no explican esa posibilidad (es decir, que un escenario sea, por ejemplo, concebido o imaginado no justifica el juicio de que el escenario es posible); el contenido de un juicio de posibilidad puede construirse de diversas maneras, y su grado de posibilidad es independiente de cómo se lo construya. Sin embargo, como se verá luego, la manera en que se construye el escenario puede ser un indicio del tipo de justificación que podría tener el juicio de posibilidad que le corresponde: que se pueda imaginar que un artefacto de ciertas características produciría tal o cual efecto al funcionar puede ser indicio, por ejemplo, de que es posible que el funcionamiento de ese artefacto puede resultar en esos efectos, pero no necesariamente explica por qué ese juicio estaría justificado.

¿Cómo explicar la justificación de los juicios sobre esos casos hipotéticos? La propuesta que voy a hacer aquí es que lo que los justifica es la capacidad de unificarlos con el conocimiento sobre el mundo real. La capacidad humana de juzgar correctamente, me parece, está sólidamente fundada en las condiciones en que se tiene conocimiento del mundo circundante. Como Nozick (2001) y otros han sugerido, esta facultad humana es producto del proceso de adaptación de los seres humanos al medio. Es sólo mediante abstracción que construye la idea de que tenemos conocimiento modal sobre lo que podría ser posible en otros mundos posibles aparte del real. Pero esos mundos posibles son o deberían ser maneras como el mundo real podría ser.

No todos los casos hipotéticos pueden unificarse con el conjunto de los hechos en el mundo real. Esto sugiere que para justificar juicios de posibilidad objetivos tienen que ponerse restricciones sobre la relación entre el caso hipotético y el mundo real. Una condición que puede proponerse es

U_1 El caso hipotético p es posible si y sólo si la descripción de p y la descripción del mundo real son unificables.

Esta condición logra capturar al menos los casos cotidianos de juicios de posibilidad aceptables. Por ejemplo, permite justificar el juicio de que Felipe Morales puede ayudar a su madre con las compras de la semana, porque la descripción del caso hipotético es unificable con la descripción del mundo real.

Un problema con U_1 es que es demasiado liberal. Distintos mundos posibles podrían unificarse con el mundo real sin poder unificarse entre sí; pero parece necesario que los juicios de posibilidad aceptables puedan unificarse entre sí por un vínculo más fuerte que el de su capacidad de ser unificables con la descripción del mundo real. Esto sugiere

U_2 El caso hipotético p es posible si y sólo si la descripción de p es unificable con la suma de la descripción del mundo real y el conjunto de las descripciones de casos hipotéticos que son

aceptables.

Esta condición debe interpretarse en un sentido dinámico. Es plausible que el conjunto de los escenarios hipotéticos crezca constantemente. De esas posibilidades, se seleccionan aquellas que mejor se unifiquen entre sí y con la descripción del mundo real.

Esto aún podría ser demasiado liberal. La descripción del mundo real puede sistematizarse de varias maneras, y es plausible que se pueda unificar casos hipotéticos con varias de esas sistematizaciones. Quizás, para justificar un juicio de posibilidad hay que unificar su contenido con la mejor sistematización del mundo real. Esto sugiere

U_3 El caso hipotético p es posible s y sólo si la descripción de p es unificable con la mejor sistematización de la suma de la descripción del mundo real y el conjunto de las descripciones de casos hipotéticos que son aceptables.

Esta alternativa es atractiva porque es plausible que la mejor sistematización de los enunciados sobre el mundo real pueda soportar *naturalmente* casos contrafácticos. La evidencia a favor de esto radica en que una sistematización puede explicar y soportar predicciones y retrodicciones. El caso en que se juzga correctamente que algo que aún no ha pasado va a pasar, y el caso en que se juzga correctamente que algo que no es verdad en el mundo real podría pasar algo son análogos y quizás estén conectados.²² Si puedo decir que va a llover mañana, puedo decir que es posible que llueva mañana (pero no viceversa).

Hay un sentido, sin embargo, en que podría querer decirse que U_3 es demasiado estricto: también es plausible que sistematizaciones que no sean la mejor soporten juicios de posibilidad correctos, que sin embargo no puedan justificarse bajo la mejor sistematización *disponible*. Al considerar cuál es la mejor sistematización disponible hay que considerar muchos factores, y puede ocurrir que las ventajas de aceptarla superen el problema de que no pueda justificar la totalidad de los juicios de posibilidad intuitivamente correctos. Ahora bien, es probable que *en principio* la mejor sistematización pueda justificar esta totalidad de juicios, y es plausible que la mejor sistematización disponible justifique más juicios de posibilidad que ninguna otra (incluso, es posible que justifique todas los juicios que las otras sistematizaciones disponibles justifican correctamente). Por otro lado, quizás cuando aparezcan este tipo de limitaciones en una sistematización (i.e. que una sistematización s_1 no pueda justificar todos los juicios de posibilidad que otra sistematización s_2 , aún cuando s_1 es mejor que s_2) pueda corregirse incorporando los elementos de s_2 que justifican los casos problemáticos a s_1 . Si esto es posible en principio y sin alterar la corrección de la sistematización de s_1 , lo que es plausible, no es necesario modificar U_3 .

U_1 , U_2 y U_3 representan tres distintos niveles de justificación para un juicio de posibilidad, o más bien etapas mediante las que se obtiene justificación para él. Esto sugiere que la distinción entre

²²La afinidad entre las categorías de la modalidad y la temporalidad requieren de una explicación independiente, que no se puede dar aquí.

los juicios de posibilidad subjetivos y los juicios objetivos es borrosa y gradual; intuitivamente, la objetividad de esos juicios depende de que se satisfagan condiciones más estrictas.

Tómese el siguiente enunciado para ilustrar estas observaciones:

a) Los unicornios son posibles

Según el modelo propuesto, el juicio de posibilidad que expresa el enunciado sólo podría ser correcto si el caso hipotético en que existen unicornios y la mejor sistematización de la suma de los casos hipotéticos aceptables y la descripción del mundo real son unificables. Ahora bien, por razones como las que da Kripke (1980), no es razonable pensar que un mundo posible en que existan unicornios sea unificable con el mundo real: en el mundo real, los unicornios *son* seres ficticios, y sobre los seres ficticios se dan varias condiciones que impiden pensar que podrían existir, incluso si se puede imaginar un caso hipotético en que existan seres con todas las propiedades que se han atribuido a ellos (i.e., los seres ficticios son esencialmente ficticios). Puede describirse esta situación desde otro ángulo: no se podría entender cómo un ser ficticio podría existir en el mundo real, lo que significa a su vez que el caso hipotético tampoco puede unificarse con otros mundos posibles aceptables.

¿Es el modelo demasiado complicado como para acomodar las características del juicio de posibilidad cotidiano, que por lo general parece requerir de poco esfuerzo (por ejemplo, cuando se juzga, presumiblemente de manera objetiva, que es posible que se haga algo, como puede ser lavar la loza, escribir un ensayo, etc.)? La idea de que en cada situación en que se emita un juicio de posibilidad tenga que evaluarse la unificabilidad del caso hipotético que presenta y conjuntos potencialmente infinitos de enunciados (o de leyes de las que pueden derivarse esos enunciados) es difícil de aceptar. Creo que en la mayoría de los casos cotidianos las condiciones que justifican los juicios se dan por sentado. Creo que hay dos mecanismos a los que se puede apelar para explicar esto. En primer lugar, es posible que al evaluar la corrección del juicio se considere si juicios pasados similares al que se evalúa han sido correctos, y que se razone inductivamente que si esos lo son, también lo es este. Otra posibilidad es que sobre ciertas clases de enunciados sobre temas de los que se tiene entendimiento objetivo (como parecen ser los juicios sobre posibilidades físicas o mecánicas, sobre las que se tienen modelos explicativos que entregan gran cantidad de entendimiento) se reduzca el estándar de justificación o se emplee una norma menos estricta que la de la verdad para justificar la afirmación del juicio (podemos no saber que tal o cual cosa sea verdaderamente posible, y aún así tener justificación para afirmarlo).²³ En algunos casos el entendimiento subjetivo de algo es evidencia suficiente para justificar los juicios de posibilidad.

²³Por ejemplo, MacFarlane (2003) propone que el valor de verdad de una afirmación varía según el contexto de evaluación del juicio. Él utiliza este modelo de la norma de la afirmación para explicar cómo juicios como “habrá una batalla naval mañana” pueden tener varios valores de verdad en varios contextos, sin que el no poder saber que el juicio es verdadero en un momento implique que no se puede afirmar. MacFarlane propone que una afirmación compromete a poder dar una justificación del enunciado en el futuro.

Se puede proponer el siguiente criterio heurístico para determinar qué clases de enunciados permiten este tipo de justificación de hecho. La capacidad de hacer modelos “ejecutables” de los casos hipotéticos es evidencia de que se los entiende. Un modelo ejecutable puede describirse como una serie de pasos precisos que llega a un resultado específico en que se satisface la situación que el caso hipotético propone (además, puede permitir derivar qué es lo que sucedería como consecuencia de que se satisfaga el caso hipotético). La capacidad de construir esta clase de modelo es evidencia de la riqueza de nuestro conocimiento y entendimiento de los conceptos involucrados en el caso hipotético, y es evidencia de que se satisfacen las condiciones de justificación de los juicios de posibilidad que se han propuesto porque la noción de ejecutabilidad tiene cierta relación conceptual con la de actualidad: el mundo real es donde se realizan los procesos que invariablemente resultan en situaciones efectivas (por definición), y por lo tanto, que dan pie a juicios verdaderos. El concepto de ejecutabilidad, por lo demás, puede aplicarse de maneras distintas según el tópico de que se trate: se puede hablar de cierta ejecutabilidad mecánica (en que los procedimientos son mecánicos), ejecutabilidad física (en que los procedimientos pueden registrar efectos), ejecutabilidad lógica (en que el procedimiento es la prueba del enunciado), etc. En todos estos casos, lo que explica la ejecutabilidad es la capacidad de unificar el caso hipotético.

Me gustaría ahora examinar dos problemas, que están relacionados.

¿Por qué no existe la práctica de abandonar los modelos que no han probado ser ejecutables? Por ejemplo, en filosofía a menudo se presentan casos hipotéticos como posibles aunque no existen modelos de ejecutabilidad para justificar los juicios de posibilidad correspondientes. Un ejemplo (intencionalmente polémico): se ha propuesto que es posible que una computadora tenga conciencia. Pero no se tiene ningún modelo de ejecutabilidad satisfactorio que permitiría justificar esta posibilidad. Algunos de los proponentes de esta idea sugieren que no se puede descartar la posibilidad, y que la ciencia “algún día entregará” los modelos que son necesarios. ¿Por qué no se descarta el caso hipotético, si no parece tenerse justificación para mantenerlo? Quizás porque la hipótesis se emplea como un elemento subjetivamente unificador para la visión de mundo, por así decirlo, de las personas que la defienden. La posibilidad objetiva de que una computadora posea conciencia sería evidencia de la unificabilidad de muchísimos fenómenos de los que en el presente no se tiene una explicación unificada. El caso hipotético se sostiene porque se cree que una teoría de la que se pueda derivar esa posibilidad sería la mejor sistematización de la descripción del mundo real y de los otros mundos posibles aceptables, o al menos mejor que las que se tienen de momento. Esto puede dirigir la acción de un agente epistémico, pero no justifica sus creencias ni hace permisibles sus afirmaciones. Como regla general, si el sujeto que afirma no puede dar él mismo evidencia positiva de la justificación de sus juicios en algún momento (en vez de delegar esa evidencia a otros), sus afirmaciones no pueden ser aceptadas.

¿Cómo evaluar la situación de casos hipotéticos que no estén unificados? ¿Debería decirse que esos casos hipotéticos no son objetivamente posibles? ¿Debería decirse tan sólo que esas posibili-

dades no pueden afirmarse? En este modelo, lo objetivamente imposible es lo que necesariamente no puede ser unificado (por defecto, no produce ni siquiera entendimiento subjetivo). ¿Cómo puede justificarse un juicio de que algo no puede ser unificado necesariamente? Que un caso hipotético no esté unificado no significa que sea objetivamente imposible. Pero si no se puede afirmar su imposibilidad, ¿cómo podría haber juicios objetivos de imposibilidad? Podría ser que una sistematización momentáneamente inaccesible a un sujeto pueda justificar el juicio de que ese caso hipotético es posible. En este caso no podría descartarse objetivamente ningún caso hipotético. Si no se tienen criterios para descartar casos hipotéticos, podría querer introducirse elementos imposibles como elementos unificadores; esta es, en efecto, una versión modal del infame argumento del “Dios de los huecos” (en que de la premisa de que hay una brecha en el conocimiento de las causas naturales de un fenómeno se deriva que la causa debe ser supernatural).

Quizás pueda resolverse la dificultad haciendo las siguientes observaciones. Lo que se acepte como posible debe ser incompatible con algunos casos hipotéticos. Así, pues, se puede decir que algo es al menos subjetivamente imposible. La incompatibilidad con lo objetivamente posible sugiere una imposibilidad objetiva. El problema sólo aparece si se acepta que la sistematización que sirve de base para los juicios de posibilidad puede reemplazarse completamente. Pero esto es implausible; es probable que muchos de sus elementos sean estables, e incluso, que no pueda aceptarse una sistematización a menos que se asuman esos elementos.

3.3. OBSERVACIONES FINALES

Ahora es conveniente retomar la discusión de las cuestiones metafilosóficas. Si, como se propuso, el objetivo de la filosofía es la obtención de entendimiento y la actividad filosófica consiste en la exploración de espacios conceptuales, ¿qué se puede decir sobre ello la epistemología del juicio de posibilidad que acaba de ser esbozada?

Si no se tiene entendimiento, ¿cómo decidir qué es realmente posible?, ¿cómo podemos esperar entender qué es la verdad, cuál de los incontables escenarios que podemos concebir e imaginar es el correcto, o incluso *si* alguno lo es? El esbozo de una epistemología del juicio de posibilidad que se ha trazado nos permite responder estas preguntas: no se puede hacer ninguna de esas cosas sin entendimiento objetivo. ¿Cómo es que se llega a la condición epistémica en que podemos considerar opciones sobre “cómo habría sido el mundo”? Mediante la construcción de modelos sobre los que se pueden hacer preguntas cuya respuesta puede aumentar nuestro entendimiento subjetivo; llamaré “básico” a este tipo de entendimiento, porque toda otra forma de entendimiento debe desarrollarse por sobre este nivel. En un sentido preciso, estos modelos no necesitan ser considerados como objetivamente posibles. Determinar su posibilidad concreta es algo que debería ganarse, por así decirlo, mediante cierto esfuerzo. Las afirmaciones de la forma “es posible que...” no expresan necesariamente, como consecuencia, juicios verdaderos que expresen conocimiento; la norma de la afirmación de estos juicios no es su verdad, sino tan sólo la garantía de que asumiendo sus contenidos

se puede ganar cierto entendimiento. En consecuencia, la actitud que se debe tomar frente a las afirmaciones de posibilidad es la de la suspensión del juicio (en el sentido de que no deben afirmarse ni negarse), a menos que se tengan razones para creer que incorporando sus contenidos al conjunto de creencias que deben sostenerse puede obtenerse algún grado de entendimiento objetivo (por ejemplo, si se tienen razones para creer que incorporándolos se puede obtener una sistematización de las creencias que es cualitativamente mejor que las que hay disponibles sin que se los incorpore); e incluso en ese caso no es necesariamente obvio que se pueda afirmar que el contenido de los juicios es verdadero.

A causa de la subjetividad del entendimiento filosófico básico, no puede decirse que los juicios filosóficos tengan necesariamente como base el mundo real. Esto no socava el valor *a priori* de estos juicios, sino sólo su *alcance* (pueden ser juicios verdaderos sobre otros mundos posibles e incluso sobre mundos propiamente imposibles, aunque este último punto introduce complicaciones adicionales). Dicho de otra manera: no se sabe necesariamente sobre qué mundo son verdaderos esos juicios, pues la posibilidad de esos escenarios se evalúa desde su propia perspectiva, y el que se unifiquen o no, o en qué grado, con la descripción del mundo real es ortogonal (es decir, separable) a la cuestión de si son “realmente” posibles; por lo mismo, no se los puede afirmar como hechos sobre el mundo real (esto es importante porque a menudo se apela a esos casos contrafácticos para explicar hechos del mundo real; la posibilidad de la velocidad de la luz fuera distinta, por ejemplo, no tiene ningún significado a menos que pueda decir que el hecho posible es posible *en este mundo*). Todo esto sugiere, me parece, que la pregunta de si la filosofía debe responder a cuestiones sustantivas sobre la realidad es, por lo tanto, meramente interna a las teorías metafísicas prescriptivas; no hay una respuesta objetivamente correcta, porque hay investigaciones posibles viables (que sean válidas, o incluso valiosas, es otra cuestión)²⁴ en las que es irrelevante si los conceptos se aplican o no en el mundo real (o incluso, si pueden aplicarse).

²⁴Dennett (2006) discute el valor de investigaciones filosóficas análogas a las que haría alguien que investigara un juego idéntico al ajedrez salvo porque el rey se puede mover dos lugares en cada dirección, en vez de sólo uno; su veredicto es que no vale la pena realizarlas. En Dennett (2011) retoma el tema para describir cierto tipo de investigaciones filosóficas como una suerte de “antropología axiomática sofisticada”, en que se “tratan las cosas que dice la gente común como axiomas o teoremas, y se trata de volver consistente el conjunto de los datos”, sin tener la ilusión de que la teoría que se axiomatiza es verdadera. Allí propone que sería positivo que estas investigaciones se realizaran como propedéutica para teorías filosóficas sustantivas, que puedan acomodar correctamente, empleando la frase de Sellars (1963), “las imágenes científica y manifiesta”:

... the perspective I would recommend is that of the diplomatic anthropologist, not the metaphysician intent on limning the ultimate structure of reality. The ontology of everyday life is now teeming with items that, like fatigues, sit rather awkwardly in the world of atoms and molecules. If we can understand how this population explosion came about, and why it is so valuable to us as agents in the world, we can perhaps discharge our philosophical obligations without ever answering the ultimate ontological question. To me it looks more and more like professional make-work, an artifact of our reasonable but ultimately optional desire for systematicity rather than a deeper mystery in need of solving.

La actividad filosófica en general es tanto “analítica” como “sintética”. El entendimiento de las disposiciones a creer y las creencias que se tienen requieren de, por así decirlo, el desmenuzamiento de las estructuras que les dan sentido. Mediante este análisis se obtiene entendimiento sobre cómo pueden vincularse conceptos de forma general, y se adquiere la capacidad de construir sistemas de creencias y conceptos que pueden dar sentido a lo que resulta intelectualmente problemático (lo que motiva las preguntas filosóficas). El filósofo trata de hacer fértiles sus ideas; pero sólo pueden hacerse fértiles si se las puede entender. La actividad filosófica implica también, por lo tanto, un intento por disolver los malentendidos, el error y la obscuridad. Con esta observación se entra al campo de lo prescriptivo. Si la descripción que se ha propuesto (esbozado, en realidad) de la actividad filosófica es correcta, estos son requisitos inevitables de cualquier filósofo que pretenda alcanzar su objetivo de *entender*. Las maneras de llegar a esta meta pueden ser varias, pero las restricciones son las mismas para todos.

El objetivo de esta tesis era presentar el problema epistemológico de los juicios de posibilidad y su alcance en el campo de la actividad filosófica, y también sugerir una manera de ver la naturaleza de esta actividad; es lo que acaba de completarse. Además, se ha desarrollado una perspectiva sobre cómo abordar la cuestión de la justificación de los juicios de posibilidad, apelando a las nociones de entendimiento y de unificación. Esta perspectiva proporciona, como se ha visto, razones para adoptar esa manera de ver la filosofía; con esto, la motivación de este trabajo queda hasta aquí satisfecha. Los juicios de posibilidad siguen rodeados de cuestiones abiertas, que superan por mucho lo que se podría llevar a cabo en una sola investigación. La actividad filosófica es, por fuerza, un intento de aproximación.

Bibliografía

- Adorno, Theodor. 2001. *Kant's Critique of Pure Reason (1959)*. Ed. Rolf Tiedelmann. Stanford: Stanford University Press.
- Alvarez, Claudia. 2007. "Does Philosophy Improve Critical Thinking Skills?." (Tesis) Department of Philosophy, Faculty of Arts, University of Melbourne.
- Baumberger, Christoph. 2011. "Understanding and its relation to knowledge." *Epistemology: Contexts, Values, Disagreement. Papers of the 34th International Wittgenstein Symposium*, 16–18. Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society.
- Bealer, George. 1987. "Philosophical limits of scientific essentialism." *Philosophical Perspectives* I: 280–365.
- . 1996. "A priori knowledge and the scope of philosophy." *Philosophical Studies* 81: 121–142.
- . 2000. "A theory of the a priori." *Pacific Philosophical Quarterly* 81: 1–30.
- . 2002. "Modal epistemology and the rationalist renaissance." *Conceivability and Possibility*, ed. T. Gendler & J. Hawthorne, 71–125. Oxford: Oxford University Press.
- Chalmers, David. 2002. "Does conceivability entail possibility?." *Conceivability and Possibility*, ed. T. Gendler & J. Hawthorne, 145–200. Oxford: Oxford University Press.
- Chisholm, Roderick. 1973. *The Problem of the Criterion: The Aquinas Lecture*. Milwaukee: Marquette University Press.
- van Cleeve, James. 1998. "Foundationalism, epistemic principles and the cartesian circle." *Descartes*, ed. John Cottingham, 101–131. Oxford: Oxford University Press.
- Copeland, B. J. 2002. "The genesis of possible world semantics." *Journal of Philosophical Logic* 31: 99–137.
- Darwin, Charles. 1876. *The Origin of Species*. Londres: John Murray.
- Dennett, Daniel. 2006. "Higher-order truths about chess." *Topoi*: 39–41.
- . 2011. "Kinds of things: towards a bestiary of the manifest image." *Scientific Philosophy and Metaphysics*. Oxford: Oxford University Press.

- Descartes, René. 1991. *The Philosophical Writings of Descartes*. Vol. III, The Correspondence. Cambridge: Cambridge University Press.
- . 1995. *Los principios de la filosofía*. Madrid: Alianza.
- . 2007. *Discurso del Método/Meditaciones Metafísicas*. Austral.
- Devitt, Michael. 2011. “Experimental semantics.” *Philosophy and Phenomenological Research* 82: 418–435.
- Divers, John. 2002. *Possible Worlds*. Londres: Routledge.
- Elgin, Catherine. 2007. “Understanding and the facts?.” *Philosophical Studies* 132: 33–42.
- Frege, Gottlob. 1974. “Sobre sentido y significado.” *Escritos Logico-Semanticos*. Madrid: Tecnos.
- Friedman, Michael. 1974. “Explanation and scientific understanding.” *The Journal of Philosophy* 71: 1–19.
- Garber, D. 1996. “Descartes physics.” *The Cambridge Companion to Descartes*, ed. John Cottingham. Cambridge: Cambridge University Press.
- Glymour, Clark. 1980. “Explanations, tests, unity and necessity.” *Nous* 14: 31–50.
- Grimm, Stephen. 2010. “Understanding.” *The Routledge Companion of Epistemology*. Londres: Routledge.
- Hempel, Carl. 1965. *Aspects of Scientific Explanation*. Nueva York: Free Press.
- Hempel, Carl & Oppenheim, Paul. 1948. “Studies in the Logic of Explanation.” *Philosophy of Science* 15: 135–175.
- Hendricks, Vincent F. 2005. *Mainstream and Formal Epistemology*. Cambridge: Cambridge University Press.
- Hill, James M. & Cox, Barry J. 2012. “Einstein’s special relativity beyond the speed of light.” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*. doi:10.1098/rspa.2012.0340.
- Ichikawa, Jonathan & Maitra, Ishani & Watherson, Brian. 2011. “In defense of a kripkean dogma.” *Philosophy and Phenomenological Research* 85: 56–68.
- van Inwagen, Peter. 2001. “Modal epistemology.” *Ontology, Identity and Modality*, 243–258. Cambridge: Cambridge University Press.
- Jackson, Frank. 1982. “Epiphenomenal qualia.” *The Philosophical Quarterly* 32: 127–136.
- Khalifa, Kareem. 2011. “Understanding, knowledge and scientific antirealism.” *Grazer Philosophische Studien* 83: 93–112.
- Kitcher, Philip. 1976. “Explanation, conjunction and unification.” *The Journal of Philosophy* 73: 207–212.
- . 1981. “Explanatory unification.” *Philosophy of Science* 48: 507–531.

- . 1989. “Explanatory unification and the causal structure of the world.” *Minnesota Studies in the Philosophy of Science XIII: Scientific Explanation*, ed. Philip Salmon, Wesley Kitcher, 410–505. Minneapolis: University of Minnesota Press.
- Kratzer, Angelika. 2002. “The notional category of modality.” *Formal Semantics: The Essential Readings*. Oxford: Blackwell.
- Kripke, Saul. 1963. “Semantical considerations on modal logic.” *Acta Philosophica Fennica* 16: 83–94.
- . 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Nueva York: Cambridge University Press.
- Lewis, David. 1990. “What experience teaches us.” *Mind and Cognition: A Reader*, ed. W. Lycan. Oxford: Blackwell.
- Lipton, Peter. 2001. *Inference to the Best Explanation*. Londres: Routledge.
- . 2004. “What good is an explanation?.” *Understanding Explanation*, 1–22. Oxford: Oxford University Press.
- MacFarlane, John. 2003. “Future Contingents and Relative Truth.” *The Philosophical Quarterly* 53: 321–336.
- Machery, Edouard & Mallon, Ron & Nichols, Shaun & Stich, Stephen. 2004. “Semantics, cross-cultural style.” *Cognition* 92.
- . próximo. “If folk intuitions vary, then what?.” *Philosophy and Phenomenological Research*.
- Menzies, Peter. 1998. “Possibility and conceivability: a response-dependent account of their connections.” *European Review of Philosophy*, 3:255–277. Stanford: CSLI Publications.
- Nozick, Robert. 2001. *Invariances*. Cambridge: Belknap Press of Harvard University Press.
- Peacocke, Christopher. 1998. *Being Known*. Oxford: Oxford University Press.
- Pritchard, Duncan. 2009. “Knowledge, Understanding and Epistemic Value.” *Epistemology*. Cambridge: Cambridge University Press.
- Putting, Gary. 2009. *What Philosophers Know: Case Studies in Recent Analytic Philosophy*. Cambridge: Cambridge University Press.
- Russell, Bertrand. 1905. “On denoting.” *Mind New Series* 14: 479–493.
- Ryle, Gilbert. 2009. *The Concept of Mind*. Londres: Routledge.
- Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- . 1989. “Four decades of scientific explanation.” *Minnesota Studies in the Philosophy of Science XIII: Scientific Explanation*, ed. Philip Salmon, Wesley Kitcher, 3–219. Minneapolis: University of Minnesota Press.

- Schopenhauer, Arthur. 1998. *On the Basis of Morality*. Indianapolis: Hackett Publishing Company.
- Sellars, Wilfrid. 1963. "Philosophy and the scientific image of man." *Science, Perception and Reality*. Atascadero, CA: Ridgeview Publishing Company.
- Stroud, Barry. 1984. *The Significance of Philosophical Scepticism*. Oxford: Oxford University Press.
- Tarski, Alfred. 1956. *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press.
- Thurston, William. 1994. "On proof and progress in mathematics." *Bulletin of the American Mathematical Society* 30: 161–177.
- Unger, Peter. 2012. "Beyond Emptiness." (Manuscript) <http://philosophy.fas.nyu.edu/object/peterunger>.
- Williamson, Timothy. 2007. "Philosophical knowledge and knowledge of counterfactuals." *Grazer Philosophische Studien* 74: 89–123.
- Wisdom, John. 1936. "Philosophical perplexity." *Proceedings of the Aristotelian Society* 37: 71–88.
- Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Oxford: Basil Blackwell.
- . 1974. *Tractatus Logico-Philosophicus*. Londres: Routledge.
- Woodward, James. 1989. "The causal mechanical model of explanation." *Minnesota Studies in the Philosophy of Science XIII: Scientific Explanation*, ed. Philip Salmon, Wesley Kitcher, 357–383. Minneapolis: University of Minnesota Press.
- Yablo, Stephen. 1993. "Is conceivability a guide to possibility?." *Philosophy and Phenomenological Research* 53: 1–42.

11