

UNIVERSIDAD AUTÓNOMA DE ZACATECAS
“FRANCISCO GARCÍA SALINAS”

UNIDAD ACADÉMICA DE FILOSOFÍA

MAESTRÍA EN FILOSOFÍA



El conocimiento de la propia mente
Donald Davidson sobre autoridad de la primera persona,
externalismo y racionalidad

T E S I S

QUE PRESENTA

MARC JIMÉNEZ ROLLAND

PARA OPTAR POR EL TÍTULO DE

MAESTRO EN FILOSOFÍA

DIRECTOR: DR. JORGE ROLANDO TAGLE MARROQUÍN

Speak out, sir! I will know your thoughts.

Walter Scott, *Wandering Willie's Tale*

Agradecimientos

Mientras trabajaba en el contenido de este escrito, fui beneficiado con la extraordinaria tutela, la incondicional amistad, la estimulante discusión, la grata compañía y las diversas atenciones que incontables personas han tenido la gentileza de brindarme. No dispongo del espacio suficiente para agradecer a todas apropiadamente. En todo caso, he procurado expresarles mi gratitud, y retribuir lo mucho que me han dado, en el momento propicio. Lamentablemente, no siempre ha sido así; sé que unas cuantas palabras no enmendarán mis ofensas. Por ello, me limitaré a consignar aquí mis deudas más evidentes.

El afecto, apoyo y aliento que me ha proporcionado cada uno de los miembros de mi familia es invaluable. Ramón, Monique, Rita, Daniel y Edith han sido los más cercanos, y a quienes más debo, pero encabezan lo que debería de ser un enorme listado.

Mis maestros y colegas en la Universidad Autónoma de Aguascalientes y en la Universidad Autónoma de Zacatecas me manifestaron mayor condescendencia de la que he merecido o esperado. Agradezco, en especial, a José de Lira, Graciela Castorena, Patricia Patiño, Rosa Imelda López y Antonio Núñez por concederme el privilegio y la responsabilidad de enseñar filosofía. A quienes han sido mis alumnos en estas instituciones, les expreso gratitud por su paciencia y por mantener con vida esta búsqueda perenne. En especial, agradezco a Alejandro Mosqueda, quien tuvo la amabilidad de leer y comentar un manuscrito de este texto.

Agradezco a los miembros de la comisión revisora de esta tesis (Jorge Tagle, Mario Gensollen, Hugo Ibarra, Antonio Núñez y José de Lira), quienes tuvieron la disposición de leer y comentar este texto con una agilidad y agudeza sorprendentes. María de la Luz Jacobo y Rosa Rodríguez, con paciencia y buen humor, evitaron que desfalleciera en los trámites administrativos; asimismo, expreso mi gratitud a Ricardo Martínez, director de la Unidad Académica de Filosofía de la UAZ, por su apoyo en el proceso de entrega de esta tesis.

Me he beneficiado de conversaciones y proyectos conjuntos con Armando Cántora y Sílvia Pinto, de la Universidad Autónoma Metropolitana-Iztapalapa. Agradezco, asimismo, a Jonathan Vogel, del Amherst College, que me haya proporcionado parte del material que cito en el texto y las agudas observaciones sobre la filosofía de Davidson que tuvo a bien comunicarme.

He tenido la dicha de encontrar amigos indulgentes y admirables, entre los que me da gusto contar a María Elba Macías, Salvador Gallardo, Raquel Mercado, César Lozano, Amalia Huerta, Claudia Delgado, Alma Palacios, Alberto Medina, Martha Esparza, Caleb Olvera y Osvaldo Rosas. De diversas maneras — que algunos de ellos desconocen — han contribuido a lo que aquí escribo. Sobre los temas de esta tesis, y sobre muchos otros, he charlado incansablemente con mi entrañable amigo Mario Gensollen; aunque a menudo en pugna, sé que nuestros puntos de vista emanan de acuerdos cruciales.

Por último, puesto que se trata de la más evidente, mi deuda de gratitud con Jorge Tagle es incalculable. Tuve la fortuna de que fuese mi maestro durante una etapa fundamental de mi vida, tanto académica como personal. El trabajo que realiza en filosofía me resulta inspirador: su profunda comprensión de los problemas filosóficos y su capacidad para transmitir el interés por enfrentarlos son extraordinarias. Ha marcado profundamente mi desarrollo intelectual, aunque el contenido de estas páginas no lo muestre con justicia. Revisó con paciencia varias versiones de esta tesis, señalando omisiones y errores. Los que, a su pesar, persisten son enteramente mi responsabilidad.

A todos ellos, y a quienes no he nombrado, mi gratitud vitalicia. Espero que el resultado no los decepcione.

Índice de contenidos

Introducción	1
Capítulo 1	
El conocimiento de la propia mente como problema	7
<i>1.1. El conocimiento de la propia mente y otras formas de conocimiento</i>	8
1.1.1. El análisis del conocimiento	10
1.1.2. La singularidad del conocimiento de la propia mente	16
<i>1.2. El alcance y los rasgos característicos del conocimiento de la propia mente</i>	20
1.2.1. Alcances y límites del conocimiento de la propia mente	20
1.2.2. Autoridad de la primera persona y asimetría epistémica	25
<i>1.3. Las fuentes de la autoridad de la primera persona I: explicaciones de acceso epistémico</i>	31
1.3.1. La concepción cartesiana	32
1.3.2. La teoría de la familiaridad (<i>acquaintance</i>)	34
1.3.3. La teoría del sentido interno	36
<i>1.4. Las fuentes de la autoridad de la primera persona II: explicaciones no epistémicas</i>	39
1.4.1. El conductismo lógico	41
1.4.2. El neoexpresivismo	44
<i>1.5. Donald Davidson: la autoridad de la primera persona como condición de la interpretación</i>	47
1.5.1. La autoridad de la primera persona en el lenguaje	49
1.5.2. La diferencia epistémica: interpretación radical	53
1.5.3. La explicación davidsoniana y los retos de una teoría explicativa	56
<i>1.6. Problemas y perspectivas en la concepción davidsoniana</i>	58
Capítulo 2	
Externalismo, autoengaño y autoridad de la primera persona	63
<i>2.1. Externalismo: motivaciones y alcance</i>	64
2.1.1. Externalismo e internalismo semánticos	65
2.1.2. Externalismo semántico y anti-individualismo	66
2.1.3. Argumentos a favor del externalismo (ortodoxo)	68
<i>2.2. El debate sobre la incompatibilidad entre externalismo y autoridad de la primera persona</i>	72
2.2.1. El problema del logro	73
2.2.2. El problema de la consecuencia	75

2.3. <i>Externalismo y autoridad: la concepción davidsoniana</i>	79
2.3.1. El externalismo davidsoniano	81
2.3.2. Respuestas a los argumentos incompatibilistas	83
2.4. <i>Autoengaño e irracionalidad</i>	90
2.4.1. Tres ejemplos de autoengaño	90
2.4.2. Modelos intencionales y motivacionales del autoengaño	93
2.5. <i>Autoengaño y autoridad de la primera persona</i>	96
2.5.1. Autoengaño y racionalidad	97
2.5.2. La mente dividida	100
2.6. <i>El conocimiento de la propia mente, la racionalidad y el proyecto epistemológico</i>	101
Capítulo 3	
El lugar del conocimiento de la propia mente en la epistemología	105
3.1. <i>La epistemología como proyecto filosófico</i>	106
3.1.1. Las motivaciones de una teoría del conocimiento	107
3.1.2. Cinco problemas de la epistemología	109
3.1.3. Objetivos y estrategias	111
3.2. <i>El lugar del conocimiento de la propia mente en la epistemología cartesiana</i>	114
3.2.1. Objetivos y supuestos de la epistemología cartesiana	115
3.2.2. Justificación: fuerza y estructura	117
3.2.3. El fundacionismo y el conocimiento de la propia mente	120
3.3. <i>Tres retos de la epistemología cartesiana</i>	124
3.3.1. El problema del escepticismo	124
3.3.2. Tres formas de escepticismo	126
3.3.3. El fracaso de la epistemología cartesiana	131
3.4. <i>La naturalización de la epistemología</i>	134
3.4.1. El naturalismo filosófico	134
3.4.2. La epistemología naturalizada y el objetivo descriptivo	136
3.4.3. Falibilismo, holismo y externalismo	140
3.5. <i>El conocimiento de la propia mente en la epistemología naturalizada</i>	146
3.5.1. Tres variedades de conocimiento	147
3.5.2. La indeterminación de lo mental	151
3.5.3. La indispensabilidad del conocimiento de la propia mente	154
Conclusiones	157
Referencias	161

Introducción

The trouble is, as so often in philosophy, it is hard to improve intelligibility while retaining the excitement.

(Davidson, 1974b: 183)

Sabemos mucho acerca de nuestra propia mente. Sabemos sin importar cuán confusas sean nuestras ideas al respecto qué es lo que queremos, cuáles son nuestras creencias, esperanzas, anhelos, temores y somos, en general, capaces de atribuirnos a nosotros mismos correctamente una gran variedad de estados psicológicos. En consecuencia, cada uno de nosotros es considerado, salvo contadas excepciones, como un informante confiable del contenido de su propia mente. Este hecho, aparentemente trivial, ha sido motivo de importantes controversias en filosofía. Parecería ocioso ponerlo en duda; no obstante, algunos influyentes pensadores lo han hecho. En contraste, una tradición nada despreciable ha sugerido que debemos considerarlo como el punto de partida, el único posible, de nuestra imagen del mundo.

En esta tesis me propongo investigar este hecho. Para decirlo de manera sucinta, me propongo investigar cómo hemos de entender el conocimiento que cada uno tiene de sus propios pensamientos y cuál es el papel que desempeña en nuestra concepción del mundo y del ser humano. Ofreceré razones para disipar las dudas que gravitan en torno suyo; pero también intentaré mostrar que algunas otras nos indican que su importancia filosófica se sitúa, en realidad, en una locación distinta a la que se le ha asignado en la cartografía filosófica.

Permítaseme hacer una breve digresión. A la luz de lo indicado en el párrafo anterior, el lector podría estar intrigado por cuáles son las razones que mueven a alguien a discutir o, al menos, a hacerlo con la extensión con la que aquí lo hago en torno a dudas infundadas sobre un asunto que no tiene la importancia capital que en ocasiones se le ha otorgado. Es una creencia difundida la de que tal es la tarea en la que invierten su tiempo quienes se dedican al quehacer filosófico. Así descrita, la filosofía no es un ejercicio al cual uno dedicaría gran parte de su tiempo, mucho menos uno en el que se buscaría hacer carrera u obtener un grado. No creo, por supuesto, que la filosofía consista en esto. Por ello, y a manera de justificación, he de abundar un poco más sobre lo antes dicho.

Aunque considero que las dudas en torno al conocimiento de la propia mente están desorientadas, no me parece que sean infundadas. En general, no pienso que las dudas que se presentan en filosofía carezcan de sustento. Acudiendo a una bien conocida afirmación wittgensteiniana: “Uno duda por razones específicas” (1979: § 458); y sólo en el trasfondo de cuantiosas suposiciones cobra sentido la duda. Y lo tiene, ciertamente, la duda filosófica. Por otra parte, el terreno de las ideas es sinuoso y el conjunto de aquello que consideramos como no problemático se modifica constantemente. Una imagen imprecisa, sólo ocasionalmente presentada de manera explícita, sobre la relación que guardamos con nuestra ‘vida interna’ ocupó durante mucho tiempo un lugar medular en la reflexión filosófica; sólo recientemente se ha visto la necesidad de examinar críticamente dicha imagen. Aunque sería exagerado asumir que el tema que examino en las siguientes páginas se sitúa en el centro de las controversias actuales, creo que es un terreno propicio para poner a prueba algunas de las ideas y herramientas filosóficas que han visto la luz en la época contemporánea.

En publicaciones recientes (*e.g.*, Gertler, 2011 y Goldberg, 2011) se ha insistido en la necesidad de desarrollar una epistemología del conocimiento de la propia mente. Coincido con este diagnóstico. No obstante, me parece que los esfuerzos

contemporáneos por dar continuidad a este proyecto son poco satisfactorios. En su mayoría, asumen como neutral una noción de “acceso privilegiado” que considero problemática y es, quizá, uno de los vestigios de la tradición cartesiana. Aunque esa noción encapsula algunos de los rasgos que hacen especial a esta forma de conocimiento, no creo que haya que elevarla al estatus de *datum*.

Por el contrario, algunos autores contemporáneos han insistido en caracterizar las peculiaridades de nuestras autoadscripciones de pensamientos enfatizando su importancia para el razonamiento práctico y la evaluación moral de la acción. Aunque estos elementos son importantes, y sin duda deben formar parte de una explicación satisfactoria de la indispensabilidad del conocimiento de la propia mente, en ocasiones enturbian sus rasgos epistemológicos. Ésta es, en parte, la razón por la que tales aspectos brillan por su ausencia en esta tesis. Se trata, sin duda, de una de sus varias deficiencias; pero me permite acentuar las cuestiones relativas al conocimiento que a menudo se pasan por alto.

En este texto defiendo la concepción davidsoniana del conocimiento de la propia mente, que ocupa una posición intermedia entre estas alternativas extremas. No intento defender que se encuentre exenta de críticas o que carezca de deficiencias, pero creo que ha sido injustamente ignorada como alternativa viable frente a enfoques más populares. Este desdén es producto de la creencia, ampliamente difundida, de que la explicación de Davidson no da cabida a los aspectos más relevantes del conocimiento de la propia mente. Mi propósito es mostrar que esta creencia es errónea y que la concepción davidsoniana tiene mucho que decir en su favor. Los tres capítulos de este escrito se orientan, desde diversas perspectivas, a la consecución de este objetivo.

En el capítulo 1 presento algunas motivaciones para considerar que hay algo inquietante y problemático en la idea de que tenemos conocimiento de nuestra propia mente. En el proceso, intento identificar algunos de los rasgos distintivos de esta forma de conocimiento. Ésas son las tareas habituales del refinamiento o

clarificación filosófica de una intuición. En seguida, paso a examinar algunas de las explicaciones sustantivas de estos rasgos que predominan en la discusión contemporánea. Presento las críticas a las que están sujetas estas explicaciones, mostrando que son inadecuadas para dar cuenta de los rasgos distintivos del conocimiento de la propia mente. A continuación, expongo el examen davidsoniano de la intuición de que el conocimiento de la propia mente es especial. La conclusión, a favor de la cual se argumenta a lo largo de los tres capítulos, es que esta intuición puede ser filosóficamente respaldada si se reconoce, en general, el estrecho vínculo entre el lenguaje, lo mental y la comunicación, que se hace explícito al examinar la naturaleza de la interpretación.

En el capítulo 2 discuto dos retos que enfrenta la explicación davidsoniana. El primero es lo que se ha pensado que constituye una de las objeciones (o cuando menos una de las restricciones) filosóficas más importantes a la intuición de que tenemos un conocimiento especial del contenido de nuestros pensamientos: el externalismo sobre la intencionalidad. Defiendo la posición davidsoniana apelando a su caracterización del externalismo, que confronto con las versiones ortodoxas. El segundo reto consiste en ofrecer una explicación de una forma de irracionalidad que pone en entredicho uno de los rasgos distintivos del conocimiento de la propia mente: el autoengaño. Siguiendo nuevamente a Davidson, sugiero que este fenómeno puede hacerse compatible con la intuición prefilosófica de que sabemos, de manera especial, cuáles son nuestros pensamientos. Argumento, además, que el externalismo y el autoengaño nos permiten robustecer la afirmación de que nuestras autoadscripciones de estados mentales manifiestan un logro cognitivo auténtico.

En el capítulo 3 confronto una influyente concepción de la epistemología, de inspiración cartesiana, y del papel que en ésta desempeña el conocimiento de la propia mente con la explicación davidsoniana. Argumento que la epistemología cartesiana es sólo una de las diversas maneras de examinar filosóficamente el co-

nocimiento; asimismo, sondeo los problemas que enfrenta y sugiero, como lo ha hecho buena parte de la reflexión contemporánea, que es preciso abandonarla. En su lugar, se propone investigar los problemas de la epistemología desde otra perspectiva teórica y metodológica: el naturalismo filosófico. Desde esta perspectiva es posible reconsiderar el papel del conocimiento de la propia mente en nuestra concepción del ser humano y del mundo en el que habita.

Una advertencia final sobre el contenido de las citas. Aunque hay versiones castellanas de algunas de las obras que por su claridad, precisión o elegancia se citan en el texto, para preservar el contenido original de los pasajes en numerosas ocasiones he preferido emplear traducciones propias. A menudo consulté con provecho las traducciones publicadas en castellano, pero en la mayoría de los casos me parecieron deficientes. De modo que, a menos que se indique lo contrario, las traducciones de las citas son propias.

Capítulo 1

El conocimiento de la propia mente como problema

We know in a way no one else can what we believe, fear, want, value, and intend. We know how things seem to us, how they look to us, feel to us, smell and sound to us to be. We know these things in a way we can never know about the world around us. Whether or not we are sometimes wrong about the contents of our own minds, whether or not we can be in doubt about our own sensations and thoughts, one thing is certainly true of such beliefs: they cannot be generally mistaken. If we think we have a certain thought or sensation, there is a strong presumption that we are right.

(Davidson, 1990: 193)

El interés filosófico por el conocimiento de la propia mente no es reciente. Una larga y venerable tradición —encauzada por intereses teóricos diversos— ha suscitado, de manera intermitente, profundas cuestiones en torno suyo. Por sus peculiaridades, a veces se le ha situado al centro de la cartografía filosófica; aunque en otras ocasiones su excentricidad ha sido considerada como un motivo para exorcizarlo de nuestra imagen del mundo. Al igual que muchos otros importantes problemas filosóficos, el que ocupa a este escrito es producto de una tensión, o de un cúmulo de tensiones, cuyos indicios permean de manera ubicua nuestra concepción cotidiana del ser humano y del mundo en el que habita.

Encontramos problemas al intentar conciliar una imagen del mundo en la que nos representamos como agentes libres con otra, también muy arraigada, en la que los fenómenos son gobernados por regularidades y es posible explicar la

aparición de cualquier suceso — la acción humana incluida — apelando a los patrones repetitivos de la legislación natural. Asimismo, varios aspectos del conocimiento de la propia mente nos resultan, en el mejor de los casos, sumamente inquietantes.

En este capítulo se presentan algunos de los rasgos que suscitan interés filosófico en torno al conocimiento de sí o de la propia mente (*self-knowledge*). Esto permitirá reconocer qué aspectos de los hechos que se agrupan bajo estas designaciones deben ser tomados en cuenta como requisito mínimo para ofrecer una explicación satisfactoria de este fenómeno. En la sección 1.1 se motiva, a partir de una elucidación del conocimiento, la idea de que tiene sentido hablar de las auto-atribuciones de estados mentales como uno de nuestros logros epistémicos. En la sección 1.2 se ofrece una delimitación del alcance del conocimiento de la propia mente y se caracterizan algunos de sus rasgos más importantes. En las secciones 1.3 y 1.4 se examinan con algún detalle varias explicaciones sustantivas de estos rasgos, indicando las dificultades que cada una enfrenta. La sección 1.5 está dedicada a la exposición de la propuesta davidsoniana, que difiere en aspectos cruciales de las explicaciones precedentes. Argumento que, no obstante, proporciona lo que se requiere de una teoría explicativa. En la última sección de este capítulo se exponen algunos de los problemas que enfrenta la concepción davidsoniana, los cuales serán objeto de discusión del capítulo 2.

1.1. El conocimiento de la propia mente y otras formas de conocimiento

Hay un aspecto de nuestra vida mental que todos damos por supuesto sin mayor cuestionamiento. Es algo tan familiar que normalmente no suscita ninguna clase de duda.¹ Se manifiesta en gran parte de nuestra conducta, especialmente en nues-

¹ Se ha sugerido incluso que el escepticismo concerniente al conocimiento de sí podría resultar incoherente (*e.g.*, Ebbs, 2001 y 2005, pero véase también Brueckner, 1997).

tra conducta lingüística. Por lo general, cada uno de nosotros *sabe* cuáles son sus pensamientos; cada uno *sabe* lo que quiere decir con sus palabras. El que sepamos estas cosas parece ser especial en el siguiente sentido: aunque alguien más puede también llegar a saberlas, su conocimiento no tiene el mismo grado de seguridad que el nuestro.²

Tal como ha sido presentado en el párrafo anterior, el asunto que nos ocupa parece guardar una estrecha relación con el conocimiento. Y es difícil ver el asunto de otra manera.³ Después de todo, si queremos *saber* lo que alguien más piensa, o lo que quiso decir al proferir una oración, es natural que nos dirijamos a *esa persona* en calidad de informante confiable para dirimir la cuestión; no parece haber una mejor manera de hacerlo. No obstante, al confrontarla con otras formas de conocimiento, ésta resulta bastante peculiar.

Ciertamente, sabemos mucho. O, por lo menos, fuera de la discusión filosófica en torno al escepticismo, somos bastante prolíficos en nuestras atribuciones de conocimiento. Sabemos cosas acerca del mundo fuera de nuestra mente; de los acontecimientos que acaecen en nuestro entorno inmediato, aunque también de los que se producen en regiones distantes del espacio-tiempo. En ocasiones, incluso sabemos lo que piensan otras personas, por qué actúan de cierta manera y cuáles son sus intenciones. Asimismo, si lo sugerido líneas arriba es correcto, sabemos mucho acerca de nuestra propia mente; *e.g.*, sabemos cuáles son nuestras creencias, deseos e intenciones en el momento actual.

² Quizá una excepción notable a este hecho se dé en la política, donde parece que los portavoces de un dignatario se encuentran en mejor posición que éste para saber cuáles son sus opiniones sobre algún asunto público, e incluso llegan a corregir sus declaraciones sobre lo que piensa o “quiso decir” al respecto. (Encontramos un caso emblemático de este tipo de situaciones en la política mexicana. Rubén Aguilar, portavoz del ex presidente Vicente Fox, se dedicaba a este tipo de prácticas cada mañana en los Pinos). Quizá haya que desestimar la pertinencia de estos casos como contraejemplos. Parecen mostrar, no que los dignatarios ignoran lo que piensan o lo que quieren decir, sino que no saben que es inoportuno o poco razonable pensar o decir ese tipo de cosas.

³ Pero véase la discusión en el apartado 1.4.2.

1.1.1. El análisis del conocimiento

Pero, si lo pensamos con detenimiento, notaremos que son cosas muy dispares las que agrupamos entre nuestros logros cognitivos. Su diversidad se manifiesta en varios de sus rasgos. Conciernen a diferentes tipos de cuestiones: el mundo y la mente. Algunos son independientes de lo que creemos mientras que en un sentido particular otros no.⁴ Además, están sujetos a diferentes formas de corrección y son respaldados de maneras distintas. ¿Por qué los agrupamos, entonces, dentro de una misma categoría?, ¿en qué sentido son todos éstos formas de conocimiento?, ¿en qué consiste decir que *sabemos* acerca de estas cosas?

El punto aquí es uno que ha ameritado notable atención filosófica: precisar la naturaleza de lo que llamamos (o bien especificar lo que entendemos por) conocimiento o saber. Dada la inmensa literatura en torno a esta cuestión, lo que sigue es inevitablemente y en el mejor de los casos esquemático.⁵ Lo que se requiere aquí puede tomar la forma de un análisis del conocimiento. Tradicionalmente, esta tarea se ha caracterizado como un intento por precisar condiciones que sean individualmente necesarias y conjuntamente suficientes para la aplicación del concepto de ‘conocimiento’ o ‘saber’.⁶ Las condiciones de adecuación de dicho análisis suelen especificarse mediante el siguiente esquema bicondicional:

(AC) S sabe que p si y sólo si _____.

El concepto a analizar (o *analysandum*), que aparece en el lado izquierdo del bicondicional, esquematiza las atribuciones de conocimiento (donde S está en lugar

⁴ Tenemos “...la idea intuitiva de que la verdad, *aparte de unos cuantos casos especiales*, es completamente independiente de nuestras creencias; como a veces se formula, nuestras creencias podrían ser tal como son y aún así la realidad y por ende la verdad acerca de la realidad ser muy diferente.” (Davidson, 1990: 40. Cursivas mías.)

⁵ Sigo de cerca, en esta exposición, la presentación de Steup (2006).

⁶ Para los propósitos de esta discusión, y siguiendo a la tradición dominante en la epistemología contemporánea, nos concentraremos en lo que se denomina saber proposicional; es decir, lo que lingüísticamente se representa mediante una conjugación del verbo ‘saber’ (a la que puede anteponerse una expresión que designe a un individuo, o a varios) seguida de una cláusula subordinada que comienza con la preposición ‘que’ y se complementa con una oración declarativa.

de un individuo racional o sujeto epistémico, y p representa una oración declarativa). El espacio en blanco debe ser completado con el *analysans*: una lista de condiciones individualmente necesarias y conjuntamente suficientes que sean, al menos, extensionalmente correctas (que preserven la verdad del enunciado bicondicional). Para evaluar la corrección del análisis, es conveniente separar el bicondicional del esquema (AC) en dos condicionales:

(AC₁) Si S sabe que p , entonces _____.

Si al completar el espacio en blanco de (AC₁) con un tipo de condiciones el esquema resultante es verdadero para todas sus instancias de sustitución (*i.e.*, para cualquier sujeto que ocupe el lugar de S y cualquier oración en lugar de p), entonces se dice que las condiciones que ocupan dicho espacio son *necesarias* para el conocimiento; análogamente, si el condicional resulta falso para alguna instancia de sustitución (si algún S sabe que p sin que las condiciones del consecuente sean satisfechas), entonces no se trata de condiciones necesarias.

Consideremos un ejemplo. Supóngase que completamos el esquema (AC₁) de la siguiente manera: si S sabe que p , entonces es p es una verdad necesaria; *i.e.*, supóngase que pensamos que para saber que p se requiere que sea imposible que p sea falsa. Para evaluar si ésta es una condición necesaria del saber, veamos una instancia de sustitución. En lugar de S , emplearemos el pronombre ‘yo’ (el lector puede utilizar este pronombre para referirse a sí mismo) y en lugar de p usaremos la oración “Estoy aquí ahora”. El antecedente del condicional sería “Yo sé que estoy aquí ahora”. Presumiblemente, ésta es una oración verdadera para cualquiera que la profiera, incluso si ignora sus coordenadas geográficas o ha perdido la noción del tiempo. No obstante, cualquiera puede apreciar que el consecuente es falso: quien profiere la oración “Estoy aquí ahora” podría estar en otro lugar en ese momento, incluso podría no existir en absoluto. En consecuencia, que p sea una verdad necesaria no es una condición necesaria para que S sepa que p .

El segundo condicional que se desprende de (AC) tiene como propósito identificar condiciones suficientes para el conocimiento o saber:

(AC₂) Si _____, entonces S sabe que p .

Si al llenar el espacio en blanco de (AC₂) con un tipo de condiciones el esquema resulta verdadero para todas sus instancias de sustitución, entonces se dice que las condiciones enunciadas en el espacio en blanco son *suficientes* para el conocimiento; si las condiciones son satisfechas pero S no sabe que p , entonces las condiciones no son suficientes para el conocimiento.

Retomemos la condición examinada en ejemplo anterior. Quizá, que p sea una verdad necesaria es una condición suficiente, aunque innecesaria, para que S sepa que p . Si así fuera, cualquier instancia de sustitución de “Si p es una verdad necesaria, entonces S sabe que p ” debería ser verdadera. Sea ‘Juvencio’ una instancia de S y “ $3 - \sqrt{1 - 2\sqrt{1 + 3\sqrt{1 + \dots}}}$ ” una instancia de p . Si los teoremas del álgebra pueden ser considerados como verdades necesarias, entonces que 3 sea el resultado del radical jerarquizado infinito $\sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + \dots}}}$ es una verdad necesaria; pero es razonable suponer que a menos que sea un prodigio o esté familiarizado con los hallazgos de Ramanujan Juvencio no sabe que $3 - \sqrt{1 - 2\sqrt{1 + 3\sqrt{1 - \dots}}}$. Por ende, que p sea una verdad necesaria no es una condición suficiente para que S sepa que p .

Idealmente, un análisis satisfactorio del concepto de ‘conocimiento’ o ‘saber’ debe reunir (mediante un enunciado conjuntivo) condiciones que satisfagan, individualmente, el esquema (AC₁), y que, en conjunto, satisfagan el esquema (AC₂).⁷

Un prominente análisis del conocimiento que a menudo se atribuye a Platón (*Teeteto*, 201d; *Menón*, 98)⁸ lo define en términos de *creencia verdadera justificada*.

⁷ Podría pensarse que encontrar condiciones suficientes es todo lo que requiere un buen análisis. Por ello, no está de más indicar la importante función que desempeñan las condiciones necesarias: evitar que ‘excrecencias’ conceptuales, condiciones irrelevantes, sean incluidas en el *analysans*.

⁸ Aunque es un análisis que el mismo Platón rechazó y del que se ha modificado la tercera condición con respecto a la formulación original (véase, *e.g.*, el comentario en Zagzebski, 2009: 7).

De modo que, de acuerdo con lo que se denomina análisis (o definición) ‘tripartita’ del conocimiento (ATC), el esquema (AC) puede ser completado como sigue:

(ATC) S sabe que p si y sólo si (i) S cree que p , (ii) p es verdadera y (iii) S está justificado para creer que p .

La condición (i) especifica un estado mental del sujeto epistémico; la (ii) refiere a una condición de logro con respecto al contenido de dicho estado mental y aquello que representa; la (iii), por su parte, tiene como propósito vincular las condiciones (i) y (ii) de manera que la verdad de la creencia no sea producto de un ‘mero accidente’. Sin duda, es la condición (iii) la que mayor discusión epistemológica ha suscitado. Esto se debe, entre otras cosas, a que es extremadamente complicado precisar lo que se requiere para evitar la clase apropiada de fortuna epistémica (de creencias que son verdaderas por casualidad), preservando los nexos conceptuales del conocimiento con la acción racional.

El esquema (ATC) ha sido objeto de numerosas críticas en la discusión epistemológica reciente. Famosamente, en su influyente artículo «Is Justified True Belief Knowledge?», Edmund Gettier presentó dos situaciones en las que intuitivamente se admitiría que un individuo satisface las condiciones establecidas por (ATC), en las que, no obstante, no se le atribuiría conocimiento; *i.e.*, mostró dos instancias de sustitución de las condiciones enunciadas en (ATC) en las que no se preserva la verdad de (AC₂). De modo que lo que Gettier (1963) razonablemente mostró es que el análisis suministrado por (ATC) no reúne condiciones suficientes para la aplicación del concepto de conocimiento. En atención a esta crítica, los defensores del (ATC) han optado por seguir dos estrategias: (a) modificar la condición (iii), haciéndola inmune a los contraejemplos-Gettier (*e.g.*, Plantinga, 1993); o (b) añadir una cuarta condición que complete el análisis (*e.g.*, Harman, 1973).

Por otra parte, una respuesta alternativa ha sugerido que la condición (iii) no es necesaria⁹ para el conocimiento, especialmente si (iii) se explica en términos que involucren información accesible al sujeto (o dependiente de sus estados mentales) y si lo que (iii) se propone evitar es que la verdad de la creencia sea producto de “fortuna epistémica”, *i.e.*, que la conexión entre creencia y verdad sea accidental. Quizá lo único que se requiere es que la verdad de la creencia haya sido causada apropiadamente por el estado de cosas en cuestión (*e.g.*, el análisis de la condición de “rastreo” de Nozick, 1981) o que los procesos — en la mayoría de los casos, procesos no conscientes — que dieron origen a la creencia sean fiables, *i.e.*, procesos que generalmente dan lugar a creencias verdaderas (*e.g.*, los análisis fiabilistas de Goldman, 1976; Papineau, 1993).

Un tercer diagnóstico es que (ATC) falla debido a que no es posible, en absoluto, ofrecer *un* análisis del conocimiento. Esto puede deberse a que ‘conocimiento’ y ‘saber’ tienen diversos significados (*e.g.*, a que se trata de conceptos profundamente incoherentes, como sostiene Schiffer, 1993 y 2004), o bien a que se trata de conceptos básicos, primitivos, a partir de los cuales podemos elucidar otros conceptos que se consideraban menos problemáticos, como creencia y justificación (ésta es la vía emprendida por Williamson, 2000 y 2011).

No es mi propósito, en este breve espacio, intentar resolver la intrincada discusión de cuál es el análisis correcto de ‘conocimiento’ o ‘saber’, o la de si hay siquiera un análisis tal. Sin embargo, para fines expositivos, es conveniente precisar qué se tiene en mente al hablar de conocimiento en la presente discusión.

Pese al problema apuntado por Gettier, en lo que sigue — como en buena parte de la discusión contemporánea (véase, *e.g.*, Sosa, 1980: 3) — se considerará que alguna reconstrucción de (ATC) ofrece, si no un análisis de conocimiento, al

⁹ Al menos es así como Brandom (2000: 97) sugiere que leamos esta discusión. Otra manera de ver el asunto es considerarla como una forma distinta de proseguir la estrategia esbozada en el inciso (a) del párrafo anterior, generalizando la condición (iii) para que incluya procesos no conscientes o independientes de los estados mentales del sujeto epistémico.

menos una elucidación reflexiva¹⁰ del concepto de ‘conocimiento’ o ‘saber’. De modo que, de acuerdo con la manera en que se empleará este concepto, que S sepa que p involucra que: (i) S crea que p , (ii) p sea verdadera y (iii) la creencia de S de que p esté respaldada apropiadamente (en un sentido que no se precisará con detalle aquí) por otras creencias de S .

Esta caracterización es, sin duda, insatisfactoria en varios sentidos. No obstante, excluye algunas de las alternativas arriba descritas. Por una parte, se rechaza una formulación burda del fiabilismo, puesto que la condición (iii) incluye explícitamente estados mentales (creencias) del sujeto epistémico;¹¹ la razón para esto es que se considera que el conocimiento guarda una estrecha relación con la acción racional y las condiciones meramente fiables de adquisición de creencias no parecen preservar este vínculo.¹² Por otra parte, se considera que la sugerencia de Schiffer no dista mucho en sus consecuencias de una prematura abdicación ante el escepticismo. Finalmente, pese a cierta afinidad con la propuesta de Williamson (que es atractiva y merece atención en sus desarrollos), no se adoptan aquí todos sus compromisos, siguiendo la consigna de que en filosofía — como en política — no es prudente ser demasiado optimista ante las promesas futuras.

¹⁰ Asumo que una elucidación reflexiva es el producto de “...trazar relaciones conceptuales entre un concepto dado y otros más” sin que esto equivalga a una ofrecer una definición que haga posible la eliminación del concepto elucidado. Retomo esta terminología de Tagle (2007), aunque ahí se discute en relación con el concepto de ‘verdad’. Creo que, en varios sentidos, esto se asemeja a la propuesta de Williamson, pero éste se ha concentrado en la dirección opuesta de la elucidación: mostrar cómo el de ‘conocimiento’ permite clarificar otros conceptos (como ‘creencia’, ‘verdad’ y ‘evidencia’).

¹¹ Esto no quiere decir que se rechace, sin más, el externalismo en epistemología. Una forma distinta del externalismo en epistemología se discute en las secciones 3.4 y 3.5.

¹² Ésta es una lección que, considero, puede extraerse del contraejemplo de Laurence Bonjour en torno a la clarividencia. Un agente epistémico puede formarse una creencia verdadera a partir de un proceso fiable (la clarividencia, en el ejemplo de Bonjour); sin embargo, si no considera que dicho proceso es fiable, si no tiene creencias que respalden, para el agente, la fiabilidad del proceso (*e.g.*, si no cree que la clarividencia sea una manera racional de adquirir o justificar creencias), parece razonable que no actúe sobre la base de dicha creencia. En este sentido, el fiabilismo rompería los lazos que intuitivamente encontramos entre el conocimiento y la acción racional — *i.e.*, la acción orientada de acuerdo a nuestro conocimiento o nuestras creencias mejor respaldadas (véase Bonjour, 1985: 37-40).

1.1.2. La singularidad del conocimiento de la propia mente

Una vez aclaradas sus similitudes, podemos retornar a la discusión la diversidad de nuestros logros cognoscitivos. Al inicio de esta sección se mencionaron algunas diferencias importantes entre estos logros. De entrada, dijimos que podían concernir a distintos tipos de cuestiones: el mundo y la mente. Esto, a la luz de las aclaraciones precedentes, quiere decir que la oración (que se abrevia con la letra *p* en los esquemas previos) incrustada en las atribuciones de conocimiento o saber, así como en las condiciones correlativas en su *analysans* o elucidación, puede representar estados o sucesos del mundo, o bien estados o sucesos mentales de algún individuo racional (alguien más o uno mismo).

En este punto es conveniente atajar una posible objeción; aunque la retomaremos más adelante, en el apartado 1.4.2. Podría considerarse que es incorrecto (o que no es una práctica ordinaria) incrustar oraciones sobre estados mentales en las cláusulas subordinadas de enunciados de saber o de creencia; *i.e.*, no es común digamos cosas como “Juan sabe que Pedro cree que la Tierra es plana” o “Sé que quiero que la violencia disminuya”. Ciertamente, no es muy común que formulemos oraciones como éstas. No obstante, creo que lo hacemos.¹³ Por otra parte, la falta de evidencia lingüística no debería, sin más, decidir la cuestión acerca de si es correcto o no incrustar oraciones sobre estados mentales en tales cláusulas subordinadas. La razón para ello puede ilustrarse si hacemos uso de una distinción entre atribuciones y adscripciones: las últimas son el correlato lingüístico de las primeras, cuando se verbalizan.¹⁴ En este sentido, aunque las adscripciones de conocimiento o creencias acerca de estados mentales (propios o ajenos) sean poco comunes, las atribuciones correspondientes podrían ser mucho más abundantes, aunque en forma implícita. Considérese la siguiente situación:

¹³ Se dice, *e.g.*, que alguien no sabe lo que quiere, en oposición a otras personas que sí lo saben.

¹⁴ Sigo en esto la distinción “no estándar” de Tyler Burge: “Hablo de *atribuir* una actitud, contenido o noción y de *adscribir* una cláusula que comienza con ‘que’ u otra parte del lenguaje. Las adscripciones son los análogos lingüísticos de las atribuciones” (Burge, 1979: 103).

Durante un curso introductorio de filosofía, el profesor presenta a sus alumnos el problema de la libertad y el determinismo en la filosofía estoica. Mientras expone los argumentos que apoyan la postura determinista, observa atentamente las reacciones que su presentación suscita en uno de sus estudiantes: con la mirada atenta al expositor, manifiesta señas continuas de disgusto. El profesor se dirige a él y, mientras golpea repetidamente su hombro, le formula la siguiente pregunta:

- ¿Puedo elegir dejar de hacer esto?
- Si los argumentos a favor del determinismo son correctos, no puede dejar de hacerlo profesor... O, al menos, eso no tiene nada que ver con lo que usted elija contestar con dificultad, debido a la distracción que le ocasionan los golpes.
- En efecto, eso diría el determinista. Pero... ¿tú qué crees?
- *Creo que puede dejar de golpearme* responde, algo molesto porque no haya concluido aún el ejemplo. Ciertamente, *quiero que deje de hacerlo. Y creo que le convendría*, pues sus golpes comienzan a molestarme y *tengo la intención de responder a sus agresiones* añade en tono decidido y amenazante.

El profesor deja de golpearlo, y se aleja (con cautela) para continuar con su exposición. Presentará ahora las motivaciones a favor de la libertad de acción.

Aunque el ejemplo ilustra una práctica pedagógica que podría considerarse censurable, no nos preocuparemos aquí por la calidad moral de los profesores de filosofía. Más bien, enfatizaremos algunos rasgos del intercambio lingüístico entre alumno y profesor. Aunque en ningún momento aparecen oraciones de saber o de creencia que tengan oraciones sobre estados mentales incrustadas como cláusulas subordinadas (es decir, no hay adscripciones de saber o creencia sobre estados mentales), es conveniente suponer, para explicar racionalmente la conducta del profesor, que éste efectúa las atribuciones correlativas a partir de las respuestas del alumno. Una explicación de esta índole podría ser la siguiente. Cuando el profesor le pregunta al alumno qué es lo que cree, éste profiere oraciones sobre sus creencias y siendo más prolífico de lo que se le solicitaba deseos e intenciones. La información adicional no es gratuita: el alumno la aporta

con el fin de que el profesor se anticipe a sus acciones. Cada una de sus aserciones sobre lo que cree, desea o se propone hacer expresa una creencia sobre sus estados mentales.¹⁵ Aunque ninguna de sus afirmaciones constituye, por sí sola, una agresión por parte del alumno, parece prudente considerarlas como una amenaza. Asumiendo que se encuentra frente a un sujeto racional, el profesor se percata de que si el alumno tiene la intención de hacer algo y cree que puede hacerlo, es muy probable que lo haga. Es por ello que interrumpe sus agresiones y, cautelosamente, se distancia del alumno.

Otra de las distinciones apuntadas al inicio entre los diversos logros cognoscitivos es que algunos dependen de manera especial de nuestras creencias. Esta distinción es relevante si uno considera que la verdad es, en un sentido importante, una condición no epistémica; *i.e.*, si suscribe "...la idea de que la verdad es objetiva; la idea de que la verdad, y por ende la realidad, son (excepto por casos especiales) independientes de lo que cualquiera cree o puede saber" (Davidson, 1988: 184). Si algunas creencias pueden versar acerca de estados mentales (creencias, entre ellos) y en efecto hay sucesos en los que esos estados son satisfechos por un individuo, entonces hay un sentido claro en el que el conocimiento de la propia mente y de otras mentes depende lo que alguien crea: para ser conocimiento tales creencias han de ser verdaderas y, por ende, deben darse los estados mentales (*e.g.*, las creencias) sobre los cuales tratan. En el caso particular del conocimiento de la propia mente (a diferencia del de otras mentes) que una persona sepa que tiene una creencia depende, específicamente, de que *esa* creencia sea una de las que tiene la persona en cuestión.

¹⁵ Esto es lo que sugiere Davidson cuando, al hablar de la diferencia entre decir asertivamente "La nieve es blanca" y decir asertivamente "Creo que la nieve es blanca", señala: "Las condiciones de verdad de las aserciones no son las mismas, pero cualquiera que comprenda la primera aserción sabe las condiciones de verdad de la segunda [...] Esto se debe a que cualquiera que comprende el habla reconoce las aserciones, y sabe que alguien que hace una aserción se representa a sí mismo como creyendo lo que dice" (Davidson, 1991: 209).

Alguien podría objetar, no obstante, que la afirmación de que algunas oraciones sobre estados mentales son verdaderas está comprometida con la existencia de objetos o sustancias inmatrimales: mentes, almas o espíritus. Creo que la manera más simple de responder a esta inquietud es señalando que la cuestión de si las oraciones sobre estados mentales pueden ser verdaderas es, en buena medida,¹⁶ neutral frente a la discusión en torno al problema mente-cuerpo. Decir que una oración en la que se atribuyen estados mentales a un individuo es verdadera equivale a decir que ese individuo cae bajo la extensión de un predicado mental, el cual puede o no identificarse con algún predicado físico, puede o no constituir una clase natural. De modo que la cuestión de si una oración sobre estados mentales puede ser verdadera es compatible con varios puntos de vista en torno a la relación entre mente y cuerpo, como el dualismo, el fisicalismo de corte reduccionista e incluso con algunas formas de no reduccionismo. Es ejemplar la manera en que Davidson presenta este punto: “No hay cosas tales como las mentes, pero las personas tienen propiedades mentales, lo que equivale a decir que ciertos predicados psicológicos son verdaderos de ellas” (Davidson, 1994: 231).¹⁷

El tercer aspecto en el que se señaló que nuestros logros cognoscitivos manifiestan diversidad parece involucrar a la tercera condición de nuestro análisis o elucidación del conocimiento, la cuestión del respaldo (o justificación, en la terminología tradicional) que deben poseer nuestras creencias verdaderas para contarse como casos de conocimiento. Es en este aspecto en el que el conocimiento de la propia mente manifiesta su singularidad, pues, a diferencia del conocimiento del

¹⁶ Digo “en buena medida” porque algunas posiciones extremas, como el materialismo eliminativo, defenderán simple y llanamente que las adscripciones de estados mentales que emplean el vocabulario de la psicología popular (*i.e.*, que hablan de creencias, intenciones, deseos y estados psicológicos similares) son siempre (y todas) falsas (véase Churchland, 1981 y 1988). Lo que afirmo a continuación es, ciertamente, incompatible con posiciones como éstas.

¹⁷ Sobre decir que Davidson tiene una posición respecto a estas cuestiones (explícitamente defendida en sus 1970 y 1993), como señala poco más adelante del pasaje citado: “Los sucesos mentales son, desde mi punto de vista, físicos (lo que no equivale, por supuesto, a decir que no son mentales)” (Davidson, 1994: 231).

mundo y de otras mentes, no parece requerir la clase de observación e inferencia que éstos precisan. No obstante, para apreciar en qué medida esto caracteriza al conocimiento de la propia mente se requiere que hagamos algunas precisiones sobre su alcance y sus peculiaridades.

1.2. El alcance y los rasgos característicos del conocimiento de la propia mente

Hasta ahora se han indicado algunas razones para considerar que es legítimo hablar de conocimiento (en el sentido antes descrito) con respecto a algunos aspectos nuestra propia vida mental. Sin embargo, no se ha precisado de qué aspectos se trata (*i.e.*, cuál es el alcance de dicho conocimiento) ni se ha especificado con claridad en qué sentido esta forma de conocimiento es especial y, por ende, digna de atención filosófica. En lo que sigue se harán algunas precisiones sobre estos puntos que pondrán de manifiesto su interdependencia.

1.2.1. Alcances y límites del conocimiento de la propia mente

Aunque se ha hablado del conocimiento de “la mente”, a estas alturas debería ser claro con esto no se pretende hablar del conocimiento de una entidad o sustancia. Debería ser claro porque ya se ha señalado que nuestro interés se ubica en el conocimiento o saber proposicional sobre estados mentales, no en el saber sobre objetos (“conocer algo o a alguien”) o en el saber operativo (“saber hacer tal o cual”). De modo que el alcance del conocimiento de la propia mente no se extiende a lo que puede entenderse bajo la consigna délfica “conócete a ti mismo” (*γνωθι σεαυτόν*), que Sócrates hiciera suya la cuestión de si sabemos quiénes

somos.¹⁸ Tampoco involucra el conocimiento del “yo” como cierta clase de objeto privado al cual sólo uno podría dirigir su atención. Con su habitual agudeza, Hume reorientó la atención filosófica del conocimiento de la propia mente, entendido como el conocimiento de una entidad, al conocimiento de los propios estados mentales cuando indicó:

Por mi parte, cuando me adentro más íntimamente en lo que denomino *mi yo* [*myself*], siempre tropiezo con alguna percepción particular u otra, de calor o frío, de luz o sombra, de amor u odio, de dolor o placer. Nunca puedo atrapar a mi yo en ningún momento sin una percepción, y nunca puedo observar nada sino la percepción. (Hume, 1739/1969: 300)

Siguiendo el espíritu de esta observación humeana, al hablar del conocimiento de la propia mente nos referiremos al conocimiento que cada uno de nosotros tiene de sus pensamientos o estados mentales. Esta forma de conocimiento parece ser peculiar frente a la forma en que otras personas pueden saber lo que pensamos.

Uno de los rasgos más sobresalientes del conocimiento de la propia mente es que no parece requerir el mismo tipo de respaldo del que depende nuestro conocimiento del mundo externo y de las otras mentes. A diferencia de estas formas de conocimiento, el de la propia mente involucra una presuposición especial de verdad, no parece depender de inferencias a partir de la observación ni es, generalmente, propenso al tipo de escrutinio crítico al que son vulnerables otros logros epistémicos. No obstante, estos rasgos no se aplican a todas nuestras autoatribuciones de estados mentales. Como extensamente lo discute Gertler (2011: 59-86), sobre buena parte de nuestra vida mental no hay una presuposición de que estemos en lo correcto ni podemos prescindir de evidencia derivada de la observación de la conducta. Lo que se presume queda fuera de este dominio privilegiado incluye “...nuestra posesión de ciertos rasgos de carácter o temperamento, nuestras posi-

¹⁸ Como se presenta en (Platón, *Fedro*: 229e): “...siguiendo la inscripción de Delfos, no he podido conocerme a mí mismo. Me parece ridículo, por tanto, que el que no se sabe todavía se ponga a investigar lo que ni le va ni le viene”.

bles reacciones afectivas ante ciertas situaciones, las fuentes causales de nuestros sentimientos y acciones, nuestras emociones y estados anímicos, y nuestras actitudes latentes o disposicionales” (Gertler, 2011: 82). Si todo lo anterior queda fuera de su alcance, ¿sobre qué aspectos de nuestra psicología sería razonable esperar que tengamos un tipo de conocimiento que fuese filosóficamente especial e interesante? El consenso entre los filósofos es que los límites de nuestro conocimiento privilegiado sobre la propia mente se extienden, a lo sumo, a los siguientes dominios:

- (i) sensaciones y rasgos de la experiencia fenoménica consciente,
- (ii) pensamientos y actitudes proposicionales,¹⁹ y
- (iii) la posición relativa en el momento actual de las partes del propio cuerpo (véase Gertler, 2011: 82 y Goldberg, 2011: 306).

Pese a sus peculiaridades, no parece problemático asimilar el tercer grupo de estados mentales al primero (lo hace, *e.g.*, Gertler, 2011). Después de todo, se trata de cierto tipo de sensaciones. Sin embargo, el consenso filosófico en la actualidad es que hay diferencias significativas entre los miembros del primer grupo y los del segundo. La literatura en torno al primer grupo ha proliferado en las últimas décadas de manera inusitada.²⁰ No obstante, en la discusión subsiguiente me limitaré a hablar del segundo grupo de estados mentales, *i.e.*: las actitudes proposicionales.

¹⁹ Como se indicó en el párrafo anterior, esto no incluye estados latentes o disposicionales; también excluiré estados pasados (en los que interviene la memoria) o futuros (que, a menudo, involucran inferencias de carácter teórico o práctico). En lo que sigue al hablar de atribuciones de pensamientos y actitudes proposicionales asumiré que nos referimos a estados mentales *presentes* (*i.e.*, contemporáneos al momento de atribución) a menos que se indique lo contrario.

²⁰ Este interés no es accidental, pues se trata de estados mentales que han resistido encarnizadamente intentos de reducción en lo concerniente al debate mente-cuerpo (a manera de ilustración véase, la selección de artículos que integran el volumen *La naturaleza de la experiencia*, 2003). Otra de las motivaciones tras este interés queda bellamente ilustrada en el siguiente pasaje de Jaegwon Kim (2006: 11): “Para la mayoría de nosotros, una vida plena, una vida que vale la pena vivir, es una que es rica en y está llena de consciencia cualitativa. Consideraríamos a una vida como empobrecida y no completamente satisfactoria si nunca incluyera experiencias de cosas como el aroma del océano en una fresca brisa matutina, la fragancia de un campo de lavandas en flor, y el vibrante y estratificado paisaje sonoro proyectado por un cuarteto de cuerdas”.

La razón para esto es que a menudo la discusión sobre el conocimiento de los estados mentales del primer grupo no concierne al saber proposicional, sino al saber sobre objetos o al saber operativo.²¹

Algunas precisiones son requeridas en este punto. Los estados mentales agrupados en el dominio que hemos especificado habitualmente se describen por medio del uso de verbos psicológicos como ‘pensar’, ‘creer’, ‘esperar’, ‘temer’, ‘sospechar’, ‘suponer’, ‘imaginar’ o ‘dudar’. La anterior dista mucho de ser una lista exhaustiva. Además, como definición, tiene el inconveniente de ser patentemente circular (véase Davidson, 1975 y Hacker, 1997 para una caracterización en líneas similares). Por ello, será conveniente introducir alguna terminología técnica.

Fue el filósofo austriaco Franz Brentano quien, en su libro *Psychologie vom empirischen Standpunkt*, introdujo la noción de ‘intencionalidad’ con el fin de echar luz sobre este asunto. Señalaba que lo que distingue a los fenómenos psíquicos (pensamientos o estados mentales, en la terminología de nuestra discusión), de los fenómenos físicos es que los primeros tienen un carácter intencional:

Todo fenómeno psíquico está caracterizado por lo que los escolásticos de la Edad Media han llamado la inexistencia intencional (o mental) de un objeto, y que nosotros llamaríamos, si bien con expresiones no enteramente inequívocas, la referencia a un contenido, la dirección hacia un objeto (por el cual no hay que entender aquí una realidad), o la objetividad inmanente.

[...] Esta inexistencia intencional es exclusivamente propia de los fenómenos psíquicos. Ningún fenómeno físico ofrece nada semejante. [...] Podemos definir los fenómenos psíquicos diciendo que son aquellos fenómenos que contienen en sí, intencionalmente, un objeto (Brentano, 1874: 81-82. Sigo la traducción de Gaos).

La terminología de Brentano es propensa a producir confusiones. De su caracterización retomaremos únicamente la indicación de que los estados mentales refieren a un contenido; éste puede especificarse, en adscripciones explícitas de

²¹ Sobre este tipo de consideraciones véanse, *e.g.*, las perspicaces sugerencias en Lewis (1988).

estados mentales, por medio de oraciones declarativas precedidas de la preposición ‘que’. Los fenómenos psíquicos, o al menos buena parte de ellos,²² son, pues, intencionales en tanto refieren a un contenido.²³ Bertrand Russell, en «La filosofía del atomismo lógico», ofreció una caracterización similar y más afín a la nuestra al hablar de actitudes proposicionales (véase Russell, 1918: 192-193). La expresión lingüística ‘actitudes proposicionales’ tiene la ventaja de que, aunado a que captura la intuición de que el rasgo fundamental de los estados mentales es poseer contenido, puntualiza que “el contenido es lo que se especifica por oraciones o proposiciones en cláusulas-que cuando atribuimos estados intencionales a agentes” (Bilgrami, 1992: 1).²⁴

El conocimiento de la propia mente parecería involucrar, de este modo, que tengamos una actitud proposicional cuyo contenido sea, a su vez, una actitud proposicional. Requiere que haya creencias (*i.e.*, un tipo de actitudes proposicionales) acerca de otros estados mentales (que aquí hemos limitado a actitudes proposicionales).²⁵ A partir de las precisiones precedentes podemos ahora examinar

²² Las excepciones más notables, aquéllas con las que Brentano enfrentó mayor dificultad en todo caso (las sensaciones de dolor, el cosquilleo, las punzadas de los celos, entre otras), fueron previamente excluidas de nuestra discusión.

²³ Una discusión introductoria, sucinta pero bien informada, de la ‘tesis de Brentano’ puede encontrarse en Crane (2008: 73-78).

²⁴ Ésta no es la única manera de caracterizar a los estados mentales intencionales, aunque probablemente sea la más popular. En lo que sigue asumiremos una caracterización como ésta. Considero que la defensa de lo que se defiende esta disertación se vería robustecida si se adoptara el análisis paratáctico de las adscripciones actitudes proposicionales (véase Davidson, 1968, 1975 y 1989). Según este análisis, la forma lógica de las adscripciones involucra la preferencia de dos oraciones ‘S cree esto’: ‘*p*’. No obstante, creo que adoptar esta caracterización complicaría innecesariamente la discusión. Además, como Guido Filippi hizo (lamentablemente) explícito en su traducción de (Davidson, 1968), hay un aspecto del análisis paratáctico (la función de la partícula ‘*that*’) que no es intuitivamente compartido por las adscripciones expresadas en lengua española.

²⁵ Otras combinaciones de actitudes proposicionales son posibles. Podemos desear tener creencias distintas a las que de hecho tenemos (si consideramos que son en parte responsables de nuestra miseria actual recuérdese, *e.g.*, el *dictum* anglosajón: «*Ignorance is bliss*»), o que nuestras intenciones hubiesen sido laudables (cuando, de hecho, no lo fueron); podemos temer que nuestros deseos nos destruyan, o que nuestras creencias sean erróneas. Puesto que la discusión actual refiere únicamente al caso del *conocimiento* de la propia mente, nos limitaremos a la discusión de creencias sobre actitudes proposicionales.

qué se ha considerado que hace especial al conocimiento de la propia mente y por qué se considera que amerita alguna clase de escrutinio filosófico.

Aunque las atribuciones de conocimiento de la propia mente permean muchas de nuestras prácticas ordinarias,²⁶ surgen tensiones cuando intentamos ser más específicos acerca de en qué sentido esta forma de conocimiento es especial y qué es lo que da cuenta de su peculiar estatus epistémico (examinamos algunas explicaciones robustas de esto último en las secciones 1.3 y 1.5). Estas cuestiones, por sí mismas, generan ya cierta perplejidad filosófica. Sin embargo, el conocimiento de la propia mente también ha recibido atención por parte de los filósofos debido a que parece un lugar propicio para fundamentar el resto de nuestras creencias.²⁷ Después de todo, mientras nuestro conocimiento del mundo y de las otras mentes parece requerir observación y/o inferencias para conectar a la creencia con aquello que la hace verdadera, en el caso del conocimiento de la propia mente, por otra parte, se ha supuesto que el estrecho vínculo entre la creencia y su contenido parece cerrar la brecha que existe entre creencia y verdad, entre apariencia y realidad. Pero, ¿en qué sentido es especial el conocimiento de la propia mente?

1.2.2. Autoridad de la primera persona y asimetría epistémica

Hay dos rasgos, ambos al parecer relacionados con la tercera condición para el saber, en los que el conocimiento de la propia mente parece ser excepcional. El primero es que las adscripciones de estados mentales en primera persona (que, presumiblemente, expresan creencias acerca de los propios estados mentales) tienen un tipo especial de autoridad, que la proferencia de dichas adscripciones proporciona una razón para asumir la presencia del estado mental en cuestión.

²⁶ Una muestra ejemplar de ello se encuentra en Ryle (1949: 163-167). Aunque quizá algunos de los casos que ahí se presentan sean exclusivos de la lengua inglesa, como nos advierte Eduardo Rabossi en su traducción de *The Concept of Mind* (véase la versión castellana 2005: 186, nota del traductor).

²⁷ Un examen de esta concepción del papel del conocimiento de la propia mente en la epistemología se presentan en la sección 3.2.

Llamaremos a este rasgo ‘autoridad de la primera persona’. El segundo es que la atribución en primera persona de estados mentales en el momento actual, a diferencia de la atribución de estados mentales a otras personas, es directa o no está mediada por la observación e inferencia a partir de la conducta. Denominaremos a este rasgo ‘asimetría epistémica’. Ambos rasgos precisan ciertas cualificaciones.²⁸

Existen matices en el estatuto que se otorga a la autoridad de la primera persona. Puede considerarse, en primer lugar, la posición *infalibilista*. El infalibilismo sobre la atribución de estados mentales sostiene que siempre que un sujeto epistémico normal (que no adolece de defectos cognitivos) efectúa una atribución de estados mentales en primera persona ésta goza de una garantía de verdad; *i.e.*, no puede estar equivocado al respecto. Esto situaría en las atribuciones de estados mentales en primera persona en una categoría única, pues haría que *todas* las atribuciones de estados mentales fuesen verdaderas (siempre y cuando se efectuaran de manera sincera, por sujetos epistémicos competentes).

No obstante, hay razones para ser más reservados con respecto al estatuto de la autoridad de la primera persona. Una de ellas es que parece haber circunstancias en las que de hecho nos equivocamos al atribuirnos estados mentales.²⁹ Puede pensarse, *e.g.*, en afirmaciones como la siguiente: “Claro, él *dice* que cree que las mujeres son iguales a los hombres con respecto a las habilidades que el trabajo requiere, pero no creo que él realmente lo crea: nunca ha contratado a una mujer para este puesto en sus treinta años de estar a cargo” (el ejemplo es de Goldberg, 2011: 307). Otra razón para rechazar el infalibilismo es de naturaleza conceptual. Como Wittgenstein señaló, parece haber una íntima conexión conceptual entre el saber y la posibilidad del error.³⁰ La raíz de esta conexión parece situarse en las

²⁸ Lo que se presenta a continuación sigue de cerca las distinciones y caracterizaciones que emplean Goldberg (2011: 306-308), Gertler (2011: 60-70) y Gultenplan (1994a: 291), aunque existen algunas diferencias terminológicas entre estas presentaciones y de ellas con lo aquí expuesto.

²⁹ Esto es parte de lo que se argumenta en el capítulo 2.

³⁰ Que Wittgenstein sostuvo este punto de vista se insinúa en los siguientes pasajes: “«Sé...» puede querer decir «No dudo...» pero no quiere decir que la frase «Dudo...» sea *carente de sentido*,

condiciones para la adquisición misma del lenguaje: aprender a emplear un concepto implica que éste tenga condiciones (públicas) de corrección y, por ende, que sea posible emplearlo incorrectamente.³¹ Si esta intuición es correcta no tendría sentido hablar de saber donde no fuesen posibles el error y la duda.

Una posición más moderada que el infalibilismo con respecto a la autoridad de la primera persona consiste en considerar que las atribuciones de estados mentales en primera persona son *incorregibles*. La tesis de la incorregibilidad señala que, aunque el error es posible, un tercero no puede mostrarnos que dichas atribuciones son erróneas. No es del todo clara cuál sería la importancia epistemológica (en qué sentido establecería un vínculo entre creencia y verdad) de la tesis de la incorregibilidad (véase Goldberg, 2011: 307). Asimismo, aunque este estatuto es más débil que el de la infalibilidad, aún parece establecer demasiado. Parece excluir la posibilidad de revisar las atribuciones de estados psicológicos, ya que no es claro que tenga sentido afirmar que son sólo corregibles desde la perspectiva de la primera persona. Después de todo, corregir una atribución es una práctica normativa y, por ende, social; no parece que tenga sentido decir que es privada.³²

Un sentido más apropiado en el que puede hablarse de la autoridad de la primera persona, que captura lo que nos parece especial acerca del conocimiento de la propia mente, consiste en señalar que las atribuciones de estados mentales en primera persona gozan de una *presunción de verdad*, *i.e.*, que cuando alguien se adscribe de manera sincera un pensamiento particular dicha adscripción justifica la suposición de que efectivamente tiene ese pensamiento. Aunque el error y la corrección son posibles, en general consideramos que alguien que se adscribe

que la duda esté excluida lógicamente. / Decimos «Sé» cuando también podemos decir «Creo» o «Supongo»; cuando podemos convencernos de algo.” (Wittgenstein, 1953: II, p. 505) “Podría suceder que siempre que dijera «Lo sé», se revelara falso.” (Wittgenstein, 1969: § 580)

³¹ Éste, me parece, es un corolario que pueden extraerse del argumento contra el lenguaje privado.

³² “Por tanto, ‘seguir la regla’ es una práctica. Y creer que se sigue la regla no es seguir la regla. Y, por tanto, no se puede seguir ‘privadamente’ la regla, porque de lo contrario creer que se sigue la regla sería lo mismo que seguir la regla” (Wittgenstein, 1953: § 202).

un pensamiento hace una atribución verdadera. La presunción de verdad ciertamente introduce una diferencia entre las atribuciones de estados mentales en primera persona y las atribuciones en tercera persona; y la diferencia es epistemológica. La diferencia radica en que cuando se hace una adscripción verdadera de estados mentales en primera persona, la creencia que esta adscripción representa (la atribución correspondiente) *cuenta como conocimiento*. Esto no ocurre con las atribuciones de estados mentales a otras personas: para que una adscripción de estados mentales a un tercero *cuenta como conocimiento* exigimos que, además de que la adscripción sea verdadera, quien la efectúa tenga evidencia suplementaria para respaldarla (*e.g.*, de su conducta).

Pasemos ahora a discutir la asimetría epistémica. Frecuentemente se caracteriza a este rasgo como una condición de acceso privilegiado, entendido como el análogo de lo que desempeña el papel de la observación y la inferencia en el caso del conocimiento de otras mentes. Así, parecería que la condición más robusta es la *omnisciencia* o *luminosidad* (como la ha denominado Williamson en 2000 y 2011). Decir que los estados mentales gozan de luminosidad para la primera persona equivale a señalar que siempre que un sujeto racional se encuentra en un estado mental se atribuirá tal estado mental. A veces se expresa esto diciendo que una persona “intima” con sus propios estados mentales (aunque no es del todo claro lo que esto quiera decir). En todo caso, hay razones para suponer que la luminosidad es una condición demasiado robusta.³³ Considérese, *e.g.*, la siguiente afirmación: “Su conducta sugiere que ella cree que los hombres son incompetentes en el cuidado de los niños aunque ella nunca admitiría tal cosa, incluso para sí misma” (el ejemplo, nuevamente, es de Goldberg, 2011: 308). Algunos estados mentales inconscientes (como deseos o ‘pulsiones’, pero también creencias

³³ Williamson mismo (tanto en su 2000: chap. 1, como en su 2011: 212) ofrece casos en los que la condición de luminosidad no se cumple; no obstante, han sido dejados de lado puesto que él considera la no luminosidad de ciertos estados fenoménicos, los cuales han sido excluidos de la presente discusión.

no declaradas) parecen desempeñar un papel importante en la acción o, por lo menos, suelen ser buenos expedientes para la explicación de la acción (recuérdese el ejemplo del profesor, en el apartado 1.1.1). Negar, en defensa de la luminosidad, que se trate de estados mentales auténticos menguaría de manera significativa nuestros recursos explicativos.

Una condición más débil de acceso es la condición de *atención selectiva*. Ésta explota una analogía entre el conocimiento de la propia mente y el del mundo al asumir que la información que nos permite generar creencias sobre los propios estados mentales, al igual que con la que formamos creencias sobre el mundo externo, puede ser fragmentaria. Una forma de ilustrar esto es la situación familiar de “conducir en automático”: al conducir un automóvil uno es capaz permanecer en el camino, no impactarse con otros autos (o peatones) e incluso sostener una conversación, sin prestar atención a todos los detalles del entorno visual. “Uno puede «llegar a» algún punto y percatarse de que ha conducido muchas millas sin consciencia de conducir o, quizá, de nada más” (Armstrong, 1968: 93). En esta analogía, cierta clase de ‘sentido interno’ nos proporcionaría información sobre los propios estados mentales (tal como la percepción, o ‘sentido externo’ proporciona evidencia sobre algunos de los sucesos que acaecen en nuestro entorno). Esto permitiría, en principio, que el conocimiento de la propia mente fuese tanto incompleto como falible. Sin embargo, la base misma de esta analogía parece objetable: no hay órganos introspectivos, análogos a los sentidos de la percepción externa, por lo que no es en absoluto claro a qué se deberían las limitaciones y la falibilidad del conocimiento que, presuntamente, proporcionarían. Además, la tesis del ‘sentido interno’ enfrenta a otra clase de preocupación: en el caso de la percepción, un criterio de decisión para la aplicación de conceptos son los estándares intersubjetivos de corrección; no es claro que la hipótesis del sentido interno pueda tener un correlato análogo.

Hay una razón más general para rechazar la suposición de que tenemos una forma de acceso privilegiado (sea ésta luminosidad o atención selectiva) a los propios estados mentales esencialmente distinta a la que empleamos para atribuir estados mentales a otras personas. Esta suposición, en el mejor de los casos, amenazaría la posibilidad de que exista univocidad en el uso que hacemos de los conceptos psicológicos (de que lo que designamos por medio de los verbos psicológicos sea lo mismo en el caso de la primera y la tercera personas). En el peor de los casos, esta suposición conduciría a una forma radical de escepticismo sobre las otras mentes.³⁴ En todo caso, no parece necesario entender la asimetría epistémica como una condición de acceso privilegiado. Lo que se requiere es una explicación del carácter de no mediado que tienen las atribuciones de estados mentales; dar cuenta de esto forma parte de lo que se requiere para ofrecer una explicación satisfactoria de la autoridad de la primera persona.

De modo que hemos señalado dos rasgos que, incluso en sus formulaciones más débiles, hacen especial al conocimiento de la propia mente: la autoridad de la primera persona, entendida como una presunción de verdad que otorga justificación a las autoadscripciones de actitudes proposicionales; y la asimetría epistémica, que consiste en el carácter no mediado por observación de la conducta e inferencia, presente en las atribuciones de estados mentales en primera persona. No obstante, indicar que el conocimiento de la propia mente manifiesta autoridad de la primera persona y asimetría epistémica es distinto de ofrecer una explicación de por qué posee estos rasgos.

En lo que resta de este capítulo examinaremos algunas explicaciones sustantivas del conocimiento de la propia mente y consideraremos las dificultades que

³⁴ Recuérdese el famoso ejemplo del escarabajo de Wittgenstein: “Supongamos que cada uno tuviera una caja y dentro hubiera algo que llamamos «escarabajo». Nadie puede mirar en la caja de otro; y cada uno dice que él sabe lo que es un escarabajo sólo por la vista de *su* escarabajo. Aquí podría muy bien ser que cada uno tuviese una cosa distinta en su caja [...] la caja podría incluso estar vacía” (1953: § 293).

enfrentan. En la sección 1.3 se presentan algunas concepciones epistémicamente robustas, que consideran a la noción de ‘acceso privilegiado’ como desempeñando un papel central en la explicación de la autoridad de la primera persona. La sección 1.4 se ocupa de explicaciones que se ubican en las antípodas de las anteriores, al negar un estatuto epistémico a la autoridad y la asimetría. En la sección 1.5 expongo lo que considero una alternativa más razonable, pues preserva el carácter epistémico de la autoridad de la primera persona sin depender de una condición de acceso privilegiado; me refiero, por supuesto, a la posición davidsoniana. Finalmente, en la sección 1.6 se presentan algunas de sus dificultades más significativas que enfrenta esta posición.

1.3. Las fuentes de la autoridad de la primera persona I: explicaciones de acceso epistémico

Aunque sus tesis han sido objeto de numerosas y severas críticas, puede afirmarse que fue Descartes quien formuló gran parte de los problemas que han ocupado a la filosofía desde el siglo XVII hasta nuestros días.³⁵ Uno de ellos —y no el menos importante— concierne al conocimiento que puede tener una persona de sus estados mentales. Descartes situó al conocimiento de la propia mente al centro de su proyecto filosófico; en esto pareció seguirlo buena parte de la tradición filosófica posterior.³⁶ La certeza y la especial conexión con la verdad que se le asignaba

³⁵ Aunque quizá ésta sea una afirmación extravagante en lo que concierne a la totalidad del quehacer filosófico, parece aplicarse con cierta plausibilidad a dos áreas de interés predominante: la epistemología y la filosofía de la mente. En especial, en lo que respecta al conocimiento de la propia mente, Descartes parece haber dado un giro importante a la discusión, desde los asuntos preponderantemente éticos que ocupaban a la filosofía griega, a cuestiones de corte epistémico y metafísico (*Cfr.* Gertler, 2011: 28).

³⁶ Con importantes acotaciones, podría sugerirse que la primacía del conocimiento de la propia mente en la empresa epistemológica persiste en pensadores tan influyentes como Locke, Berkeley, Hume y Kant, así como en el positivismo lógico y la fenomenología de inicios del siglo XX. En el capítulo 3 se sugiere que éste es un rasgo dominante de lo que denomino ‘epistemología cartesiana’. Aún es común, para aproximarse a esta tradición, introducir el problema del conocimiento con

a este tipo de conocimiento parecían concederle un estatuto apropiado para constituir el fundamento del resto de nuestro conocimiento concerniente tanto al mundo externo como a las otras mentes. Aunque una exposición más detallada del proyecto epistemológico cartesiano se presenta en el capítulo 3, es conveniente hacer algunas acotaciones.

1.3.1. La concepción cartesiana

Es común adjudicarle a Descartes las posiciones más robustas de entre las arriba descritas, a saber: que la autoridad de la primera persona es *infalible* y que lo que explica la asimetría epistémica es la *luminosidad* u *omnisciencia* de los propios estados mentales. Podemos llamar a esta combinación de posiciones la tesis de la *transparencia*.³⁷ Probablemente esta concesión no resista el escrutinio exegético.³⁸ Sin embargo, no me ocuparé aquí de la difícil tarea de intentar desentrañar cuál era exactamente el punto de vista de este autor en torno al asunto, ni de examinar si tal concepción es siquiera consistente.³⁹ Lo que haré será más bien exponer y

afirmaciones como la siguiente, claramente afín al problema cartesiano: “Si piensas al respecto, el interior de tu propia mente es la única cosa de la que puedes estar seguro” (Nagel, 1987: 8).

³⁷ Es conveniente aclarar que la denominación ‘tesis de la transparencia’ también se ha empleado en esta discusión para caracterizar a otra posición con respecto al conocimiento de la propia mente. De acuerdo con esta terminología, se dice que un estado mental es *transparente* si tiene la cualidad de “...permitir al sujeto captar el estado mental al considerar su objeto más que el estado mental mismo. *E.g.*, si las creencias son transparentes puedes captar que crees que está lloviendo al «mirar a través de» la creencia (transparente) y reflexionar directamente sobre su objeto, *viz.*, la lluvia” (Gertler, 2011: 281). Aunque no discutiremos esta tesis en particular, sí hablaremos de algunos puntos de vista que parecen apoyarse en ella en el apartado 1.4.2 (véase la nota 50).

³⁸ Hay indicios claros de que Descartes no suscribiría una posición tan robusta como la luminosidad en lo que concierne a la asimetría epistémica. *E.g.*, en el *Discours de la méthode* (1637: Partie III) señala: “...hay pocas personas que deseen decir todo lo que piensan, [...] muchas lo ignoran, pues siendo diferente el acto del pensamiento en virtud del cual se cree algo de aquel otro por el cual se conoce que se tiene tal creencia, frecuentemente se da el uno sin el otro.” [El pasaje es el siguiente: «...il y a peu de gens qui veulent dire tout ce qu'ils croient, [...] plusieurs l'ignorent eux-mêmes, car l'action de la pensée par laquelle on croit une chose, étant différente de celle par laquelle on connaît qu'on la croit, elles sont souvent l'une sans l'autre.»] No obstante, en los apartados 3.2.2 y 3.2.3 se sugiere que el estatuto de infalibilidad es una exigencia del proyecto epistemológico cartesiano.

³⁹ Un examen de esta índole se realiza en Broughton (2007).

evaluar un conjunto de ideas en torno al conocimiento de la propia mente que fueron inspiradas por la filosofía de Descartes, a las cuales denominaré la ‘concepción cartesiana’ del conocimiento de la propia mente.

La concepción cartesiana considera que la forma en que tenemos acceso al contenido de nuestras mentes es enteramente distinta de la manera en que conocemos el mundo extenso: podemos conocer lo que pensamos, sentimos y queremos con un tipo especial de certeza, ausente en el falible conocimiento que los sentidos nos proporcionan del mundo físico y de las personas, las otras mentes, que lo habitan. Varias metáforas ilustran esta forma especial de acceso privilegiado: se dice que conocemos nuestros propios pensamientos a través del ‘ojo de la mente’, o que son objetos en ‘el teatro cartesiano’ o que ‘le son dados a la mente’; como veremos, de estas metáforas se desprenden explicaciones distintas de la autoridad de la primera persona. Cabe señalar que este doble estándar epistémico va aparejado en la filosofía cartesiana a una dualidad ontológica entre sustancias extensas y sustancias pensantes; pero no se requiere adoptar la tesis ontológica para suscribir la dualidad epistémica.⁴⁰

Puede considerarse que de la tesis del doble estándar epistémico se desprenden dos posibles tipos de asimetría del conocimiento de la propia mente: (A_1) con respecto al conocimiento del mundo externo y el de otras mentes; y (A_2) únicamente con respecto al conocimiento de otras mentes. La asimetría (A_1) apunta a una diferencia importante en *la manera en que obtenemos información* relativa a nuestras propias mentes por oposición a la forma en llegamos a saber cosas acerca

⁴⁰ Recientemente se ha sugerido que la intuición cartesiana del *cogito* conlleva *necesariamente* alguna forma de dualismo sustancial. *E.g.*, Treanor (2006: 248) señala que: “...nos estamos engañando si pensamos que podemos aceptar esta conclusión cartesiana [la eficacia del *cogito*] sin estar aceptando alguna forma del dualismo sustancial”. Sin embargo, aunque podría considerarse que la certeza del *cogito* involucra la autoridad de la primera persona, aceptar esta autoridad no implica –al menos no si se la concibe en cierta forma y es aceptada en cierto grado– suscribir la eficacia del *cogito* y congraciarse con el dualismo sustancial. Es importante mencionar –aunque esto es algo que aún sigue siendo tema de discusión– que la supuesta certeza que el *cogito* proporciona a un individuo pensante de su propia existencia no se deriva *tout court* de la autoridad epistémica con respecto a los propios estados mentales.

del mundo externo: mientras en el primer caso no parece haber intermediarios casuales ni epistémicos, sí parece haberlos en el segundo. La asimetría (A_2) apunta al *tipo de recursos* que nos es proporcionado sobre nuestras mentes, en contraste con la que obtenemos de las otras mentes: para atribuirnos estados mentales parece que sólo empleamos ‘datos internos’, mientras que para atribuírselos a otras personas recurrimos a la observación (de lo que el otro dice o hace).

Distinguiremos dos grupos de explicaciones de la autoridad de la primera persona que se desprenden de la concepción cartesiana, dependiendo de a cuál de las asimetrías de acceso arriba indicadas pretenden hacer justicia. Llamaremos al primer grupo la teoría de la *familiaridad* o *contacto directo* (*‘acquaintance’*, en lengua inglesa), que pretende preservar tanto (A_1) como, por extensión, (A_2). Al segundo grupo lo denominaremos la teoría de la *introspección* o *sentido interno*, cuyo compromiso explícito es únicamente con (A_2). Ambos grupos de teorías tienen adherentes contemporáneos.⁴¹ A continuación se expondrán los rasgos centrales de ambos grupos de teorías explicativas y algunas de las objeciones que se les presentan (para una exposición pormenorizada véase Gertler, 2011: 87-165).

1.3.2. La teoría de la familiaridad (*acquaintance*)

La teoría de la familiaridad señala, en términos generales, que tenemos un tipo de contacto (metafísico) directo con algunos de nuestros estados mentales, y éste proporciona una robusta justificación no inferencial a nuestras atribuciones de dichos estados. De modo que, de acuerdo con la teoría de la familiaridad, nues-

⁴¹ Gertler (2011) incluye entre los defensores de la teoría de la familiaridad a Bonjour (2003), Chalmers (2003), Conee (1994), Fales (1996), Feldman (2004), Fumerton (1995), Gertler (2001), Levine (2007) y Pitt (2004); entre los partidarios de la teoría del sentido interno se encuentran Armstrong (1968), Goldman (2006), Lormand (1996), Lycan (1996) y Nichols & Stich (2003). Aunque Gertler denomina a estas posiciones ‘empiristas’, creo que eso podría generar confusiones sobre su raíz común en la concepción cartesiana; Goldberg (2011) las denomina ‘constructivas’. En ambos grupos de teorías se discute el conocimiento de los propios estados fenoménicos; por las razones expuestas en 1.2.1, dicha discusión no se considera aquí.

tros pensamientos desempeñarían la doble tarea de: (a) ser aquello con lo que se confrontan nuestras creencias de que tenemos tales pensamientos, desempeñando un rol metafísico (el de hacedores de verdad [*truth-makers*]); y (b) justificar dichas creencias, realizando una labor epistémica. Esta tarea dual explicaría, a la vez, qué origina la autoridad de la primera persona, a saber: que, en lo que concierne a los juicios sobre los propios pensamientos, no existe brecha entre apariencia y realidad (véase Gertler 2011: 95). De ser exitosa, esta teoría también mostraría por qué esta autoridad epistémica está tan estrechamente ligada a la primera persona: cada uno de nosotros estaría justificado de manera especial al atribuirse a sí mismo estados mentales porque se trata de *sus* estados mentales (y, por ende, está en contacto directo con ellos).

Es importante notar que la teoría de la familiaridad no es en absoluto atractiva si se considera que la autoridad de la primera persona tiene el estatus de transparencia (*i.e.*, infalibilidad y omnisciencia).⁴² En efecto, como ya había objetado Hobbes a Descartes atribuyéndole, al parecer, una posición similar, en esta teoría existe la amenaza de un regreso: “...aún cuando alguien pueda pensar que ha pensado [...], es sin embargo enteramente imposible pensar que se piensa o saber que se sabe, pues, en efecto, la pregunta podría ser infinita: ¿en virtud de qué sabes que sabes que sabes, etc.?” (Hobbes, en Descartes p. 141). Si uno conociera infaliblemente todos sus estados mentales, debería también saber que cree que los tiene y que cree que cree que los tiene, y así *ad infinitum*. No obstante, la teoría de la familiaridad *no implica* transparencia, sino que también es compatible con posiciones más moderadas.

Una de las inquietudes que suscita esta teoría explicativa del conocimiento de la propia mente tiene un origen distinto, relacionado con la función dual que atribuye a los estados mentales. No es claro cómo la relación que se le asigna a los

⁴² En todo caso, ya se han ofrecido razones, en la sección 1.2.2, para negar este estatuto a cualquier explicación de la autoridad de la primera persona y de asimetría epistémica.

estados mentales y a las atribuciones de éstos que se realizan en primera persona pueda ser epistémica (al menos, en el sentido en que aquí se ha caracterizado al conocimiento). Pues, ¿cómo puede el ‘contacto’ ser una forma de justificación? Generalmente, se ha considerado que la teoría de la familiaridad es más plausible cuando se aplica a estados fenoménicos que cuando se extiende a actitudes proposicionales.⁴³ Esto se debe, creo, a una ambigüedad en la expresión anglosajona para el conocimiento (*‘to know’*), que no discrimina entre conocimiento proposicional (saber que *p*) y conocimiento de objetos (conocer *O*). El contacto parece jugar un papel importante en el conocimiento de objetos;⁴⁴ su rol no es tan claro en el caso del saber proposicional. De modo que, en realidad, la teoría del *contacto directo* parece dejar en el misterio las fuentes del conocimiento de la propia mente que aquí hemos venido discutiendo. Veamos si el segundo grupo de teorías explicativas corre con mejor suerte.

1.3.3. La teoría del sentido interno

La teoría del sentido interno o introspección, a diferencia de la teoría de la familiaridad, enfatiza la relación entre el conocimiento de la propia mente y el conocimiento que, a través de los sentidos, tenemos del mundo.⁴⁵ De este modo, el introspeccionista concibe al conocimiento de la propia mente como un tipo de percepción de orden superior: se trata de una suerte de mecanismo que recibe como entradas estados mentales y produce atribuciones de estados mentales como

⁴³ Gertler (2001) expresa dudas acerca de que la teoría de la familiaridad pueda ser extendida exitosamente a estados mentales intencionales (*i.e.*, actitudes proposicionales).

⁴⁴ En cuyo caso, la teoría de la familiaridad parecería tener repercusiones directas e indeseables para la ontología de lo mental: objetos internos privados (*e.g.*, datos sensoriales, *qualia*, etcétera).

⁴⁵ Presumiblemente, Locke defendió esta tesis: “Esta fuente [...] de ideas la tiene todo hombre en sí mismo, y aunque no es un sentido, ya que no tiene nada que ver con objetos externos, con todo se parece mucho y puede llamársele con propiedad sentido interno” (1690: Book II, Chap. I, § 4). Asimismo, parece haberle otorgado a la asimetría epistémica el estatuto de omnisciencia: “...es imposible que alguien perciba sin percibir que percibe. Cuando vemos, oímos, olemos, gustamos, sentimos, meditamos o deseamos algo, es que sabemos que hacemos cualquiera de esas cosas” (Locke, 1690: Book II, Chap. XXVII, § 9).

salidas. La teoría del sentido interno parece atractiva porque permite generalizar la condición de justificación, aquello que hace a la creencia verdadera conocimiento, de una manera que involucre los mismos recursos en el caso del conocimiento del mundo, de las otras mentes y de la propia mente. En sus versiones contemporáneas esta posición ha resultado especialmente atractiva para los defensores de posiciones fiabilistas en torno al conocimiento.⁴⁶ Éstas lo representan como el producto de un proceso que generalmente da lugar a creencias verdaderas (*i.e.*, un proceso que probablemente debido a razones evolutivas genera creencias sobre los propios estados psicológicos de manera fiable). De este modo, lo que explicaría la autoridad especial de las atribuciones de estados mentales en primera persona sería la fiabilidad del proceso *causal* mediante el cual se vinculan dichas atribuciones con lo que las hace verdaderas (los estados mentales en cuestión).

Ya habíamos señalado una razón (en la sección 1.1.1) para guardar ciertas reservas hacia enfoques como éste.⁴⁷ La razón es que considerar a la mera fiabilidad como lo que distingue al conocimiento de la mera creencia verdadera parece

⁴⁶ El atractivo principal de las posiciones fiabilistas está vinculado al externalismo epistémico, que es una buena apuesta teórica debido a que el internalismo parece conducir de manera inevitable a la conclusión escéptica (véase el apartado 3.3.3). Aunque, por supuesto, esta polémica no está cerca de su resolución, como lo ilustra el siguiente pasaje de Robert Stalnaker: “Internalistas y externalistas reclamarán que el otro está dando por sentado lo que requiere ser explicado. Los internalistas ven al proyecto externalista como un proyecto motivado por el pesimismo. Su reclamación es ésta: «Puesto que no ves esperanza en razonar hacia afuera de tu mundo interno, simplemente renuncias y *asumes* que hay un mundo que responde a tu concepción interna. Te proporcionas a ti mismo algún material adicional, dándolo por sentado ya que no encuentras otra manera de hacer progresos. Decides que el trabajo honesto es tan mal pagado que robar es la única opción». Pero los externalistas rechazan esta manera de entender su proyecto. «No es», insisten, «que estemos dando por sentado lo que tú consideras dado y más aún. Eres tú, pensamos, quien está dando por sentados fenómenos que requieren una explicación. Desde nuestra perspectiva, podemos darle sentido a tu punto de partida el mundo interno sólo ubicándolo en un mundo más amplio. El problema, pensamos, no es que no se pueda responder al escepticismo, desde un punto de vista puramente interno, aunque es posible que no se pueda responder. (De hecho, argumentamos que el problema del escepticismo, visto de esta manera, es peor de lo que tú crees). El problema es más bien que el escepticismo acerca del mundo externo tiene una de sus fuentes en una aceptación acrítica, y una falsa concepción, de nuestro conocimiento del mundo interno» (2008: 4). Además, las posiciones fiabilistas parecen casar muy bien con el fisicalismo, lo cual causaba cierta inquietud con respecto a la teoría de la familiaridad (véase la nota 44).

⁴⁷ Davidson (1983: 143) hace una advertencia similar: “...una explicación causal de una creencia no muestra cómo o por qué la creencia está justificada”.

cercenar el estrecho vínculo que intuitivamente encontramos entre el conocimiento y la acción racional.

No obstante, hay otras razones, que pueden generalizarse incluso a formulaciones no fiabilistas de la teoría del sentido interno, para mostrarse renuente ante esta explicación. Una de ellas, expuesta por Shoemaker (1994) es que, si hemos de seguir la analogía que esta teoría establece entre percepción e introspección, deberíamos admitir la posibilidad de ‘ceguera respecto a la propia mente’ (*self-blindness*) o carencia de sentido interno. Ello implicaría la posibilidad de que hubiese sujetos racionales que fuesen especialmente poco confiables al atribuirse estados mentales o incluso que fuesen incapaces de hacerlo (tal como hay invidentes, cuya confiabilidad al hacer juicios visuales por obvias razones es significativamente inferior a la nuestra). Un sujeto como éste resultaría muy peculiar; haría, *e.g.*, frecuentes aserciones de enunciados problemáticos como la paradoja de Moore: “Es el caso que p , pero no creo que sea el caso que p ”.

Otra razón es que en la teoría del sentido interno el estatuto epistémico de las atribuciones de primera persona parece, de manera importante, ser una cuestión accidental. *E.g.*, si la evolución no hubiese propiciado el desarrollo de un sentido interno con mayor fiabilidad que nuestros órganos de percepción externa, nuestro conocimiento de los estados mentales de otras personas podría ser más seguro que el de la propia mente. Ésta sería una situación extravagante. Intuitivamente, no parece que tal situación sea siquiera conceptualmente posible. El defensor de la teoría del sentido interno podría replicar que nuestras intuiciones conceptuales son a menudo incorrectas. Podríamos concederle esto, aunque, a estas alturas uno podría sospechar que adoptar su punto de vista exige una revisión demasiado extensa de nuestras prácticas de atribución de estados mentales.⁴⁸

⁴⁸ Lo cual no parece distar de lo que propondría un materialista eliminativo. Esto no sugiere que la teoría del sentido interno sea más plausible a la luz del eliminativismo, pues éste rechazaría tal teoría en la formulación arriba discutida por preservar el vocabulario de la psicología popular.

Hemos examinado dos influyentes reconstrucciones de lo que se ha denominado la concepción cartesiana del conocimiento de la propia mente; ambas tienen una raíz común, *i.e.*, la idea de que lo que explica la autoridad de la primera persona involucra cierta clase de asimetría epistémica, entendida ésta en términos de acceso privilegiado. Se han indicado razones para considerar que ambas explicaciones fallan en su propósito o son, en el mejor de los casos, deficientes. Esto proporciona una razón para abandonar la condición de acceso epistémico privilegiado. A continuación, revisaremos explicaciones de lo que hace especiales a las atribuciones de estados mentales en primera persona que no consideran que éstas involucren una condición robusta de acceso privilegiado, o incluso ninguna clase de factores epistémicos, y evaluaremos sus méritos. En la sección 1.5 se argumentará que abandonar la condición de acceso epistémico privilegiado no requiere renunciar *tout court* a la suposición de que las adscripciones de estados mentales en primera persona involucran conocimiento, y que es razonable conservar esta suposición.

1.4. *Las fuentes de la autoridad de la primera persona II: explicaciones no epistémicas*

La agenda filosófica del siglo XX fue, en gran medida, anti-cartesiana.⁴⁹ Por ello, no resulta sorprendente que varias de las figuras intelectuales más sobresalientes

⁴⁹ Robert Stalnaker resume de manera excepcional esta reacción: “La imagen cartesiana de la mente, y del mundo, fue atacada desde diversas direcciones a lo largo de la mayor parte del siglo pasado. Se nos enseñó a prescindir de objetos privados, y de lenguajes privados, del mito de lo dado, del fantasma en la máquina, del teatro cartesiano, de cosas presentes a la mente. Naturalizamos nuestra epistemología: en lugar de intentar construir un fundamento a partir de los materiales que encontramos en nuestros mundos internos, se nos recomendó comenzar en medio de las cosas, observar cómo las personas de hecho emprendían la justificación de sus creencias, y explicar su conocimiento en términos de la manera en que interactúan con las cosas en el mundo que nosotros, como teóricos, encontramos ahí. Pero la bestia cartesiana es una criatura con cabeza de hidra que se resiste a ser destazada, y continúa coloreando nuestras imágenes y proyectos filosóficos. Wittgenstein, Ryle, Quine, Sellars, Davidson [...] podrán haber cortado unas cuantas cabezas cartesianas, pero siguieron volviendo a crecer” (Stalnaker, 2008: 1).

del siglo pasado defendieran una concepción diametralmente opuesta a lo que en la sección precedente se ha denominado la ‘concepción cartesiana’ del conocimiento de la propia mente. Este tipo de reacción anti-cartesiana fue alentada principalmente por la obra filosófica de Gilbert Ryle y del segundo Wittgenstein, aunque prolifera incluso en sofisticadas variantes en nuestros días.⁵⁰ Pese a que esta reacción ha tenido una gran diversidad de manifestaciones, aquí nos concentraremos únicamente en aquellas que sean relevantes para la autoridad de la primera persona.

Llamaremos al cúmulo de explicaciones que comparten los rasgos de esta reacción al cartesianismo la ‘concepción no epistémica’ de la autoridad de la primera persona. Quienes defienden este punto de vista consideran, en términos generales, que no sólo carecemos de un acceso epistémico privilegiado al contenido de nuestra propia mente, sino que la autoridad que caracteriza a nuestras autoadscripciones de actitudes proposicionales no es razón para considerarlas como un logro cognitivo significativamente distinto al conocimiento de otras mentes, o incluso podría no constituir un logro epistémico en absoluto. De modo que la concepción no epistémica sugiere que la autoridad de la primera persona o bien no es un fenómeno epistémico singular o bien ni siquiera es realmente epistémica, sino que sólo parece serlo.

⁵⁰ Con respecto al conocimiento de la propia mente, Sellars (1956) y Rorty (1970) han defendido posiciones similares a la de Ryle (1949), mientras que Finkelstein (2003 y 2010), Bar-On (2004) y Wright (1998) defienden varias versiones contemporáneas de la posición de Wittgenstein (1953). Aunque con onerosas deudas teóricas con estos próceres de la filosofía del siglo XX, Moran (2001), Bilgrami (2006), Gallois (1996), Burge (1988 y 1996) y Shoemaker (1994) sostienen posiciones más originales. Señalan que lo peculiar de nuestras atribuciones de estados mentales en primera persona es que cada uno de nosotros se encuentra en una posición única para evaluar y revisar sus estados mentales (en ocasiones, dirigiendo su mirada hacia el mundo), por lo que tenemos cierta responsabilidad de ponerlas en consonancia con nuestra evidencia y con restricciones racionales. No obstante, se ha objetado que este tipo de enfoques no son teorías en competencia acerca del conocimiento de la propia mente, sino que podrían considerarse más bien como explicaciones de otro fenómeno: el cambio racional de actitudes proposicionales (véase Gertler, 2011: 180-199).

1.4.1. El conductismo lógico

Comenzaremos con el punto de vista de Ryle y posiciones afines a la suya. A esta concepción frecuentemente se le denomina ‘conductismo lógico’. Ryle señala que la condición de acceso epistémico privilegiado con que se caracteriza al conocimiento de la propia mente es un corolario de la concepción cartesiana de lo mental:

Una contraparte natural de la teoría de que las mentes constituyen un mundo distinto del «mundo físico» es la teoría de que existen maneras de descubrir los contenidos de este otro mundo que son contrapartes de nuestras maneras de descubrir los contenidos del mundo físico. Mediante la percepción sensorial averiguamos qué existe y sucede en el espacio; de modo que lo que existe y sucede en la mente debe también ser averiguado mediante la percepción, pero una percepción de un tipo distinto y refinado, uno que no requiere del funcionamiento de los groseros órganos corporales (Ryle, 1949: 154).

En consecuencia, Ryle sugirió que el abandono del dualismo cartesiano implicaría también el rechazo de la concepción cartesiana del conocimiento de la propia mente. En sustitución de la imagen dualista, Ryle propuso que consideráramos a lo mental como un constructo lógico a partir de conductas. De este modo, nuestro vocabulario mental podría ser redefinido como (y eliminado a favor de) un vocabulario ontológicamente más inocuo sobre conductas y disposiciones conductuales. En este marco teórico, el supuesto acceso privilegiado a los propios estados mentales no involucraría realmente una diferencia en el contenido ni en la forma de conocimiento entre lo que ocurre al conocer la propia mente y lo que sucede cuando sabemos acerca de otras mentes: “El tipo de cosas que puedo descubrir acerca de mí mismo son el mismo tipo de cosas que puedo descubrir acerca de otras personas, y los métodos para descubrirlas son más o menos los mismos” (Ryle, 1949: 155).

De este modo, Ryle rechaza, en términos generales, la asimetría epistémica. Cree que al conocer nuestros pensamientos debemos también apelar a evidencia

conductual y realizar inferencias a partir de ella. Sin embargo, considera que del conductismo lógico se desprende una teoría explicativa de la autoridad de la primera persona. Aunque conocemos nuestros estados mentales exactamente en la misma forma en que conocemos los ajenos (por medio de la observación de lo que se dice y se hace), nuestras autoatribuciones de estados mentales tienen una presuposición de verdad debido a que cada uno de nosotros observa continuamente su propia conducta, desde una perspectiva más próxima (por así decirlo). Por ello, resulta natural que nuestro conocimiento acerca de los propios estados mentales se considere más acertado y parezca más confiable.⁵¹ Sin embargo, la diferencia entre el conocimiento de la propia mente y el de otras mentes sería de grado, no de naturaleza.

No resulta complicado encontrar contraejemplos a la explicación de Ryle. En muchas ocasiones podemos dar cuenta de lo que pensamos, sentimos y queremos sin recurrir a la observación de nuestra conducta; e incluso en estos casos se preserva la autoridad de la primera persona. Por contraparte, parece poco razonable afirmar que un verdugo conoce mejor los pensamientos de sus víctimas que éstas, simplemente porque se encuentra en una posición más ventajosa para observar su conducta. La explicación ryleana de la autoridad de la primera persona fallaría en estos casos. No obstante, hay una objeción más profunda a la teoría explicativa propuesta por Ryle, que sugiere el abandono del marco teórico con el que trabaja el conductismo lógico:

¿Por qué estamos dispuestos (como asumo que lo estamos) a abandonar el intento de dar definiciones explícitas de los conceptos mentales en términos de los con-

⁵¹ La variante de Rorty (1970) en muchos sentidos similar a la de Sellars (1956) señala que aunque originalmente las adscripciones de estados mentales en primera persona se efectuaban con base en la observación, después se descubrió que las personas podían adscribirse estados mentales sin usar evidencia conductual u observacional, dando mejores explicaciones de la conducta. Esto, afirma Rorty, instauró una convención lingüística de tratar a las adscripciones en primera persona de forma privilegiada. Como indica Davidson (1984: 8) esto no aclara el hecho de que las adscripciones en primera persona, no basadas en la evidencia, son unívocas con las adscripciones en tercera persona (conciernen a los mismos estados y sucesos, aunque basadas en observación y evidencia).

ductuales? Seguramente no sólo porque todos los intentos actuales son conspicuamente inadecuados. Más bien debido a que estamos persuadidos, como lo estamos en el caso de muchas otras formas de reduccionismo definicional (el naturalismo en la ética, el instrumentalismo y el operacionalismo en las ciencias, la teoría causal del significado, el fenomenalismo, y así sucesivamente el catálogo de las derrotas de la filosofía), de que hay sistema en los fracasos. Supóngase que intentamos decir, sin usar ningún concepto mental, qué es para un hombre creer que hay vida en Marte. Una línea que podríamos tomar es ésta: cuando se produce cierto sonido en presencia del hombre (‘¿Hay vida en Marte?’) él produce otro (‘Sí’). Pero, por supuesto, esto muestra que él cree que hay vida en Marte sólo si comprende el español, su producción del sonido fue intencional y fue una respuesta a los sonidos como significando algo en español; y así sucesivamente. Para cada deficiencia descubierta, añadimos una nueva salvedad. Pero sin importar cómo parchemos y ajustemos las condiciones no mentales, siempre encontraremos la necesidad de una condición adicional (siempre y cuando él *note, comprenda, etc.*) que es mental en su carácter (Davidson, 1970: 216-217).

El proyecto filosófico del conductista lógico fracasa porque no es posible analizar un estado mental particular preservando su valor para la explicación de la acción empleando únicamente términos conductuales, sin reintroducir vocabulario mental. Las acciones de un individuo, o sus disposiciones a la conducta, no dependen del hecho de que tenga una creencia particular, sino también de otros estados mentales como deseos y otras creencias. Una creencia particular puede aparecer en cualquier clase de conducta, dependiendo de otras actitudes proposicionales del individuo. Además, incluso si pudiéramos enlistar exhaustivamente las disposiciones de un individuo o sistema a comportarse de diversas maneras en varias circunstancias, nada en esta lista nos mostraría que lo que estamos proporcionando es un análisis de sus creencias, pues para que algo cuente como una creencia (u otro estado mental), debe ser capaz de combinarse apropiadamente con otros estados mentales para explicar racionalmente la acción (véase Davies, 1998: 259).

1.4.2. El neoexpresivismo

Aunque se ha asociado a Wittgenstein con el conductismo lógico, parece más razonable adjudicarle una posición más refinada, que en sus versiones actuales se denomina ‘neoexpresivismo’.⁵² De acuerdo con esta posición, la formulación lingüística de nuestros reportes sobre estados mentales produce la ilusión de la autoridad epistémica de la primera persona. No obstante, para el neoexpresivista, el carácter epistémico de la autoridad es espurio. El enclave de este punto de vista se sitúa en conjunto de tesis acerca del significado lingüístico (“Toda una nube de filosofía se condensa en una gotita de gramática” Wittgenstein, 1953: II, p. 507), que reaparece en una serie de observaciones dispersas en la obra de Wittgenstein:

Puedo saber lo que otro piensa, no lo que yo pienso.

Es correcto decir «Sé lo que piensas», y erróneo: «Sé lo que pienso» (Wittgenstein, 1953: II, p. 507)

El neoexpresivista observa que, aunque es perfectamente natural indagar cuáles son las bases sobre las que uno infiere estados mentales de otros, no es así en el caso de los propios estados mentales. Al parecer hay un uso anormal de expresiones de saber y de duda en lo que concierne a los reportes de actitudes proposicionales cuando se emplean para referirse a estados mentales en primera persona. De acuerdo con Wittgenstein, la duda y el conocimiento se inscriben en una forma específica de usar el lenguaje y el hecho de que no nos parezca razonable dudar sobre nuestros pensamientos, sentimientos y creencias no indica que los conozcamos de manera infalible, transparente, sino que no es apropiado decir que los conocemos en absoluto (lo que no significa, por otra parte, que los ignoremos).

Esto no indica que, literalmente, nuestro uso del lenguaje en estos casos sea incorrecto y deba ser reestructurado, sino que encapsula una confusión que ha

⁵² Ésta es la postura que le atribuyen, *e.g.*, Finkelstein (2003 y 2010) y Bar-On (2004). Aunque es conveniente señalar ciertas reservas hacia esta lectura — como advierte Wright (1998) — la trataré (con acotaciones similares a las indicadas con respecto a Descartes) como la concepción wittgensteiniana.

de ser disipada: la suposición de que los reportes de actitudes proposicionales desempeñan funciones lingüísticas similares en la primera y la tercera personas. Esta suposición es alentada por las similitudes sintácticas (a nivel de la gramática superficial) entre ambos grupos de reportes; pero el neoexpresivista sugiere que, como en muchos otros casos, estas similitudes encubren una diferencia semántica. Mientras las adscripciones de estados mentales en tercera persona pueden ser descriptivas (y, por ende, tener condiciones de verdad), lo que el neoexpresivista propone es que las que se formulan en primera persona desempeñan una función expresiva: son manifestaciones [*avowals*] y no descripciones de nuestra vida mental. Una forma burda de ilustrar el punto es la siguiente: si digo “*S* siente dolor de estómago”, lo que intento es describir, sobre la base de evidencia conductual, adecuadamente una situación; en cambio, si *S* profiere “Me duele el estómago”, lo que hace es manifestar lingüísticamente su dolor (lo que podría haber realizado, de manera más natural, profiriendo “¡Auch!” mientras se frota el estómago).

Wittgenstein presenta esquemáticamente la intuición que guía al neoexpresivista en el siguiente pasaje de *Zettel*:

Plan para el tratamiento de los conceptos psicológicos.

Los verbos psicológicos se caracterizan por el hecho de que la tercera persona del presente tiene que verificarse mediante observación, en tanto la primera no.

Frases en la tercera persona del presente: información. En primera persona del presente: expresión. ((No del todo cierto.))

La primera persona del presente es aún a una expresión (Wittgenstein, 1967: § 472).

De modo que, de acuerdo al neoexpresivismo, la manera de seguir adelante consiste en negar que tengamos *conocimiento* (saber proposicional) de un paisaje interno de la clase sugerida por la concepción cartesiana. No podríamos tenerlo porque una de las condiciones del conocimiento ni siquiera es atinente a los repor-

tes de actitudes proposicionales en primera persona: no pueden ser verdaderos o falsos, sino manifestaciones auténticas o fingidas.

Sin embargo, hay varias razones para no estar satisfechos con el neoexpresivismo. Una de ellas es que amenaza la presunta univocidad del vocabulario psicológico; no se limita a indicar una ambigüedad más en nuestro uso de la palabra ‘saber’, sino que asume que nuestras atribuciones de estados mentales varían su función semántica cuando se emplean en tercera y en primera persona (sugiriendo que, *e.g.*, en este último caso ni siquiera tendrían condiciones de verdad, al tratarse de manifestaciones). Esto, como ha señalado Sosa (2003: 274), es vulnerable a la objeción de que el lenguaje de las supuestas manifestaciones parece semánticamente mucho más rico que cualquier vocabulario puramente expresivo, como se manifiesta en el hecho de que puede aparecer en condicionales y otras expresiones que pueden caracterizarse como funciones veritativas.

Por otra parte, no parece claro que pueda ofrecerse un tratamiento de las actitudes proposicionales en términos de manifestaciones similar al caso del dolor. Pues, ¿cuál sería la manifestación natural de un estado mental con un contenido lingüísticamente articulado (como, *e.g.*, “Creo que hay vida en Marte”)? Guttenplan (1996: 57) ha indicado que la concepción no epistémica “...no es una que sea fácil mantener enfocada. Tiene la tendencia ya sea de transformarse ante la vista en una clase de conductismo o bien de desaparecer totalmente”. Esto se debe a que si tratamos las autoadcripciones de estados mentales como meras manifestaciones, entonces o bien asumimos que pueden caracterizarse en términos puramente conductuales o bien simplemente ignoramos el fenómeno que nos proponíamos explicar (la presunción de que el hablante está, generalmente, en lo correcto al adscribirse estados mentales).

Hemos examinado dos enfoques no epistémicos que intentan ofrecer una teoría explicativa de las peculiaridades involucradas en las atribuciones de estados mentales en primera persona. No obstante, han resultado insatisfactorios en la

medida en que no dan cuenta apropiadamente de los fenómenos que nos proponíamos explicar: la autoridad de la primera persona y la asimetría epistémica. Estos enfoques rechazan la aproximación cartesiana al problema de la asimetría epistémica, entendida como condición de acceso privilegiado. Como vimos en la sección 1.3, hay buenas razones para tomar distancia en este punto con respecto a la concepción cartesiana. Sin embargo, el neoexpresivismo se aparta aún más al negar que las autoadscripciones de estados mentales puedan tener un estatuto epistémico; esto sería razonable si le otorgáramos a la autoridad de la primera persona el estatuto de la infalibilidad, pues parece que el conocimiento involucra la posibilidad del error y la duda. Pero hemos visto que puede tratarse de una condición más modesta, por lo que no tenemos, en este punto, una buena razón para afirmar sin más que las atribuciones de pensamientos en primera persona están excluidas de nuestro dominio cognoscitivo.

En lo que sigue examinaremos la propuesta de Donald Davidson, que promete ofrecer una explicación sustantiva del conocimiento de la propia mente sin ser objeto del tipo de críticas que se han presentado a las posiciones hasta aquí examinadas.

1.5. Donald Davidson: la autoridad de la primera persona como condición de la interpretación

Donald Davidson examinó el conocimiento de la propia mente y sus intrigantes características en numerosos ensayos,⁵³ y fue concediéndole un papel cada vez más importante en su proyecto filosófico. En este ámbito, Davidson se propone explicar qué es lo que hace peculiar al conocimiento de la propia mente frente a otras formas de conocimiento y cómo se relaciona con ellas. En esta sección nos ocuparemos de presentar la teoría explicativa de Davidson del conocimiento de la

⁵³ Entre los más importantes habría que contar sus (1984), (1987), (1988), (1989), (1990), (1991) y (1998).

propia mente; su situación al interior de la empresa epistemológica se examina en el capítulo 3. Para captar con claridad la magnitud de la tarea de ofrecer una teoría explicativa es conveniente resumir las condiciones que, hasta este punto, consideramos que debe satisfacer.

Señalamos en la sección 1.1 que lo que se pretende explicar guarda, intuitivamente, relaciones con el conocimiento; *i.e.*, pese a sus peculiaridades, se encuentra estrechamente relacionado con nuestros logros cognitivos respecto al mundo y a las otras mentes. Mínimamente, involucra creencia, verdad objetiva y un vínculo, dependiente de otras creencias del individuo, que conecte apropiadamente las condiciones anteriores. En la sección 1.2 indicamos que lo que aquí denominamos ‘conocimiento de la propia mente’ concierne al conocimiento, por parte de un sujeto, de sus propios estados mentales en el momento actual, limitados éstos a actitudes proposicionales. Indicamos también que esta forma de conocimiento exhibe dos rasgos especiales: la autoridad de la primera persona, entendida como una presunción de verdad que acompaña a las autoadscripciones de estados mentales, y la asimetría epistémica, concebida como el carácter directo o no mediado por observación e inferencia a partir de la conducta de nuestras atribuciones de pensamientos en primera persona. En la sección 1.3 se ofrecieron razones para rechazar explicaciones de estos rasgos que apelan a una clase de acceso epistémico privilegiado. Por una parte, veíamos que la teoría de la familiaridad no explica cómo la presencia de ciertos estados mentales puede fungir como justificación, en el sentido requerido por el saber proposicional, de las atribuciones en primera persona. Por otra parte, la teoría del sentido interno no explica adecuadamente por qué la autoridad está tan estrechamente ligada a la primera persona. En la sección 1.4 vimos que, pese a que puede resultar atractivo *prima facie*, rechazar la condición de acceso epistémico privilegiado no debería considerarse equivalente a negar los rasgos epistémicos peculiares del conocimiento de la propia mente. En especial, no es conveniente soslayar la condición de asimetría epistémica ape-

lando a una diferencia de grado o cantidad de evidencia, como sugería el conductismo lógico. Asimismo, consideramos que es deseable que una explicación de la autoridad de la primera persona preserve la univocidad del vocabulario psicológico, sin ser negligente a sus aspectos epistémicos, lo que nos pareció un defecto en el neoexpresivismo.

Como puede apreciarse, la tarea de ofrecer una teoría explicativa del conocimiento de la propia mente no es sencilla; no obstante, en este apartado me propongo argumentar que, tal como ha sido caracterizado hasta este punto, se trata de un reto que la explicación de Davidson enfrenta de manera satisfactoria. En los capítulos siguientes trataré de mostrar que esta tarea es muchísimo más compleja de lo que hasta ahora pudiera parecer, aunque defenderé que la explicación davidsoniana continúa siendo prometedora.

1.5.1. La autoridad de la primera persona en el lenguaje

Davidson comienza por señalar que el problema de ofrecer una teoría explicativa del conocimiento de la propia mente puede situarse dos niveles: uno epistémico y otro concerniente al lenguaje. En el primer caso, la teoría debe explicar qué respalda la creencia de una persona de que tiene cierto estado mental en el momento presente, de una manera que no está disponible, en principio, a un tercero; en el segundo caso, la teoría debe decirnos por qué una presunción de verdad, ausente en las adscripciones en de estados mentales en tercera persona, acompaña las preferencias de oraciones sobre actitudes proposicionales que un hablante hace en primera persona. Podemos ver esto con más claridad si empleamos la distinción, presentada al inicio de este capítulo, entre atribuciones y adscripciones de estados mentales.⁵⁴ El problema epistémico concierne a la justificación de las atribuciones

⁵⁴ Aunque en la discusión precedente esta distinción no haya jugado un papel muy relevante, a partir de ahora comenzará a desempeñar un rol mucho más serio.

de estados mentales en primera persona (*i.e.*, se centra en la asimetría epistémica); el lingüístico, por su parte, refiere a la presunción de verdad, por parte de un oyente, sobre las autoadscripciones de actitudes proposicionales que hace un hablante (*i.e.*, se enfoca en la autoridad de la primera persona). Davidson sugiere que concentrarse en el segundo problema puede ayudar a echar luz sobre el primero:

El punto puede presentarse, y la pregunta formularse, ya sea en la modalidad del lenguaje o de la epistemología. Pues si uno puede hablar con una autoridad especial, el estatus del propio conocimiento debe de algún modo concordar; mientras que si el propio conocimiento muestra alguna diferencia sistemática, las afirmaciones de conocimiento deben reflejar la diferencia. Asumo, por tanto, que si la autoridad de la primera persona en el habla puede ser explicada, habremos hecho mucho de, si no es que todo, lo que debe hacerse para caracterizar y dar cuenta de los hechos epistemológicos (Davidson, 1984: 3).

De este modo, si partimos de la autoridad de la primera persona, podemos comenzar reformulando el problema de la asimetría epistémica en términos de comunicación lingüística. “La pregunta se convierte entonces en la siguiente: ¿qué es lo que explica la diferencia entre la clase de convicción que usted tiene de que estoy en lo cierto cuando [...me *adscribo* un estado mental...] y la clase de convicción que yo mismo tengo?” (Davidson, 1984: 11) Podría considerarse que esta estrategia constituye un *détour* del problema original si no tenemos claro qué aspectos del fenómeno deseamos explicar (esto es lo que se sugiere, *e.g.*, en Bar-On, 2004: 174-178). Pero mientras haya claridad sobre los objetivos de una teoría explicativa no debería asumirse, sin más, que esta estrategia constituya un cambio de tema. Esta variación en el énfasis tiene sus ventajas, pues, al tratar el problema en su versión puramente epistémica, interfieren subrepticamente suposiciones acerca del problema lingüístico:

En la medida en que planteamos el problema en términos de la clase de garantía o autoridad que alguien tiene con respecto a afirmaciones acerca de la actitud de un agente hacia una proposición (u oración con una interpretación particular), parece

que estamos restringidos a dar cuenta de las diferencias al postular simplemente diferentes clases o fuentes de información. Alternativamente, podemos postular diferentes criterios de aplicación de los conceptos o palabras clave (‘cree que’, ‘se propone’, ‘desea que’, etc.). Pero estos movimientos no hacen sino replantear el problema, como hemos visto, y de esta manera invitan al escepticismo acerca del conocimiento de las mentes de otros (o de nuestra propia mente). Pero si planteamos el problema en términos de relaciones entre agentes y preferencias, podemos evitar el punto muerto (Davidson, 1984: 11).

Lo que tácitamente permea la discusión, si nos centramos en la cuestión epistemológica, es la suposición de que tanto el hablante como el oyente se encuentran en la misma situación en lo que concierne a *parte de la evidencia relevante* para la atribución de actitudes proposicionales: el significado de las preferencias (autoadscripciones de estados mentales) del hablante. De la relevancia de esta evidencia da testimonio el hecho de que la individuación de actitudes proposicionales requiere de la estructura lingüística fina incrustada en las cláusulas subordinadas. Que una atribución de actitudes proposicionales (creencias, deseos, intenciones, etc.) sea correcta depende, en parte, de que el sujeto de la atribución pueda reconocer la oración subordinada como expresando el contenido de su pensamiento. Pero un hablante puede no reconocer como el contenido de su pensamiento una oración que preserva, desde la perspectiva del oyente, el significado de la oración que el hablante profirió como la cláusula subordinada de una autoadscripción de cierta actitud proposicional.⁵⁵ Quizá un ejemplo illustre este punto:

⁵⁵ Esto es una forma (tortuosa, quizá) de indicar que las actitudes proposicionales generan contextos intensionales o de opacidad semántica: “Una manera de reconocer que estamos atribuyendo una actitud proposicional es notar que las oraciones que usamos para hacer la atribución pueden cambiar de verdadera a falsa si, en las palabras que seleccionan el objeto de la actitud, sustituimos alguna expresión referencial por otra expresión que se refiere a la misma cosa. [...] Si usáramos palabras como ‘creer’, ‘pensar’ y ‘proponerse’, prescindiendo del rasgo de la opacidad semántica, habría una cuestión acerca de si estamos usando esas palabras para atribuir actitudes proposicionales. Pues desde hace mucho se ha reconocido que la opacidad semántica distingue al habla acerca de actitudes proposicionales del habla acerca de otras cosas” (Davidson, 1982: 97-98).

H y *O*, dos notarios públicos, caminan juntos por la acera. *O* le comenta a *H* que ha estado teniendo continuas jaquecas y que tiene dificultades para leer, en especial las letras pequeñas de los contratos. *H* manifiesta sus simpatías, mientras *O* le pregunta: “¿Cuál crees que sea el problema?” *H* responde: “Creo que *necesitas visitar a un oculista*. Veamos si hay uno por aquí”. Unos minutos más tarde, *O* se detiene en seco, sabiendo que su búsqueda ha terminado. Frente a ellos se abren las puertas de un consultorio con la inscripción rotulada: «OFTALMÓLOGO». Le hace una seña a *H*, pero éste le responde con un gesto de perplejidad: “Lo que necesitas es un doctor de ojos”, comenta, “no uno de oídos, nariz y garganta”. *O* suelta una carcajada: “Pero, amigo”, responde, “un oftalmólogo es un doctor de ojos. El de oídos, nariz y garganta se llama otorrinolaringólogo”. Cabizbajo, *H* confiesa: “No lo sabía”.

Desatendamos el problema clínico de *O* y las deficiencias educativas de *H*; centrémonos en el intercambio lingüístico. *O* asume que la siguiente atribución es correcta: *H cree que necesito visitar a un oftalmólogo*. Después de todo, *H* profirió la oración “Creo que necesitas visitar a un oculista” y, como *O* sabe, los oculistas se hacen llamar ‘oftalmólogos’. No obstante, como descubre más tarde, su atribución es errónea: *H* no cree que *O* deba visitar a un oftalmólogo, pues (erróneamente) *H* cree que los oftalmólogos se ocupan del cuidado de oídos, nariz y garganta. Si *O* hubiese preguntado: “¿Crees que necesito visitar a un oftalmólogo?”, *H* habría respondido en forma negativa, pues aunque comparta propiedades semánticas con la cláusula subordinada de su adscripción original no reconocería a la oración “necesitas visitar a un oftalmólogo” como expresando el contenido de su creencia. Esto se refleja en la conducta de *H*, quien no cree que hayan encontrado a un oculista cuando se detienen frente al consultorio.

Puesto que la conducta lingüística del hablante y sus respuestas al entorno son toda la evidencia con la que cuenta el oyente para atribuirle pensamientos al hablante, si asumimos que ambos comprenden igualmente el significado de las preferencias de éste, la pregunta que nos resta para dar cuenta de la autoridad de la primera persona es: ¿con qué evidencia adicional cuenta el hablante? Pero esto

apunta a trazar la diferencia entre el conocimiento de la propia mente y el de otras mentes en términos de un acceso epistémico privilegiado. Y no queremos hacer esto.

1.5.2. La diferencia epistémica: interpretación radical

Si, como sugiere Davidson, centramos nuestra atención en el problema lingüístico, veremos con más claridad que hay una diferencia importante entre la situación del hablante y la del oyente. Este último no sólo tiene que descifrar cuáles son los pensamientos del hablante a partir de su conducta lingüística, sino que debe investigar también lo que significan sus palabras:

Surge aquí un reto básico. Un hablante tiene a una oración por verdadera como resultado de dos consideraciones: lo que él asume que la oración significa, y lo que él cree que es el caso. El problema es que lo que, de manera relativa, es directamente observable para un intérprete es el producto de dos actitudes inobservables, creencia y significado. ¿Cómo pueden distinguirse los papeles de estos dos factores explicativos y extraerse a partir de la evidencia? (Davidson, 1990a: 61)

La clave para ver la diferencia entre la situación del hablante y la del oyente reside en considerar al habla como un caso de acción intencional, que involucra la producción de preferencias con el propósito de establecer la comunicación. Puesto que comprender las preferencias lingüísticas de un hablante involucra interpretar sus acciones, el oyente no puede simplemente asumir que sabe qué significan dichas preferencias (incluso cuando ha descubierto que son sinceras), a menos que esté en condiciones de especificar qué estados mentales expresan; y, a la inversa, no puede asumir que está en condiciones de especificar los estados mentales expresados por las preferencias del hablante, a menos que sepa lo que tales preferencias significan. El oyente se encuentra comprometido con una inferencia que involucra *interpre-*

tación radical: debe intentar comprender al hablante sin conocer de antemano cuáles son sus creencias ni qué significan sus palabras (véase Davidson, 1973a).

El oyente observa las interacciones causales entre el hablante y su entorno para desarrollar hipótesis empíricas sobre el contenido de sus creencias y el significado de sus preferencias. Sin conocimiento previo del significado de las palabras del hablante y del contenido de sus creencias, para que la comunicación tenga lugar el oyente debe asumir, adicionalmente, que el hablante es *interpretable*, *i.e.*, que usa "...un repertorio finito de sonidos distinguibles aplicados consistentemente a objetos y situaciones que él cree que son evidentes para su oyente" (Davidson, 1984: 13). Si el oyente no presupone esta clase de sistematicidad en las preferencias del hablante, entonces, incluso con la observación de la conducta del hablante y de sus relaciones causales con el entorno, no tiene nada que interpretar (véase Davidson, 1987: 38). Descubrir cómo se relacionan las preferencias del hablante con el entorno que ambos comparten cómo determinar sus condiciones de verdad, es parte importante de la tarea del intérprete si quiere comprender a su interlocutor. Para hacer esto, el intérprete debe calibrar las preferencias del hablante con las suyas, bajo la guía de lo que a sus luces es racional.

Por contraparte, el hablante, en general, no se encuentra en tal situación; no tiene que interpretar radicalmente sus propias preferencias para saber lo que significan o lo que él cree: "...no puede preguntarse si generalmente quiere decir lo que dice" (Davidson, 1984: 12). Por supuesto, si el hablante tiene la intención de ser comprendido por su audiencia, debe contribuir procurando hacer aserciones sinceras y proporcionado indicios de cómo su lenguaje se relaciona con situaciones que son familiares para quien lo escucha o bien que éste puede captar; pero, para el hablante, la cuestión de calibrar su propio lenguaje y sus estándares de racionalidad o bien no se presenta o bien resulta trivial. De este modo, cuando un hablante se adscribe actitudes proposicionales de forma veraz, el intérprete debe asumir que tales adscripciones son verdaderas en virtud de que el hablante es in-

terpretable *i.e.*, en virtud de las creencias e intenciones que se manifiestan en su conducta lingüística. Así, si el hablante hace preferencias verdaderas sobre sus actitudes proposicionales, puede decirse que la creencia que expresan sus aserciones está justificada y, por ende, el hablante *sabe lo que quiere decir con sus palabras y cuáles son los estados mentales que se adscribe*.

En la exposición anterior se encuentra encapsulada la explicación davidsoniana de la autoridad de la primera persona. Ésta toma la forma de una inferencia a la mejor explicación.⁵⁶ Su punto de partida es el hecho de que, generalmente, la comunicación interpersonal es exitosa (sin conceder esto, el escéptico ni siquiera puede hacer inteligible su duda). Lo que mejor explica la posibilidad de la comunicación es la suposición, por parte del intérprete, de que las creencias del hablante sobre el significado de sus palabras explican, racionalmente, sus preferencias lingüísticas. Así, si el hablante es interpretable, presenta como verdadera una autoadscripción de actitudes proposicionales (*i.e.*, si su adscripción es sincera) y se presume que tal adscripción expresa el contenido de sus creencias, entonces hay una fuerte presuposición de que tiene conocimiento sobre sus estados mentales. De modo que hay una presunción, esencial para la posibilidad de interpretación y, por ende, para la comunicación misma, de que el hablante sabe lo que cree cuando afirma que cree algo. Aunque nos hemos limitado a examinar el caso de adscripciones explícitas de estados mentales, una explicación similar puede generalizarse para el caso de las atribuciones: el intérprete debe asumir que si el hablante hace una aserción (y, por ende, sabe que presenta como verdadera la oración que profiere) y sabe lo que quiere decir con ella, entonces sabe lo que cree.

⁵⁶ Se ha insinuado también que podría tratarse de un argumento trascendental (*e.g.*, Maker, 1991; Hacker, 1997: 290; y Cutrofello, 1999). Sin embargo, el mismo Davidson (1999) considera que ésta no es una presentación idónea y que en principio el argumento no fue planteado de esa manera. He preferido exponerlo como una inferencia a la mejor explicación por sugerencia del doctor Jorge Tagle Marroquín. Para una discusión crítica en torno a los argumentos trascendentales véase la selección de Isabel Cabrera en *Argumentos trascendentales* (1999); la caracterización germinal de la inferencia a la mejor explicación se encuentra en Harman (1965).

Veamos si esta caracterización del fenómeno satisface las condiciones que le impusimos a una teoría explicativa del conocimiento de la propia mente.

1.5.3. La explicación davidsoniana y los retos de una teoría explicativa

En primer lugar, se requería que una explicación adecuada diera cuenta del fenómeno como un logro cognitivamente sustancial, a la par de nuestro conocimiento del mundo y de otras mentes. Esto requiere que sea objetivamente evaluable como verdadero o falso. Al respecto, Davidson señala: “[I]as tres variedades de conocimiento son objetivas [...] en el sentido de que su verdad es independiente de que se crea que son verdaderas. [...] Incluso en el caso de creencias acerca de nuestras propias creencias [...] tales creencias pueden ser erróneas” (Davidson, 2001: xiii). Pero, ¿en qué sentido podría decirse que el hablante tiene creencias falsas acerca de sus propias creencias? Apreciar la respuesta a esta interrogante estriba, nuevamente, en no perder de vista el problema lingüístico. Lo que significan las palabras del hablante “...depende en parte de las pistas para la interpretación que le ha dado al intérprete, o de la evidencia que justificadamente cree que el intérprete tiene. El hablante puede estar equivocado acerca de lo que sus propias palabras significan. Ésta es una de las razones por las que la autoridad de la primera persona no es infalible” (Davidson, 1984: 13).

Por otra parte, la explicación debería dar cuenta de los rasgos característicos del conocimiento de la propia mente, a saber: de la autoridad de la primera persona (*i.e.*, la presunción de verdad que acompaña a las autoadscripciones de actitudes proposicionales) y de la asimetría epistémica (*i.e.*, el carácter directo, o no mediado por inferencias a partir de la conducta, de nuestras autoatribuciones de pensamientos). La explicación de estos rasgos se encuentra anclada, para Davidson, en las condiciones de la interpretación. El oyente debe asumir que lo que el hablante quiere decir con sus palabras cuando se adscribe una actitud proposicional es pro-

ducto de sus creencias y de su intención de ser comprendido, *i.e.*, que sus preferencias son interpretables. Aunque el intérprete no está constreñido a suponer que la oración que expresa el contenido de las actitudes proposicionales del hablante es verdadera,⁵⁷ sí debe suponer en la medida en que propicie la comunicación que dicha oración expresa el contenido de los pensamientos del hablante y, por ende, que sus adscripciones son verdaderas. Esto da cuenta de la autoridad de la primera persona, *i.e.*, del hecho de que las autoadscripciones de actitudes proposicionales constituyan una razón para creer que el estado mental sobre el que tratan efectivamente es el caso.

La explicación davidsoniana esclarece, de una manera muy natural, porqué la autoridad está tan estrechamente asociada a la primera persona; esto depende, nuevamente, de situar el problema en las condiciones de la interpretación: “el conocimiento de los contenidos de nuestras propias mentes debe, en la mayoría de los casos, ser trivial. La razón es que, dejando de lado casos especiales, el problema de la interpretación no puede surgir” (Davidson, 1991: 217). Es la condición de hablante lo que le confiere la autoridad a quien se adscribe estados mentales, no una misteriosa relación (epistémica o metafísica) con su vida interna.

Pero, ¿en qué sentido esta explicación elucida la asimetría epistémica? Situémonos nuevamente en el problema lingüístico. Para que el oyente esté justificado en atribuirle un estado mental al hablante debe comprometerse en una inferencia que involucra *interpretación radical*; el hablante, en cambio, sólo debe asegurarse de que sus preferencias sean *interpretables*:

Para poner el asunto en su forma más simple: no puede haber una garantía general de que un oyente está interpretando correctamente a un hablante; sin importar cuán fácil, automática, irreflexiva y exitosamente un oyente comprenda a un hablante, él es propenso a errores serios. En este sentido especial, él siempre puede ser conside-

⁵⁷ El oyente puede, *e.g.*, dudar que las creencias del hablante sean verdaderas, cuestionar la factibilidad de que sus deseos se realicen, o considerar que sus esperanzas y temores son infundados.

rado como interpretando a un hablante. El hablante no puede, en la misma manera, interpretar sus propias palabras. Un oyente interpreta (normalmente sin reflexión o pausa) sobre la base de muchas pistas: la educación, el lugar de nacimiento, la inteligencia y profesión del hablante, la relación de del hablante con objetos cercanos y distantes, y así sucesivamente. El hablante, aunque debe tener muchas de estas cosas en mente cuando habla, pues a él le corresponde tratar de ser comprendido, no puede preguntarse si generalmente quiere decir lo que dice (Davidson, 1984: 12).

Hay que notar que ésta es una elucidación de la asimetría epistémica y no apela a una forma de acceso epistémico privilegiado. Hay una diferencia entre lo que justifica las atribuciones en tercera y en primera personas: el oyente debe hacer hipótesis empíricas sobre el lenguaje del hablante (su idiolecto) a partir de su conducta públicamente observable; el hablante, en general, no necesita hacer esto.⁵⁸ Sin embargo, esto no se debe a que el hablante tenga un contacto especial con sus propios pensamientos o a que disponga de información privilegiada sobre el contenido de su mente. El carácter no mediado por observación e inferencia del conocimiento de la propia mente es consecuencia de que, para descifrar lo que significan las propias palabras no se requiere (en general) interpretación radical. Este conocimiento es, al igual que el del mundo externo y el de otras mentes, *empírico* (véase Davidson, 1991: 205), ya que lo que da el significado que tienen a las palabras del hablante (y el contenido que tienen a sus pensamientos) es aquello a lo que regularmente las aplica con el fin de ser comprendido por el oyente (véase Davidson, 1987: 37).

1.6. Problemas y perspectivas en la concepción davidsoniana

Hasta ahora hemos visto que la teoría explicativa que ofrece Davidson del conocimiento de la propia mente tiene mucho que la recomiende. Da cuenta en forma satisfactoria de los rasgos especiales de este conocimiento, que identificamos en

⁵⁸ “La asimetría descansa en el hecho de que el intérprete debe, mientras no lo hace el hablante, confiar en lo que, si se hiciera explícito, sería una complicada inferencia al interpretar al hablante” (Davidson, 1984: 13).

la sección 1.2. No requiere para ello de la postulación de dualismos metafísicos o epistémicos, como los que precisaban las teorías de acceso privilegiado que revisamos en la sección 1.3. Asimismo, no tiene ningún compromiso con la tesis conductista de que pueden definirse explícitamente los términos mentales empleando vocabulario conductual, ni involucra hacer modificaciones sustantivas en la semántica de los verbos psicológicos cuando se aplican en primera y en tercera persona, como lo hace el neoexpresivismo, según veíamos en la sección 1.4. De modo que, por el momento, la explicación davidsonina cumple todas las condiciones que le hemos impuesto.

Además, es importante notar que la teoría de Davidson caracteriza al fenómeno que nos ocupa de manera epistémica, como una forma de conocimiento, en el sentido acotado en la sección 1.1.⁵⁹ En este aspecto la explicación davidsoniana, aunque se distancia de ella en muchos sentidos, preserva uno de los rasgos de la concepción cartesiana: la suposición cognitiva. Con ella, hereda parte de la agenda de la tradición epistemológica en la que se inscribe el cartesianismo. Si el conocimiento de la propia mente constituye en efecto un logro epistémico, ¿cuáles son las relaciones que éste guarda con nuestro conocimiento del mundo y de las otras mentes?, ¿es posible reducir una de estas formas de conocimiento a las otras? Para responder a estas interrogantes hemos de hablar de algunos de los rasgos más originales e influyentes de la filosofía de Donald Davidson. Un examen más sosegado de estos rasgos nos permitirá comprender con mayor claridad el lugar que le asigna al conocimiento de la propia mente en su proyecto epistemológico. No obstante, postergaremos esta exposición hasta el capítulo 3.

⁵⁹ Esto es algo que sus críticos en ocasiones no reconocen. Brie Gertler, *e.g.*, no la considera como una explicación sustantiva del *conocimiento* de la propia mente (véase Gertler, 2008 y 2011). Sanford Goldberg manifiesta una actitud similar cuando señala: “Davidson tuvo mucho menos que decir sobre la naturaleza positiva de los mecanismos a través de los cuales los sujetos tienen tal conocimiento de la propia mente. Lo poco que dijo [...] sugiere que rechazaría la apelación a cualquier clase de componente constructivo en la explicación de la autoridad de la primera persona” (Goldberg, 2011: 312).

En el próximo capítulo nos ocuparemos de otra de las aristas que enfrenta la explicación davidsoniana, en la medida en que persevera en la suposición cognitiva. Como indicamos en la sección 1.2, parece haber una profunda conexión entre el conocimiento y la posibilidad del error. Ésta fue una de las razones por las que no consideramos que fuese razonable asignarle a la autoridad de la primera persona el estatuto de infalibilidad. Señalábamos, además, que la posibilidad de error está conectada con que haya condiciones públicas de corrección; por ello la incorregibilidad nos parecía, también, una condición demasiado robusta. Davidson parece concordar en este punto:

Aunque hay autoridad de la primera persona con respecto a las creencias y otras actitudes proposicionales, el error es posible [...] y también lo es la duda. De modo que no siempre tenemos conocimiento indubitable o con certeza de nuestras propias actitudes. Tampoco son incorregibles nuestras afirmaciones acerca de nuestras propias actitudes proposicionales. Es posible que la evidencia disponible para otros desafíe nuestras autoadscripciones (Davidson, 1984: 4).

No obstante, una cosa es señalar que error y corrección son posibles; otra es explicar cómo son posibles. A la teoría davidsoniana de la autoridad la acompaña una teoría del error: una persona no sabe cuáles son sus pensamientos si *no sabe lo que significan sus palabras* o si *no es interpretable*. ¿Bajo qué condiciones puede ocurrir esto?

Como se indicó en la sección precedente, el hablante puede estar equivocado sobre lo que significan sus palabras. Esto es posible, de acuerdo con Davidson, debido a que "...lo que las palabras de una persona significan depende, en los casos más básicos, de las clases de objetos y sucesos que causaron que la persona considerara que las palabras eran aplicables; de manera similar para aquello de lo que tratan los pensamientos de la persona" (Davidson, 1987: 37). Esta tesis se conoce como externalismo.

Asimismo, señalábamos en la sección previa que una de las condiciones para que la comunicación sea posible es que el hablante sea interpretable, pero

“[o]bviamente el hablante puede fallar en este proyecto por momentos; en ese caso podemos decir, si nos parece, que no sabe lo que sus palabras significan” (Davidson, 1984: 13). Una situación paradigmática en la que el hablante falla en el proyecto de hacerse interpretable, en lo que concierne a sus propios estados mentales, se presenta cuando sostiene creencias irracionales, *i.e.*, en el caso del autoengaño.

El externalismo y el autoengaño representan dos retos para la explicación davidsoniana del conocimiento de la propia mente. Esto se debe a que, aunque en los casos arriba señalados debemos diagnosticar que el hablante no sabe lo que piensa, las fallas en el conocimiento de nuestros estados mentales no deberían, en general, socavar la autoridad de la primera persona. En el siguiente capítulo veremos cómo Davidson sugiere que debemos tratar estos casos.

Capítulo 2

Externalismo, autoengaño y autoridad de la primera persona

...it doesn't follow, simply from the fact that meanings are identified in part by relations to objects outside the head, that meanings aren't in the head.

(Davidson, 1984: 31)

...however wild the pattern of beliefs we are willing to attribute to the self-deceiver, we must not fall into contradiction ourselves in describing his confusion.

(Davidson, 1997: 216-217)

En el capítulo anterior se confrontaron diversas concepciones del conocimiento que cada uno tiene de sus propios estados mentales y se ofrecieron argumentos para desestimar algunas de estas explicaciones. Asimismo, se presentaron razones para respaldar la reconstrucción teórica que ofrece Donald Davidson de la autoridad de la primera persona. Se señaló que para explicar lo que hace especial al conocimiento de sí frente a otras clases de conocimiento (su presunción de verdad y su carácter de no mediado por observación e inferencia) no se requiere apelar a una forma de acceso epistémico privilegiado, sino situarlo en las condiciones de la interpretación. No obstante, se sostuvo que hay aspectos epistémicos involucrados en el conocimiento de la propia mente.

En este capítulo se considerarán dos retos que enfrenta la explicación davidsoniana de la autoridad de la primera persona: el externalismo y el autoengaño. Ambos se desprenden de considerar a la autoridad de la primera persona como

involucrando aspectos epistémicos. En la sección 2.1 se examinan sumariamente las motivaciones y argumentos a favor de la tesis de que el significado lingüístico y por extensión el contenido mental está determinado por factores independientes al sujeto, *i.e.*, la tesis del externalismo. En la sección 2.2 se presentan las razones que sugieren que existe una tensión entre el externalismo y el conocimiento de la propia mente. En la sección 2.3 se defiende que la reconstrucción davidsoniana del conocimiento de la propia mente es compatible con su formulación del externalismo, y se ofrecen también razones para preferir esta última por encima de formulaciones alternativas. En la sección 2.4 se expone la discusión en torno a cómo definir el autoengaño, así como razones para desestimar algunas de sus caracterizaciones. En la sección 2.5 se presenta la explicación de Davidson del autoengaño, como parte de su teoría del error en relación con el conocimiento de la propia mente. La sección 2.6, como preludeo a los contenidos del capítulo 3, elabora las conexiones de lo discutido en este capítulo con la concepción davidsoniana de la racionalidad y con el proyecto epistemológico.

2.1 *Externalismo: motivaciones y alcance*

Como es bien sabido, uno de los problemas centrales en la agenda filosófica del siglo XX fue tratar de echar luz sobre la naturaleza de la representación, en particular de la representación lingüística. Se ha sugerido incluso que la centralidad de este problema es uno de los varios factores que dieron cohesión a la tradición analítica en la filosofía.¹ Aunque puede polemizarse la veracidad de esta hipótesis histórica, es innegable la atención que se le concedió al problema del significado en los albores de la filosofía del siglo pasado.

¹ Esto sugiere, *e.g.*, Dummett cuando escribe: “[s]ólo con Frege se estableció finalmente el objeto propio de la filosofía, a saber: primero, que la meta de la filosofía es el análisis de la estructura del *pensamiento*; segundo, que el estudio del *pensamiento* ha de ser claramente distinguido del estudio del proceso psicológico de *pensar*; y, finalmente, que el único método apropiado para analizar el pensamiento consiste en el análisis del lenguaje. Como he argumentado, la aceptación de estos tres principios es común a toda la escuela analítica” (1975: 458).

2.1.1. Externalismo e internalismo semánticos

Uno de los productos de los debates en torno al lenguaje y al significado — uno que ha capturado el interés de los filósofos desde hace cerca de cuatro décadas — fue la introducción en la discusión filosófica una tesis sumamente polémica: el externalismo semántico. Una formulación de esta tesis podría ser la siguiente:

(ES): El significado de algunas de las palabras que emplea un hablante no está sólo determinado por las ideas con las que lo asocia o por sus estados físicos internos, sino también por factores independientes del (o externos al) hablante (véase Lau & Deutsch, 2010 para una caracterización similar).

El internalismo semántico, en contraste, afirma que algunas propiedades semánticas, como el significado y la referencia lingüísticos, están determinadas únicamente por propiedades o sucesos intrínsecos al hablante. Empleando esta terminología, lo que el externalismo semántico sostiene es la negación del internalismo,² *i.e.*, que algunas propiedades semánticas sobrevienen (o dependen) de propiedades relacionales o de sucesos que involucran tanto al hablante como a su entorno.

Podemos ilustrar la distinción entre propiedades intrínsecas y relacionales por medio de algunos sencillos ejemplos. La propiedad o condición de ‘tener cáncer’ parece ser una propiedad intrínseca de un individuo, que involucra un crecimiento irregular e incontrolable de sus células, pues se trata de la misma condición sin importar qué factores la hayan causado (sean genéticos, ambientales o derivados de hábitos del individuo). La propiedad o condición de ‘tener una picadura de mosquito’, por otra parte, es una propiedad relacional, ya que encontrarse en esa condición no involucra sólo estados del individuo, como tener irritación cutánea, sino que requiere haber sido picado por un mosquito. De ma-

² Aunque podría suceder que una propiedad relacional no fuese lo opuesto a una propiedad intrínseca, ya que muchas propiedades parecen ser tanto relacionales como intrínsecas (véase Wheatherson, 2006). Para los propósitos de esta exposición se ignora esta posibilidad.

nera análoga, el externalismo semántico sostiene que el significado de algunas expresiones lingüísticas depende de factores en el entorno del hablante.

Puede ubicarse una de las motivaciones originales del externalismo semántico en la tesis wittgensteiniana de que “[e]l significado de una palabra es su uso en el lenguaje. / Y el *significado* de un nombre se explica a veces señalando a su *portador*” (Wittgenstein, 1953: § 43. Sigo la traducción de Alfonso García Suárez y Ulises Moulines). La idea subyacente a esta intuición es que, si el significado ha de desempeñar una función al explicar la comunicación humana debe tener un aspecto público. Davidson lo expresa de la siguiente manera: “[s]ea lo que sea que haya en el ‘significado’, sea lo que sea acerca del habla que la hace útil para la comunicación, debe ser transmitido por la conducta públicamente observada de los usuarios del lenguaje” (Davidson, 2003: 284-285. Una motivación similar se presenta en Bilgrami, 1992: 200-213).

2.1.2. Externalismo semántico y anti-individualismo

Si asumimos (como lo hicimos en la sección 1.5) que la individuación de las actitudes proposicionales requiere de la estructura lingüística fina incrustada en las cláusulas subordinadas, veremos que hay una clara conexión entre el externalismo semántico, tal como se ha caracterizado líneas arriba, y otra tesis influyente sobre la naturaleza de la representación, *i.e.*: el externalismo sobre el contenido mental intencional o anti-individualismo. La siguiente podría ser una formulación tentativa del anti-individualismo:

(AI): El contenido de los enunciados mentales intencionales de un individuo *depende*, al menos en parte, de relaciones entre ese individuo y su entorno; en otras palabras, la individuación de pensamientos intencionales depende de las relaciones que una persona guarda con su entorno (físico y, en algunos casos, social; *e.g.*, su contexto histórico o las prácticas de la comunidad

lingüística a la que pertenece). (Para una caracterización similar véase Ludlow & Martin, 1998: 1; Burge, 1988: 111; y Gertler, 2008 y 2011: 81-82.)

Las relaciones de prioridad conceptual entre (ES) y (AI) dependen, en buena medida, de la posición que se adopte frente a la discusión sobre la relación entre lenguaje y pensamiento. Por lo general se considera que los argumentos que apoyan una de estas tesis pueden extenderse de manera natural a la otra.³ No obstante, es preciso hacer una acotación sobre este punto. En algunos casos (AI) es una tesis mucho menos polémica que (ES). Esto se debe a que algunas actitudes proposicionales, aquellas que están estructurada *de re*, claramente dependen de las relaciones del individuo con su entorno; en otros casos, cuando se estructuran *de dicto*, la tesis no es tan trivial. La distinción *de re/de dicto* es una distinción sobre alcance (*scope*);⁴ en este caso, sobre el alcance de los verbos psicológicos. Podemos ilustrar este punto si consideramos algunas adscripciones de actitudes proposicionales:

- (1) Juan cree de su vecino que es un espía.
- (2) Juan cree que su vecino es un espía.

La oración (1) representa una atribución de creencia estructurada *de re*, *i.e.*, la creencia de Juan es acerca de su vecino (sea como sea que él lo caracterice) y depende, para ser verdadera, de la existencia del vecino de Juan. Que sea correcto atribuirle a Juan la creencia expresada en (1) depende no sólo de cuáles sean las creencias de Juan, sino de que de hecho tenga un vecino. La oración (2), en cambio, está estructurada *de dicto*, *i.e.*, la creencia de Juan es acerca de una oración

³ Al respecto, Robert Stalnaker señala: “Uso el término ‘anti-individualismo’ introducido, en el sentido relevante, por Tyler Burge para la tesis de que la propiedad de tener un pensamiento con cierto contenido intencional es una propiedad relacional en lugar de una propiedad intrínseca del que piensa. La tesis es a menudo etiquetada como ‘externalismo’, pero no quiero confundir esta tesis específica con la aproximación externalista más general a la intencionalidad. Hay conexiones: creo que la tesis anti-individualista está motivada por la aproximación general, pero uno podría aceptar una formulación externalista del problema de la intencionalidad, mientras defiende el individualismo acerca del contenido mental” (Stalnaker, 2008: 114, n. 5). En lo que sigue, se habla en general del externalismo como una tesis que incluye el anti-individualismo.

⁴ Para una discusión sobre la ambigüedad de alcance de los verbos psicológicos, véase Quine (1956).

y, aparentemente, Juan podría tener esa creencia incluso si no tuviese vecinos (si, *e.g.*, alguien fingiese que habita en la casa contigua). La verdad de (2) parece depender únicamente de cuál sea el estado mental de Juan, sin importar si tiene vecinos.⁵ Las actitudes estructuradas *de dicto* son las que han suscitado mayor controversia, por lo que al hablar de (AI) nos centraremos en ellas.

Dada la estrecha afinidad entre (ES) y (AI), en lo que sigue me referiré a ellas o a una posición más general que las incluye a ambas indistintamente como ‘externalismo’, a menos que la distinción sea requerida. Aunque los argumentos en favor de esta postura — incluso en su formulación más débil — no son aún concluyentes, se exploran ya las consecuencias que tendría en otros campos *si fuese* verdadera. Se ha sugerido que sus implicaciones, si fuese correcta, serían relevantes para teorías (tanto empíricas como filosóficas) de la mente y de la acción (véase Lau & Deutsch, 2010), así como para la epistemología. En este último ámbito, se considera que podría proporcionar una novedosa respuesta al problema planteado por el escepticismo cartesiano,⁶ aunque también ha generado importante controversia en lo relativo al conocimiento de los propios estados mentales.⁷

2.1.3. Argumentos a favor del externalismo

Los argumentos más conocidos a favor del externalismo emplean experimentos mentales. Estos argumentos consisten en el planteamiento de un escenario con-

⁵ Para una discusión más detallada de la distinción *de re/de dicto* en torno a las actitudes proposicionales, puede consultarse con provecho (McKay & Nelson, 2010).

⁶ Famosamente, Putnam sugirió (en su 1981: 5-21) que una consecuencia del externalismo semántico es que no podríamos *decir o pensar* que nos encontramos en alguno de los escenarios contrafácticos que propone el escéptico cartesiano (el sueño perpetuo, el engaño por parte del genio maligno o siendo cerebros en cubetas manipulados por un super-científico). Para un panorama de esta discusión puede consultarse con provecho Brueckner (2012). Véase también lo discutido en el capítulo 3.

⁷ Un muestra de la inmensa literatura que ha generado esta vertiente del debate puede apreciarse en numerosas compilaciones, *e.g.*, *Externalism and Self-Knowledge* (1998), *New Essays on Externalism and Self-Knowledge* (2003) e *Internalism and Externalism in Semantics and Epistemology* (2007). Nos ocuparemos de esta discusión en la sección 2.2.

trafáctico concebible⁸ (que a menudo se describe como un ‘mundo posible’), el cual varía en algunos aspectos relevantes del mundo actual, seguido de una apelación a nuestras intuiciones conceptuales⁹ (sobre lo que diríamos en tal situación). Como señala Dennet “...estos experimentos mentales imitan a los experimentos empíricos en su diseño, intentando aislar una interacción crucial entre variables al mantener otras variables constantes” (1996: 76).¹⁰ No obstante, a diferencia de los experimentos de las ciencias, los experimentos mentales no pretenden establecer una relación empírica o meramente fáctica, sino probar un punto conceptual o una relación que se sostiene incluso en condiciones contrafácticas.

Los experimentos mentales que apoyan la tesis del externalismo mantienen constantes las propiedades intrínsecas de los hablantes o pensadores, mientras varían sus condiciones externas. Esos experimentos, como explica Burge (1988: 112), mantienen constante la historia causal de una persona (su movimiento corporal, sus estímulos superficiales y su química interna, pero también los enunciados con los que describiría sus estados mentales), *i.e.*, sus propiedades intrínsecas, variando el entorno (físico o social) en el que la persona interactúa (las propiedades extrínsecas). A partir de esto luego se muestra que los significados de algunas de las oraciones que emplean o algunos de los pensamientos que sería correcto atribuirles varían. Esta variación indica que, en el mejor de los casos, deben

⁸ Siguiendo el famoso *dictum* de Hume, se considera frecuentemente a la concebibilidad como una guía más o menos segura de la posibilidad: “[c]ualquier cosa que la mente concibe con claridad incluye la idea de existencia posible, o en otras palabras, nada de lo que imaginamos es absolutamente imposible” (Hume, 1739: 81). Eric Marcus (2004: 477) ha señalado que hay dos formas importantes de impugnar estos argumentos: “...se puede negar que lo que se pretende concebible de hecho lo sea [...]; o se puede negar que la concebibilidad establezca posibilidad”. Yablo (1993) discute con detalle la cuestión de si la concebibilidad que ofrece respaldo para inferencias acerca de la posibilidad.

⁹ Al respecto Frank Jackson (1998: 30, n. 4) ha señalado que cada uno de nosotros, en tanto hablante típico, puede representar un caso ejemplar de qué intuiciones conceptuales son plausibles y cuáles no.

¹⁰ Aunque es común emplear experimentos mentales en la argumentación filosófica (son célebres, *e.g.*, los de Jackson, 1982 y 1986; Chalmers, 1996; así como los que se presentan a continuación), varios filósofos han objetado su efectividad o relevancia (*e.g.*, Dennett, 1991 y 2005; Papineau, 1993) al señalar que “[u]n problema con tales experimentos es que la variable dependiente es la ‘intuición’ son bombas de intuiciones y la contribución de la imaginación es más difícil de controlar de lo que los filósofos han supuesto” (Dennett, 1996: 76). Véase la nota 22 para la opinión de Davidson sobre los experimentos mentales.

considerarse principios de individuación de propiedades semánticas o de contenidos mentales a partir de propiedades extrínsecas a la persona en cuestión.

En su célebre (1975), Hilary Putnam presentó un experimento mental a favor del externalismo.¹¹ En él, nos pide que imaginemos un mundo (la Tierra Gemela) físicamente idéntico al actual con la única diferencia de que en él se encuentra el compuesto XYZ, con las mismas propiedades superficiales del agua, en sustitución de lo que en nuestro mundo es H₂O. Continúa sugiriendo que consideremos la existencia en ese mundo de un sujeto *S'*, quien es un duplicado físicamente idéntico de *S*, un habitante del mundo actual. Asimismo, los lenguajes de *S* y *S'* son muy similares: sólo se distinguen en que, mientras *S* emplea la palabra “agua” para referirse a lo en nuestro mundo es H₂O, *S'* usa “agua” para designar al compuesto XYZ (el líquido que se encuentra en los mares de la Tierra Gemela y que sus habitantes consumen para saciar su sed). En seguida, se hace una apelación a nuestras intuiciones conceptuales: si *S* y *S'* profirieran la oración “Eso es agua” mientras señalan los océanos de sus respectivos planetas, ¿significarían lo mismo ambas preferencias? La respuesta intuitiva es “No”, pues sus preferencias tendrían condiciones de verdad distintas: mientras cada uno de ellos diría algo verdadero en sus respectivos planetas al proferir esa oración, su preferencia sería falsa si fuesen transportados de un planeta a otro. Esto podría extenderse a sus actitudes proposicionales: al creer (desear, esperar, temer, etc.) que tienen “agua” frente a sí, ambos individuos tendrían pensamientos distintos. De modo que el experimento mental de Putnam presumiblemente muestra que el significado de las expresiones (y el contenido de los pensamientos) que involucran términos de clase natural (como “agua”, “oro”, “tigre”, etc.) está, en parte, determinado por el tipo de sustancias que se encuentran en el entorno del hablante. Saul Kripke, en su

¹¹ Vale la pena señalar que, con excepción de Davidson (1987), ninguna de las formulaciones originales de los argumentos que a continuación se presentan emplea explícitamente el término ‘externalismo’ (Burge habla, en su lugar, de anti-individualismo). Sin embargo, todos parecen avalar alguna versión de la tesis que hemos caracterizado bajo tal designación.

(1972), había presentado argumentos similares a favor de que el contenido semántico de nombres propios y términos de clase natural está determinado por conexiones causales con los objetos en el entorno, inmediato o remoto, del hablante.

Por su parte, Tyler Burge (1979), empleando un experimento mental similar al de Putnam, añadió que la comunidad de hablantes de la que forma parte el individuo es también relevante en la determinación del significado de sus palabras y del contenido de sus pensamientos. Burge consigna esta conclusión parafraseando un célebre pasaje de John Donne: “[n]ingún fenómeno mental intencional es insular. Cada hombre es una pieza del continente social, una parte del todo social.” (Burge, 1979: 116). Asimismo, Burge indicó que esto no se limita a conceptos de clase natural (usa el caso de “artritis”, que, se sugiere, no constituye una clase natural), sino que incluye otras nociones observacionales y teóricas. Su externalismo se extiende a “[c]ualquier noción que se aplique a tipos públicos de objetos, propiedades o sucesos que sean típicamente conocidos por medios empíricos” (Burge, 1986: 223).

Donald Davidson, en su (1987), también presentó un experimento mental que se propone respaldar una forma aún más general de externalismo. Nos propone que imaginemos que, mientras camina en las proximidades de un pantano, es fulminado por un relámpago que lo reduce a sus moléculas elementales; mientras tanto, por mera coincidencia, aparece una réplica física suya: el hombre de los pantanos. Éste profiere oraciones y exhibe una conducta similar a la del Davidson original. No obstante, aunque se comporte como él, intuitivamente diríamos, sugiere Davidson, que no tiene actitudes proposicionales, pues carece de una historia causal apropiada para el aprendizaje del lenguaje.

La moraleja de los argumentos de Putnam (1975), Kripke (1972), Burge (1979 y 1986) y Davidson (1987) es: tenemos muy buenas razones para *crear* que alguna formulación de la tesis del externalismo es verdadera. O, para ponerlo en otros términos: la moraleja es que significado y contenido mental son, probablemente,

propiedades relacionales. No obstante, como indican Ludlow & Martin (1998: 1) esta tesis parece producir una tensión con la plausibilidad *prima facie* de la idea de que podemos tener un tipo especial de conocimiento sobre el contenido de nuestros estados mentales. Como señala Gertler (2003), casi todos aquellos que aceptan el externalismo acerca del contenido mental afirman que esta tensión es meramente aparente. No obstante, algunos consideran que ambas tesis son mutuamente excluyentes y, por ende, si ha de aceptarse una de ellas, deberá rechazarse la otra.¹² Pocos de los que consideran esta conjunción como paradójica están dispuestos a abandonar por completo la idea de que tenemos conocimiento privilegiado de nuestra propia mente.

En la siguiente sección se examinarán algunos de los argumentos que se han ofrecido para sugerir que el externalismo es incompatible con el conocimiento de la propia mente. En la sección 2.3 se mostrará que la formulación davidsoniana del externalismo no sólo es compatible con su teoría explicativa de la autoridad de la primera persona, sino que suministra parte de lo que se requiere para explicar las fallas en el conocimiento de la propia mente.

2.2. El debate sobre la incompatibilidad entre externalismo y autoridad de la primera persona

Se han indicado principalmente dos sentidos en los que es problemático asociar la tesis del externalismo con una explicación epistémica de la autoridad de la primera persona (véase Davies, 1998a y Brown, 2009). El primero, que se conoce como el ‘problema del logro (*achievement*)’, consiste en que el externalismo parece

¹² Existen otras alternativas para evitar la incompatibilidad: rechazar ambas tesis o introducir una tercera que elimine la tensión. Agradezco la oportuna precisión sobre este punto al doctor Jorge Tagle Marroquín. En lo que sigue asumiremos que alguna formulación del externalismo semántico es correcta, por lo que no se considerará la primera de estas alternativas; Parent (2007) parece optar por la segunda. Los argumentos que se exponen en la siguiente sección no toman en cuenta esta última posibilidad.

hacer difícil que un individuo logre tener conocimiento de sus estados mentales (*e.g.*, parece requerir que el individuo deba buscar información empírica sobre su entorno para saber lo que piensa); el segundo, que se denomina ‘problema de la consecuencia’, indica que aparentemente la combinación de externalismo y conocimiento de la propia mente tiene consecuencias absurdas (*e.g.*, que saber lo que uno piensa implica saber qué cosas existen). La discusión de ambos problemas a menudo se formula empleando modalidades epistémicas (*a priori/a posteriori*) para enfatizar el carácter especial del conocimiento de la propia mente; no obstante, en lo que sigue se asumirá que son los rasgos destacados en el capítulo 1, y no las modalidades epistémicas, lo que distingue al conocimiento de la propia mente de otras formas de conocimiento.

Aunque difieren en su diagnóstico de cuál de las tesis habría que abandonar, Paul Boghossian, Michael McKinsey y Jessica Brown han presentado varios argumentos que se proponen mostrar que el externalismo y el conocimiento de la propia mente son incompatibles. En lo que sigue veremos algunos de estos argumentos y cómo dependen de las caracterizaciones que emplean de la autoridad de la primera persona y del externalismo.

2.2.1. El problema del logro

Aunque el problema del logro ya había sido suscitado por Putnam (1975),¹³ Paul Boghossian hizo patente que esta dificultad se presenta al suponer que el contenido de nuestros pensamientos está determinado por factores externos al individuo si se suscribe también la idea de que podemos saber cuáles son nuestros pensamientos “directamente, sin el beneficio de inferencia a partir de otras creencias”

¹³ Quien ya había señalado que lo que determina el significado y la extensión “no es, en general, completamente conocido por el hablante” (1975: 165)

(Boghossian, 1989: 149).¹⁴ El problema radica en que, si las propiedades del contenido de nuestros pensamientos se individualizan por medio de factores externos (*i.e.*, si se trata de propiedades relacionales), entonces “[l]o que necesitaríamos ver, si hemos de saber meramente al mirar, no está ahí para ser visto” (Boghossian, 1989: 149).

El problema puede ilustrarse si consideramos una versión del ejemplo clásico de Putnam (1975), que se denomina el escenario del cambio prolongado (*slow-switching*). Volvemos a encontrarnos con *S*, un habitante de la Tierra, y con *S'*, su réplica física exacta en la Tierra Gemela. Si el argumento de Putnam es correcto, los pensamientos de *S* acerca de lo que él llama “agua” están determinados por H₂O (llamémoslos pensamientos sobre el concepto de *agua*), mientras que los pensamientos de *S'* sobre lo que él denomina “agua” dependen del compuesto XYZ (llamémoslos pensamientos sobre el concepto de *tagua*). Son pensamientos distintos en virtud del tipo de sustancias en sus entornos respectivos, incluso si ellos desconocen la naturaleza de dichas sustancias.

Ahora bien, supongamos que lo siguiente ocurre: sin percatarse de ello, *S* y *S'* son intercambiados, en múltiples ocasiones durante periodos prolongados de tiempo, de sus planetas nativos. “La mayoría de los [externalistas] están de acuerdo en que, si una persona permaneciera en cada situación durante suficiente tiempo, esa persona eventualmente adquiriría los conceptos apropiados para tal situación” (Boghossian, 1989: 159). De modo que, al estar en la Tierra adquieren y emplean correctamente el concepto de *agua*, mientras que sucede otro tanto en la Tierra Gemela con el concepto de *tagua*. No obstante, si hubiesen proferido la oración “Esto es agua” un día antes de ser transportados de planeta, no podrían

¹⁴ Más adelante añade: “[e]sto implica que conocemos nuestros pensamientos ya sea sobre la base de alguna forma de observación interna, o sobre ninguna base” (Boghossian, 1989: 149). Tal como lo presenta aquí, y en el resto de su ensayo, Boghossian parece creer que las mejores caracterizaciones del conocimiento de la propia mente lo describen como una forma de acceso epistémico privilegiado (la teoría del sentido interno) o como un fenómeno no epistémico (la teoría neoexpresivista). Véanse los apartados 1.3.3 y 1.4.2 para la discusión de estas teorías explicativas.

decir cuál de los conceptos estaban empleando (y, por ende, cuáles eran sus pensamientos) a menos que supiesen en qué entorno se encontraban.

Pero, ¿qué es lo que saben S y S' ? Por hipótesis, no se percatan de que sus lugares han sido intercambiados, ya que, además, la Tierra y la Tierra Gemela son idénticas en todos sus aspectos superficiales. Si el externalista está en lo correcto, tampoco saben (a menos que investiguen las propiedades químicas de sus respectivos entornos) cuál es el contenido de sus pensamientos. De modo que, al parecer, si estaban en condiciones de conocer el contenido de sus pensamientos en primer lugar (incluso sin que hubiese ocurrido el cambio prolongado de sus entornos) este conocimiento involucraría investigación empírica. Pero entonces no es claro en qué sentido el conocimiento que tenían de sus propios pensamientos pudiera ser privilegiado o especial con respecto al conocimiento del mundo externo. Boghossian cree que esto sugiere que debemos ser más modestos con respecto al alcance del externalismo o bien reconsiderar las explicaciones del conocimiento de la propia mente, pues, como indica: “[n]o concibo seriamente que no conozcamos nuestras propias mentes. Nuestra capacidad de conocer la propia mente no es un componente opcional de nuestra concepción ordinaria de nosotros mismos, una tesis que podamos ser capaces de descartar preservando todo lo que realmente importa” (Boghossian, 1989: 150).

2.2.2. El problema de la consecuencia

El problema de la consecuencia fue planteado en numerosas ocasiones por Michael McKinsey, quien lo caracteriza como una *reductio ad absurdum* (la *reductio* aparece en su 1991 y se reitera en sus 2002a, 2002b y 2007).¹⁵ McKinsey comienza

¹⁵ Jessica Brown (1995) ofrece una reconstrucción más inquietante del argumento de McKinsey, quien la emplea en exposiciones posteriores al artículo de Brown.

por caracterizar la concepción del conocimiento de la propia mente (CPM) que le parece incompatible con el externalismo:

(CPM): Es necesariamente verdadero que si una persona S está pensando que p , entonces S puede, en principio, saber *a priori* que está pensando que p .

Aunque esta formulación, puesto que habla de conocimiento *a priori*,¹⁶ no es directamente equivalente a ninguna de las que examinamos en el capítulo 1, puede asumirse que (CPM) involucra una forma de acceso privilegiado al contenido de la propia mente, distinto del empleado en la observación sensorial o en la investigación empírica, para explicar el rasgo de la autoridad de la primera persona. Es importante notar que la caracterización de McKinsey de la autoridad de la primera persona no le otorga el estatuto de infalibilidad, pues —en tanto sigue a Kripke en la distinción entre *a priori* y necesario (véase Kripke, 1972: 34-39)— (CPM) no especifica que S siempre esté en lo correcto al atribuirse pensamientos. No obstante, McKinsey parece considerar que el conocimiento de la propia mente involucra una condición de acceso con el estatuto de luminosidad (o, al menos, de omnisciencia en principio) para las actitudes proposicionales. A esto McKinsey suma una formulación del externalismo sobre el contenido mental (ECM), aún a las tesis semánticas de Kripke (1972) y Putnam (1975):

(ECM): Muchos predicados estructurados *de dicto* de la forma ‘está pensando que p ’ expresan propiedades que son amplias, en el sentido de que la posesión de tal propiedad por parte de un sujeto implica, lógica o conceptualmente, la existencia de objetos contingentes o sustancias de cierto tipo que son externas al sujeto (véase McKinsey, 2002a: 35 y 2007).

McKinsey sugiere que, si empleamos estas tesis, podemos extraer una consecuencia epistemológica absurda. Consideremos la siguiente situación.

¹⁶ Brown señala que la caracterización de la autoridad de la primera persona como conocimiento *a priori* es polémica: “...podríamos decir que una proposición es conocida *a priori* si es conocida sin depender de la experiencia perceptual. [...] Hay un amplio acuerdo sobre que lo *a priori* incluye verdades matemáticas, lógicas y algunas definicionales. No obstante, hay desacuerdo acerca de cuánto la noción de *a priori* se extiende más allá de estos casos centrales” (Brown, 2004: 23).

- (I) S está pensando que el oro es un metal precioso.

De acuerdo con (CPM) podemos decir:

- (II) En principio, S puede saber *a priori* que está pensando que el oro es un metal precioso.

Dado que el oro constituye una clase natural, el predicado *de dicto* “está pensando que el oro es un metal precioso” expresa una propiedad amplia (relacional), por lo que de (ECM) se sigue que:

- (III) La proposición que S está pensando [“el oro es un metal precioso”] implica, lógicamente o conceptualmente, la proposición E .

Donde E es una proposición sobre el mundo externo, que asevera o implica la existencia de objetos externos a S (e.g., la proposición de que el oro existe). Puesto que (III) es una consecuencia de (ECM), y ésta es una tesis filosófica (y no una tesis de la ciencia empírica), McKinsey considera que, en principio, S podría saber *a priori* que (III) es verdadera. De modo que, de la conjunción de (II) y (III) se sigue¹⁷ una consecuencia absurda:

- (IV) Si S está pensando que el oro es un metal precioso, entonces, en principio, S puede saber *a priori* que el oro existe.

(IV) es claramente inaceptable, pues — a menos que se extienda ampliamente el campo de lo que puede conocerse *a priori* — no es posible, en principio, que S sepa independientemente de cualquier experiencia que el oro existe. Afirmar lo contrario sería reivindicar una forma radical de especulación racionalista¹⁸ (véase 2002a: 35 para el ejemplo de McKinsey). Puesto que esta conclusión se desprende

¹⁷ Mediante algún principio de inferencia, que preserve la condición de conocimiento *a priori*. Éste podría consistir en una reformulación del principio de clausura bajo implicación conocida como la siguiente: “Si S sabe *a priori* que p , y S sabe *a priori* que p tiene como consecuencia lógica inevitable que q , entonces S puede saber *a priori* que q a partir de p ”. Sin embargo, el que (ECM) no sea una hipótesis de la ciencia empírica no debería considerarse decisivo sobre su estatus *a priori*.

¹⁸ Sería similar a extender al dominio de los objetos contingentes el procedimiento empleado en el argumento ontológico de San Anselmo, que pretende demostrar la existencia de Dios mediante la pura reflexión sobre su concepto. El argumento anselmiano ha generado muchas sospechas; su contraparte con respecto a los objetos contingentes parece incluso menos razonable.

de (CPM) y (ECM), más la suposición aparentemente trivial de que alguien puede tener actitudes proposicionales, se sigue que una de las dos primeras tesis debe ser abandonada: $\neg[(\text{CPM}) \wedge (\text{ECM})]$.

Si hemos de elegir entre rechazar una de estas tesis, continúa McKinsey, no hay duda: rechazemos (CPM), puesto que tenemos mejores argumentos para (ECM).¹⁹ De modo que para conocer el contenido de nuestros pensamientos deberíamos investigar empíricamente nuestro entorno físico y social. Por ende, deberíamos abandonar la suposición de que el conocimiento de la propia mente involucra una presuposición especial de verdad y no requiere observación e inferencia.

Algunos defensores del externalismo propuesto por Burge han sugerido que, para evitar la conclusión de la *reductio* de McKinsey, sería preciso matizar la formulación del externalismo de modo que no tenga como consecuencia la tesis (ECM) empleada por McKinsey. Así, no habría forzosamente incompatibilidad entre el externalismo y la autoridad de la primera persona.²⁰ Sin embargo, McKinsey ha señalado enérgicamente que esta estrategia no es adecuada, pues ignora que lo que los ejemplos de Kripke, Putnam y Burge plausiblemente muestran es (ECM) (2002a: 35 y 2007: *passim*). Es en este sentido que McKinsey señala: “...la estrategia metafísica compatibilista elimina efectivamente la base semántica del externalismo, y al hacerlo *sustrae todas las razones que había para creer en el externalismo en primer lugar*” (2002a: 35. *Cursivas mías*).

Tanto el problema del logro como el de la consecuencia constituyen retos teóricos para cualquier explicación de la autoridad de la primera persona en la medida en que sea caracterizada como un fenómeno especial de carácter episté-

¹⁹ En las primeras formulaciones de la *reductio* no era claro que McKinsey tuviese una posición sobre cuál de las tesis debería abandonarse (como señala Brueckner, 2007: 38-39). No obstante, en sus últimos ensayos McKinsey ha sido muy enfático sobre este punto (véase 2002a, 2002b y 2007).

²⁰ Esta parece ser la línea seguida por McLaughlin & Tye (1998), quienes se inspiran en la defensa de Burge (en su 1988) de que el externalismo es compatible con una forma de “acceso privilegiado” a los propios estados mentales.

mico.²¹ Ambos argumentos sugieren que si el externalismo es correcto, entonces la autoridad de la primera persona no puede ser especial (con respecto al conocimiento del mundo y de otras mentes) o bien no puede ser epistémica. En lo que sigue se sugerirá que, no obstante, ni el problema del logro ni el de la consecuencia son incompatibles con los rasgos característicos del conocimiento de la propia mente que identificamos en el capítulo 1. Esto puede apreciarse al confrontarlos con la propuesta davidsoniana, la cual, como hemos argumentado, preserva tanto el carácter especial del conocimiento de la propia mente así su estatus epistémico.

Como hemos visto, los argumentos incompatibilistas emplean explicaciones del conocimiento de la propia mente que hemos desestimado en el capítulo anterior; asimismo, asumen que la versión correcta de la tesis externalista corresponde a las formulaciones ortodoxas de Kripke (1972), Putnam (1975) y Burge (1979 y 1986). En la siguiente sección examinaremos con más detalle el externalismo que Davidson defiende y veremos que, no sólo es compatible con su concepción de la autoridad de la primera persona, sino que (como indicamos en la sección 1.6) juega un papel importante en la explicación davidsoniana: suministra una parte de su teoría del error. Las secciones 2.4 y 2.5 se ocuparán de otro aspecto de esta teoría del error: el autoengaño como forma de irracionalidad.

2.3. Externalismo y autoridad: la concepción davidsoniana

Aunque Davidson no discutió explícitamente los argumentos de Boghossian, McKinsey y Brown, hay indicios de que reconoció la tensión que, *prima facie*, se produce al aceptar tanto el externalismo como la autoridad de la primera persona. No obstante, más que un obstáculo para su explicación de la autoridad de la pri-

²¹ Esto es algo que señala oportunamente Boghossian al discutir explicaciones no epistémicas de la autoridad de la primera persona (véase 1989: 164-172).

mera persona, él mismo reconoció que el externalismo constituía una parte importante de ella:

En años recientes [...] algunos de *los hechos mismos de los que dependí para defender la autoridad de la primera persona* han sido empleados para atacar esa autoridad: se ha argumentado, sobre *lo que se piensa que son bases nuevas*, que mientras los métodos del intérprete en tercera persona determinan lo que usualmente consideramos que son los contenidos de la mente de un agente, los contenidos así determinados pueden ser desconocidos para el agente. *La explicación que ofrecí* [...] de la asimetría entre atribuciones de actitudes en primera y [tercera] personas *me parece* [...] *robustecida* por las nuevas consideraciones, o *por aquellas que parecen válidas* (Davidson, 1987: 17. Cursivas mías).

Asimismo, señaló que la posición que uno adopte frente al problema de la incompatibilidad depende, en última instancia, de bajo qué formulación se suscriba de la tesis externalista:

El problema surge por admitir que el papel de identificación e individuación de los factores externos parece conducir a la conclusión de que nuestros pensamientos podrían no ser conocidos por nosotros.

Pero, ¿se sigue esta conclusión? La respuesta *depende de la manera en la cual uno piense que la identificación de los contenidos mentales depende de factores externos* (Davidson, 1987: 25. Cursivas mías).

En esta sección examinaremos la relación entre externalismo y autoridad de la primera persona en la teoría explicativa de Davidson del conocimiento de la propia mente. Se argumentará que, tal como la caracteriza este autor, dichas tesis no enfrentan las dificultades que se identificaron en la sección previa. En el capítulo 1 se ofrecieron razones para suscribir la caracterización davidsoniana del conocimiento de la propia mente; en esta sección se presentarán razones para preferir su versión del externalismo a lo que denomino la ‘formulación ortodoxa’ (que incluye varios aspectos de las contribuciones de Kripke, Putnam y Burge).

2.3.1. El externalismo davidsoniano

La formulación davidsoniana del externalismo, que presentamos en la sección 2.1.3 con el experimento mental del hombre de los pantanos, no está respaldada únicamente por la apelación a intuiciones conceptuales acerca de situaciones posibles,²² sino que constituye un engrane más de su teoría general del pensamiento, el lenguaje y la acción (véase Davidson, 1980) con miras a echar luz sobre la comunicación humana. Al respecto, Davidson reconoce sus deudas teóricas con Quine, de quien señala incluso que una de sus obras capitales, *Word and Object* (1960), “...salvó uno podría casi decir que inventó a la filosofía del lenguaje como un tema serio” (del manuscrito preparativo para Davidson, 2003).²³

El objetivo de Quine (1960) consiste en que explicar el significado lingüístico compartido de hablantes que, por otra parte, pueden tener creencias distintas. Éste parece ser un requisito para dar cuenta de cómo es posible la comunicación entre hablantes, que es lo que se quiere, en última instancia, explicar. Aunque ésta parece ser también la motivación más importante de los artículos clásicos de Putnam, Kripke y Burge, el externalismo de Quine (o, al menos, el que Davidson le atribuye) es más robusto y está guiado por intereses explicativos distintos, sobre la base de preferencias metodológicas diversas.

²² En efecto, Davidson comenta: “[q]uizá deba insistir en que los argumentos a favor [del externalismo que defiendo] no descansan en intuiciones concernientes a qué diríamos si ciertos contrafácticos fuesen verdaderos. Ninguna ciencia ficción o experimento mental son requeridos” (1987: 29). “Siento una desconfianza general de los experimentos mentales que pretenden revelar lo que diríamos bajo condiciones que de hecho nunca se presentan. Mi versión del externalismo depende de lo que considero nuestra práctica efectiva” (1991: 199). En retrospectiva, añade incluso: “[c]onfieso que el hombre de los pantanos ahora me avergüenza. Las historias de ciencia ficción que imaginan cosas que nunca suceden proporcionan un terreno de prueba muy pobre para nuestras intuiciones concernientes a conceptos como el de persona o el de qué constituye pensamiento. Estos conceptos comunes funcionan tan bien como se requiere en el mundo tal como lo conocemos. Hay múltiples criterios para aplicar muchos conceptos importantes, y los casos imaginados son unos en los cuales estos criterios, que normalmente van en conjunto, apuntan en diferentes direcciones. Nos preguntamos, ¿qué diríamos en tales casos? ¿Quién sabe? ¿Por qué debería importarnos? A menos que los casos de hecho ocurran. Si lo hacen, decidiremos qué decir...” (Davidson, 2006: 1061).

²³ Agradezco al profesor Jonathan Vogel por haberme proporcionado este material y por tomarse la molestia en lo que debió constituir un complicado ejercicio de interpretación radical de discutir conmigo varios de estos asuntos durante su estancia en la ciudad de México en junio de 2010.

Davidson (2003) sugiere que el externalismo de Quine tiene dos principales afluentes: por una parte tiene un componente perceptual, involucrado en su concepción del aprendizaje ostensivo de los enunciados observacionales, y por otra, un componente social, que se despliega en la noción de triangulación y en la adquisición del concepto de error. La situación básica en la que puede verse la interacción de estos componentes consta al menos de tres elementos: dos sujetos y un mundo de objetos, propiedades y sucesos que los sujetos pueden discriminar por medio de la percepción. Se trata de una tríada irreducible. Cada sujeto lleva a cabo sus inducciones habituales, sean aprendidas o heredadas. Para cada sujeto hay estímulos que pueden ser clasificados en conjunto, en virtud de la similitud de las respuestas. Para que haya pensamiento y lenguaje,²⁴ cada sujeto debe ser capaz de asociar las respuestas del otro sujeto con estímulos del mundo compartido. Este triángulo interconectado constituye una condición necesaria (mas no suficiente) para la existencia de conceptualizaciones, pensamiento y lenguaje. Hace posible la creencia objetiva y otras actitudes proposicionales. La tesis del externalismo se presenta al preguntar: ¿en dónde en esta triada hay que localizar lo que individúa al pensamiento? La respuesta es: en los objetos externos, ahí donde las percepciones de ambos sujetos convergen.²⁵ Se trata, por supuesto, de una forma (en ocasiones incluso literal) de triangulación: las líneas de visión convergen en el estímulo compartido.

Aunado a esto, para explicar la posibilidad del error se requiere que los sujetos sean capaces de apreciar las reacciones del otro hacia un entorno compartido (véase Davidson, 2001: 6). Hay, pues, una doble determinación de lo externo sobre

²⁴ “Por supuesto, si hay lenguaje hay pensamiento, de modo que no puede ser más fácil explicar el primero que el segundo. Sin embargo, es útil reconocer el del algún modo sorprendente hecho de que el elemento social que es esencial al lenguaje es también esencial al pensamiento mismo. Lo que hace a esta relación entre lenguaje y pensamiento útil en el presente contexto es que hace posible estudiar el pensamiento al estudiar el lenguaje” (Davidson, 2001: 13).

²⁵ Hay una notable tensión en la obra de Quine entre esta idea de que lo que determina el contenido es un estímulo distante (a la que Davidson se propone inclinarlo) y la de que se trata de un estímulo próximo, que parece ser la versión predilecta de Quine. Sigo en esto a Davidson (2003).

el significado lingüístico y el contenido mental: ésta se localiza, por un lado, en el objeto que es la causa de las percepciones de ambos sujetos, y por otro en la conducta observable de cada uno de ellos. Así pues, el significado de lo que un hablante dice, y por ende los pensamientos que puede adscribirse por medio del lenguaje, no están accidentalmente conectados con lo que un intérprete competente puede extraer de ellos. No obstante, esto no significa que hablante y oyente estén en la misma situación en lo que concierne a determinar lo que significan las preferencias del hablante. Recuérdese que el oyente se encuentra en una situación de interpretación radical, *i.e.*, debe hacer un tipo especialmente complicado de inferencia a partir de la conducta, el entorno y la historia del hablante, para generar una hipótesis empírica sobre lo que significan sus palabras y el contenido de sus pensamientos. El hablante, por su parte, no requiere (en la mayoría de los casos) observar su conducta para determinar lo que significan sus palabras,²⁶ sino que sólo debe aplicarlas sistemáticamente a objetos y situaciones que, desde su perspectiva, faciliten la labor del intérprete.

2.3.2. Respuestas a los argumentos incompatibilistas

Esto puede ayudarnos a dar cuenta del problema del logro. Por supuesto, el hablante puede fallar en su proyecto de hacerse interpretable, si, *e.g.*, hay una ambigüedad en su uso de la palabra “agua” al aplicarla a lo que un intérprete reconoce como dos tipos distintos de sustancias. En casos como éste podríamos llegar a decir que el hablante no sabe lo que significan sus palabras (ni, por ende, cuál es el contenido de sus pensamientos). Es, en parte, por esta razón que el conocimiento de la

²⁶ “Ni el hablante ni su oyente conocen de una manera especial o misteriosa lo que significan las palabras del hablante; ambos pueden estar equivocados. Pero hay una diferencia. El hablante, después de concentrar cualquier conocimiento y habilidad que tenga a la tarea de decir lo que sus palabras significan, no puede mejorar la siguiente clase de enunciados: ‘Mi preferencia de “Wagner murió feliz” es verdadera si y sólo si Wagner murió feliz’. Un intérprete no tiene ninguna razón para asumir que ésta será *su* mejor manera de enunciar las condiciones de verdad de la preferencia del hablante” (Davidson, 1984: 13).

propia mente constituye un *logro* cognitivo: es posible fracasar en su obtención. Sin embargo, esto no significa que no haya una presunción especial en su favor, que se encuentra ausente cuando se atribuyen pensamientos a otra persona. El hablante sigue teniendo autoridad epistémica sobre el contenido de sus pensamientos debido a que las atribuciones del intérprete se encuentran en una situación mucho más comprometida.

Retomemos el ejemplo del cambio prolongado de Boghossian. Supongamos que *S* es transportado, por primera vez e ignorándolo, a la Tierra Gemela. Un intérprete competente tiene que decidir cómo hacer más inteligible, a sus luces, la conducta (tanto lingüística como no lingüística) de *S* cuando profiere la oración “Esto es agua”, al señalar a los océanos de la Tierra Gemela. Podría atribuirle la creencia falsa de que lo que tiene frente a sí es H_2O , la creencia verdadera de que se encuentra frente a XYZ o incluso la creencia verdadera de que está situado frente a H_2O o XYZ; *i.e.*, podría asumir que sus pensamientos son sobre el concepto de *agua*, el de *tagua* o el concepto disyuntivo de *agua o tagua*. Cualquiera de estas hipótesis podría hacer inteligible su conducta. Por supuesto, si el intérprete sabe que *S* ha sido transportado de planeta, atribuirle una creencia sobre el concepto de *tagua* haría falsas muchas de sus creencias anteriores (*e.g.*, que los océanos de la Tierra están compuestos de XYZ). Por otro lado, asumir que todas sus nuevas creencias son sobre *agua* haría falsas muchas de sus creencias durante su estancia en la Tierra Gemela. En el tercer caso, quizá representaríamos mejor la situación si dijésemos que *S* no tiene ni el concepto de *agua* ni el de *tagua*.

No obstante, ninguna de estas consideraciones determina cuál deba ser la interpretación correcta de sus palabras. En última instancia, parte de lo que decidiría cómo resolver el problema de cómo interpretar a *S* sería informarle de su situación y ver si continúa aplicando la palabra “agua” en ambas sustancias o si intenta discriminarlas. En alguno de estos casos podríamos decir que *S* no sabía,

después de todo, cuáles eran sus pensamientos cuando profería la palabra “agua”. Si opta por lo primero, habríamos de reinterpretar sus creencias previas; si se decide por lo segundo, podemos explicar racionalmente sus errores de aplicación del concepto relevante apelando a su falta de evidencia. En todo caso, la condición para elaborar una interpretación de lo que *S* quiere decir es asumir que sus preferencias son producto de sus creencias acerca de su entorno y de su intención de hacerlas comprensibles para el intérprete; *i.e.*, la interpretación de lo que *S* quiere decir requiere la suposición de que *S sabe lo que significan sus palabras, que sabe cuáles son sus pensamientos*. En algunos casos, podría decirse que en el escenario del cambio prolongado *S* está masivamente equivocado en sus creencias acerca de sus propios pensamientos.²⁷ Esto podría parecernos extraño. Pero no debería sorprendernos; después de todo, el escenario *es* extraño.

El problema de la consecuencia, presentado por McKinsey, requiere un tratamiento distinto. Habíamos señalado que la caracterización de McKinsey de (CPM) como conocimiento *a priori* es problemática (al menos, si se la entiende en términos de acceso privilegiado).²⁸ Si abandonamos esta suposición (como sugerimos que habría de hacerse al final de la sección 1.3), la *reductio* pierde su eficacia. No obstante, es conveniente señalar que la conclusión epistémica de la *reductio* (*i.e.*, que podríamos saber *a priori* qué cosas existen) es resultado, en buena medida, de la tesis que McKinsey no está dispuesto a abandonar: el externalismo. Esta tesis, tal como él la formula, señala que *hay* predicados que implican lógicamente o conceptualmente la existencia de objetos contingentes. Si el externalismo es una tesis que uno pueda conocer *a priori*, de ella se seguiría que podemos saber *a priori* que existen objetos; por ende, los argumentos a favor del

²⁷ Aunque no más que en sus creencias sobre el mundo o sobre lo que piensan otras personas.

²⁸ En la única ocasión en que discute la explicación davidsoniana de la autoridad de la primera persona, McKinsey no muestra que la *reductio* pueda aplicársele. Señala que esto se debe a que la concepción de Davidson del conocimiento de la propia mente es “...demasiado débil para ser de mucho interés filosófico” (1995: 176-178). *Pace* McKinsey, en el capítulo 1 defendimos que la explicación davidsoniana tiene mucho que la recomiende.

externalismo serían ya una respuesta al problema del escepticismo sobre el mundo externo. Ciertamente, esta consecuencia del externalismo es menos radical que la conclusión de la *reductio*: saber *a priori* que existen objetos es distinto de saber *a priori* cuáles son los objetos que existen. Sin embargo, ambas conclusiones deberían ser inquietantes si se consideran como el resultado de investigación *a priori*. Como se verá en el capítulo 3, el externalismo de Davidson, en tanto se considera como resultado de una inferencia a la mejor explicación, hace un balance distinto de la situación.

Hay un sentido importante en el que el externalismo de Davidson se distancia de las formulaciones ortodoxas. Este distanciamiento permite apreciar con más detalle cómo es posible responder a los problemas planteados por Boghossian y McKinsey. Mientras Kripke, Putnam y Burge formulan el externalismo como una reconstrucción teórica del significado de cierto *tipo* de expresiones lingüísticas, Davidson lo considera, en primera instancia, como una elucidación del significado en *casos* de proferencias. La diferencia entre el externalismo de Davidson y las versiones ortodoxas es de carácter metodológico: depende de cuestiones de prioridad explicativa. Mientras el externalismo por el que abogan Kripke, Putnam y Burge parece considerar que el significado es explicativamente más básico que el éxito en la comunicación, Davidson invierte esta ecuación. Considera que ‘significado’ y ‘lenguaje’ son conceptos teóricos que están subordinados a la noción más básica de comunicación exitosa.²⁹ Podría ilustrar parcialmente esta diferencia metodológica por medio de una analogía con las teorías clásica y bayesiana de la probabilidad.

²⁹ En es en este espíritu que Davidson señala: “No hay por qué asombrarse del carácter abstracto del lenguaje. El concepto de un lenguaje es de la misma clase que, y depende de, conceptos como nombre, predicado, oración, referencia, significado y observación. Todos éstos son conceptos teóricos [...]. Queremos estos conceptos cuando hablamos acerca del comportamiento verbal. Los filósofos, psicólogos y lingüistas requieren estos términos teóricos si desean describir, teorizar sobre y explicar actividades verbales. [...] En efecto, hablamos tan libremente acerca del lenguaje, o los lenguajes, que tendemos a olvidar que no hay tales cosas en el mundo; sólo hay personas y sus diversos productos escritos y acústicos. Este punto, obvio en sí mismo, es sin embargo fácil de olvidar, y tiene consecuencias que no son universalmente reconocidas” (Davidson, 1992: 108).

La teoría clásica de la probabilidad comienza por *asignar valores fijos* a los diversos sucesos posibles (*e.g.*, con respecto al lanzamiento de una moneda, asigna $\frac{1}{2}$ a cada uno de los posibles resultados de lanzar la moneda *i.e.*, cara o cruz). Luego, partiendo de estos valores, hace predicciones y genera explicaciones en virtud de este modelo, trabajando bajo el supuesto de que, entre más se aproxime el número de lanzamientos a infinito, tanto más se irá regularizando la distribución de los valores para cada una de las variables, aproximándose a *la asignación original*. Sin embargo, si en un número importante de casos la diferencia con la asignación original es significativa (digamos que de 1'000 lanzamientos, 125 caen cara y 875 en cruz), heurísticamente la teoría clásica de la probabilidad asume una desviación (que la moneda está deforme o mal equilibrada) y *reasigna valores fijos* en función del último resultado en consideración ($\frac{1}{8}$ para cara y $\frac{7}{8}$ para cruz).

Por contraparte, la teoría bayesiana inicia con una *asignación arbitraria* de valores a las diversas posibilidades (aunque, por lo general, comienza con una distribución equitativa) y *ajusta* estos valores en función de los nuevos resultados. Luego hace predicciones y genera explicaciones en virtud de este modelo, trabajando bajo el supuesto de que entre más se aproxime el número de lanzamientos a infinito, tanto más se irá regularizando la distribución de los valores para cada una de las variables, aproximándose a *la distribución real*. De modo que la teoría bayesiana no necesita comprometerse con una distribución fija de probabilidades, ni hacer suposiciones sobre las causas de una desviación significativa con respecto al estimado inicial. Podría hacerlo, pero no lo requiere ni esto amplía su capacidad explicativa.³⁰

De manera análoga al proceder de la teoría clásica de la probabilidad, las formas ortodoxas de externalismo comienzan estipulando significados compartidos por los hablantes y prosiguen intentando fijar la contribución semántica del entor-

³⁰ Agradezco a Ramón Jiménez Lara, su instructiva y amena explicación de las diferencias entre estas dos teorías de la probabilidad. Aunque temo no hacerle justicia en muchos aspectos, me ha parecido excepcional como ilustración de diferencias metodológicas.

no en la determinación del contenido de cierto tipo de expresiones, *i.e.*, fijando la manera en que el entorno determina el significado. Trabajan bajo el supuesto de que si su teoría del significado exhibe una divergencia con situaciones comunicativas típicas, será preciso modificar su teoría de cómo el entorno determina el significado compartido. Por su parte, la metodología de Davidson se asemeja al proceder bayesiano. Como Quine, Davidson no cree que sea teóricamente productivo hablar de significados o lenguaje compartidos por los hablantes; prefiere, en su lugar, hablar de idiolectos.³¹ De este modo, la investigación de cómo el entorno determina cuál es la teoría del significado correcta para un idiolecto podrá variar de hablante a hablante. A veces se considera que adoptar esta estrategia constituye una forma de escepticismo sobre el significado. Por mi parte, creo que sería más apropiado considerar que ambos enfoques (el ortodoxo y el davidsoniano), al hablar de ‘significado’, trabajan con una noción teórica distinta — lo cual no quiere decir, por supuesto, que hablen de fenómenos distintos. Si esto es correcto, elegir entre ambas nociones es una cuestión que se resuelve, como en la práctica científica, comparando los méritos explicativos de ambas teorías al dar cuenta la comunicación humana. Sugiero que el hecho de que no dé origen a problemas como los de Boghossian y McKinsey es una virtud de la noción davidsoniana.

Un mérito adicional del externalismo propugnado por Davidson es que, a diferencia de las formas más populares de externalismo (sea o no éste social), su respaldo no depende principalmente de la consideración de experimentos mentales ni de la apelación a intuiciones.³² Aunque se sirve de este tipo de estrategias expositivas, su valor es como ilustraciones coloridas o en la medida en que permitan

³¹ Un idiolecto es un lenguaje de un único hablante. Esto no quiere decir que se trate de un lenguaje privado; sólo en la medida en que puede ser públicamente interpretado cabe decir que se trata de un lenguaje (véase Davidson, 1975).

³² Kripke ha hecho explícitas sus simpatías metodológicas a este respecto: “Algunos filósofos piensan, por supuesto, que el hecho de que algo tenga contenido intuitivo es una prueba muy poco concluyente a su favor. Yo, por mi parte, pienso que es una prueba muy fuerte a favor de cualquier cosa. Hablando en definitiva, no sé realmente, en cierta forma, qué prueba más concluyente puede uno tener acerca de algo” (Kripke, 1972: 45).

hacer idealizaciones útiles. En esto Davidson, como Quine, considera recomendable minimizar la dependencia, en la argumentación filosófica, de las intuiciones prefilosóficas; defender, por encima de ellas, consideraciones empíricas y de conservadurismo teórico.³³ Es una ventaja metodológica de una tesis filosófica que, aunque no entre en conflicto con la mayoría de nuestras intuiciones, no les dé a éstas peso probatorio.

Como se ha mostrado, el externalismo de Davidson no es incompatible con su explicación de la autoridad de la primera persona. Lo que es más, es el externalismo lo que explica, para Davidson, que pueda presentarse cierta clase de errores en las autoatribuciones de estados mentales. Esto se debe a que "...lo que las palabras de una persona significan depende, en los casos más básicos, de las clases de objetos y sucesos que causaron que la persona considerara que las palabras eran aplicables" (Davidson, 1987: 37). Si la conducta lingüística del hablante no es suficiente para discriminar de manera sistemática los objetos y situaciones en su entorno a los que hace referencia, bien podría decirse que el hablante no sabe lo que piensa. Pero esto no socava la autoridad de sus autoadscripciones de estados mentales, ya que el oyente no se encuentra en una mejor situación para corregirlas, en la medida en que es un intérprete radical. El oyente puede indicarle al hablante que existen ambigüedades en su conducta lingüística con respecto a lo él, como intérprete, reconoce. Pero es al hablante a quien le corresponde suministrar las claves sobre cómo interpretarlo.

En las siguientes secciones examinaremos otro sentido en el cual puede decirse que un hablante no sabe cuáles son sus pensamientos, y por ende se equivoca al adscribirse actitudes proposicionales, *i.e.*, cuando se atribuye falsamente ciertos estados mentales. Consideraremos un caso representativo (aunque no es, por supuesto, el único) de este tipo de fallas en el conocimiento la propia mente: lo que se conoce como autoengaño. En la siguiente sección se presentarán las dificultades

³³ Sigo en esto a Tagle Marroquín (2009).

des que enfrenta la caracterización del autoengaño. En la sección 2.5 se analizará la explicación que Davidson ofrece de éste al interior de su concepción del conocimiento de la propia mente.

2.4. *Autoengaño e irracionalidad*

El autoengaño es un problema filosófico vasto que ha preocupado principalmente a la filosofía moral en lo relativo a la responsabilidad, la vida buena y la autenticidad de un individuo; sin embargo, también ha suscitado interés en los campos de la filosofía de la mente, la epistemología y la filosofía de la psicología. Para los propósitos de la presente discusión se tratará únicamente lo que concierne a estas últimas áreas, dejando de lado los asuntos morales. Específicamente, se enfatizará la relación que el autoengaño guarda con la autoridad de la primera persona. En lo que sigue se esbozarán varias formas de caracterizar el autoengaño, desde algunas de las principales perspectivas teóricas, señalando los problemas que enfrentan. En la sección 2.5 se analizará una posible forma de conciliar este fenómeno con la teoría explicativa de Davidson del conocimiento de la propia mente.

2.4.1. Tres ejemplos de autoengaño

El autoengaño (*self-deception*), junto al pensamiento desiderativo (*wishful thinking*, o ‘hacerse ilusiones’), la debilidad de la voluntad o *akrasia*, y el mal razonamiento, son consideradas habitualmente formas de irracionalidad (véase Davidson, 1997: 218; también 1982a; 1985; 1986). Virtualmente todos los aspectos del autoengaño son materia de controversia en la discusión filosófica actual, incluyendo su definición y los casos paradigmáticos. *E.g.*, Deweese-Boyd (2012), pretendiendo ser neutral frente al tema, señala que, en general, puede decirse que el autoengaño consiste en la adquisición y mantenimiento de una creencia (o, al menos, la admi-

sión de esa creencia) frente a fuerte evidencia de lo opuesto, *motivada por deseos o emociones* que favorecen la adquisición y retención de esa creencia. Sin embargo, la última parte de su definición, como veremos más adelante, favorece cierto enfoque del problema. De momento será conveniente que consideremos algunos ejemplos de lo que podrían considerarse como situaciones de autoengaño:

- (A) Al señor M, que se encuentra realizando sus actividades laborales ordinarias, se le solicita que responda lo que se le describe como “una encuesta rutinaria sobre hábitos relacionados con la salud”. Se le comenta que la encuesta es anónima y se le pide que sea veraz en sus respuestas. M acepta con diligencia y de buena fe. Lee con cuidado las preguntas y trata de responderlas con exactitud. La última de ellas dice lo siguiente: “¿Con cuánta frecuencia consume bebidas alcohólicas?”; en seguida se presentan las siguientes opciones: “(a) Nunca”, “(b) De vez en cuando”, “(c) Frecuentemente”, “(d) Diariamente”. Después de dudarlo por unos instantes, M subraya la respuesta (b). Mientras entrega la encuesta recuerda que es el esperado día de paga y trata de decidir si festejarlo con un buen whisky o terminarse la botella de ron, que ha estado consumiendo (con la discreción que exige la carestía) durante los últimos tres días.
- (B) La joven E recientemente ha decidido hacer un gran cambio en su vida: por primera vez dejará el hogar materno para compartir habitaciones con su pareja sentimental. Como se conocen desde hace muchos años, su relación ha tenido tropiezos. Incluso antes de ser pareja, E y A, quien ahora es su compañero, llevaban una relación cercana: se contaban secretos que habían decidido no compartir con otras personas. En especial, E sabía que A llevaba una vida relajada y había sido infiel con sus parejas sentimentales (en una ocasión, incluso, le sugirió a E que tuvieran un amorío). Pero “Eso es el pasado”, piensa E. Tras un mes de vivir juntos, E, que había salido a visitar a su familia, llega por la mañana a su casa y descubre, horrorizada, que su compañero yace en su lecho junto a una mujer desnuda. Tras una semana de copioso llanto, decide intentar ‘arreglar las cosas’. Aunque está herida, decide volver con su pareja con la convicción de que “No lo volverá a hacer”. En los meses siguientes, cuando E tiene que ausentarse lo llama con frecuencia y procura no salir demasiado.

(C) Una mañana X descubre alarmado un mechón de cabellos sobre su almohada. No es la primera vez que sucede; pero no deja de ser inquietante. “Debe ser el estrés”, se dice para tranquilizarse. Cuando comenta la situación con uno de sus amigos, éste le responde socarronamente: “Te estás quedando calvo”. X replica con incredulidad y no sin cierta ironía: “¿Cómo crees?, si tengo más pelo que tú”. Cuando va al supermercado, antes de tomar su champú habitual, lee con cuidado la etiqueta de otra botella, la destapa y huele su contenido. Bajo el nombre de la marca, la etiqueta indica “Tratamiento contra la alopecia”. Tras deliberarlo unos segundos, decide que, pese a que es un poco más caro, comprará el nuevo champú.

Las situaciones (A), (B) y (C) describen a personas que parecen tener creencias conflictivas. M conoce sus hábitos de consumo de bebidas embriagantes, pero expresándose con sinceridad diría que sólo bebe un trago de vez en cuando. E conoce las costumbres licenciosas de su pareja, incluso cuando está con ella; sin embargo, diría que “él no es esa clase de persona”. X se mira al espejo diariamente, pero no se incluye entre las víctimas de la calvicie. ¿Saben estas personas lo que piensan?

El interés filosófico en el autoengaño radica, entre otras cosas, en que supone una suerte de aberración epistemológica: un individuo racional adquiere y mantiene una creencia que se opone al total de la evidencia que posee. En ocasiones esto se debe a motivaciones extrarracionales: deseos, temores, inquietudes, etc. “En el lenguaje de la psicología pop[ular], somos incapaces de ‘reconocer’ las cosas que son ‘demasiado dolorosas para lidiar con ellas’ y ‘entramos en negación’. Algunas veces ocurre exactamente lo opuesto: no podemos dejar de pensar en una terrible posibilidad y, como resultado, nos convencemos de que es real” (Bach, 2009: 781). Aunque éstas pueden constituir motivaciones para adquirir o sostener una creencia, no son razones para creer que es el caso.³⁴

³⁴ “Si uno sería más feliz, estaría más orgulloso, más relajado, menos tenso si uno tuviese una creencia específica, ésa es una razón, poniendo de lado otras consideraciones, para tener la creencia. Pero tal razón no es, en sí misma, una razón para suponer que la creencia deseable es verdadera” (Davidson, 1997: 216).

2.4.2. Modelos intencionales y motivacionales del autoengaño

La forma tradicional en que se ha caracterizado el autoengaño es modelándolo sobre la base del engaño interpersonal (véase Barnes, 1997: 18-21; Bach, 2009; y Deweese-Boyd, 2012). Por lo común, el engaño interpersonal se caracteriza de esta manera: un sujeto *A* *intencionalmente* hace creer a un sujeto *B* la proposición *p*, creyendo que su contraria ($\neg p$) es verdadera o dudando de la verdad de *p*. Se considera, pues, un tipo de conducta intencional. No siempre que alguien nos proporciona información falsa somos víctimas de engaño; a veces se trata de ‘un error honesto’. Análogamente, no todas nuestras creencias conflictivas son producto del autoengaño; a veces la evidencia parece inconsistente. Caracterizar al engaño (sea interpersonal o intrapersonal) como una conducta intencional permite distinguirlo del mero error. Llamemos a éste el modelo intencionalista.

Alfred Mele (1997) señala que si concebimos al autoengaño bajo el modelo intencionalista, siguiendo la analogía con el engaño interpersonal, enfrentaremos dos paradojas: la paradoja ‘estática’ y la ‘dinámica’. La primera no es otra que la paradoja de Moore: afirmar una proposición y decir enseguida que uno no se la cree; esto, como señala Mele (1997), requiere que el autoengañado se encuentre en lo que parece un estado imposible de la mente: afirmar la contradicción. La segunda paradoja se refiere a la forma misma en que se modela el autoengaño intencional: hacer algo intencionalmente implica hacerlo a sabiendas; de modo que cuando uno se engaña a sí mismo lo hace sabiendo que quiere engañarse; pero esto supondría que uno sabe de antemano que lo que se pretende creer es falso. En el engaño interpersonal esto sería como saber que un informante *siempre* nos proporciona un falso testimonio y, a pesar de ello, *per impossibile* creer que su información es veraz. Debido a estas paradojas, algunos autores³⁵ han optado por una aproximación no-intencionalista al autoengaño. Asumen que lo único que interviene en casos auténticos de autoengaño son motivaciones (deseos, aspiraciones,

³⁵ Entre los que habría que contar a Mele (1997), Barnes (1997) y Funkhouser (2006).

temores, inquietudes, etcétera). De este modo, su estrategia es deflacionista: el autoengaño es, para ellos, sólo una forma de error motivado por uno mismo; sus manifestaciones toman la forma de otros casos de irracionalidad, mas no de literal autoengaño. Podemos denominar a éste el modelo motivacionalista.

Los ejemplos que planteamos arriba pueden ser ilustrativos del tipo de explicación que ofrecería quien adopta el modelo motivacionalista. Podría ser que M, debido las distracciones de su trabajo (y al alcohol, quizá), haya perdido la noción de cuán frecuentemente se embriaga, haciéndolo un caso de mal razonamiento. Aunque E sabe que si sigue con A le esperan situaciones ‘incómodas’, su profundo amor la lleva a actuar contra su mejor juicio, siendo víctima de debilidad de la voluntad. X desearía tener más cabello, por lo que no ha notado lo despoblada que está su coronilla ni cuán pronunciadas son sus entradas; tiene pensamiento desiderativo. Hasta aquí un teórico motivacionalista estaría de acuerdo en que los ejemplos anteriores representan un caso de lo que confusamente se denomina autoengaño. Más aún, diría que lo anterior *es todo lo que se requiere* para que los ejemplos constituyan casos de autoengaño. Puede considerarse que lo que la estrategia motivacionalista sugiere es, en parte, renunciar a explicar el autoengaño; en su lugar, se propone echar luz sobre otras formas de irracionalidad. Bajo esta perspectiva, aunque el motivacionalista reconocería que el autoengaño es como lo describe el intencionalista, negaría que haya casos de auténtico autoengaño.

En ocasiones, incluso parecería que el modelo motivacionalista intenta echar luz sobre los casos putativos de autoengaño apelando a causas fisiológicas de los sujetos involucrados. Se sugiere, *e.g.*, que la aparente inconsistencia en las creencias de un individuo puede ser similar a la condición de algunos pacientes que sufren de afasia (*i.e.*, la pérdida de la capacidad para producir o comprender preferencias lingüísticas, ocasionada por daño cerebral). En algunos casos, estos pacientes desarrollan una curiosa disposición para sobrellevar situaciones que el resto de nosotros encontraría perturbadoras: responden con risas a indicios de fingimiento o

deshonestidad. Se ha sugerido que mecanismos similares podrían explicar las conductas de un sujeto que resultan inconsistentes con sus creencias (véase, *e.g.*, Ramachandran, 1996). Al hacer esto, el modelo motivacionista pierde de vista un elemento importante: hace del autoengaño algo no-racional, suprimiendo su carácter irracional, que parece ser uno de sus elementos esenciales: “[p]ues lo irracional no es meramente lo no racional, lo que reside fuera del ámbito de lo racional; la irracionalidad es una falla al interior de la casa de la razón” (Davidson, 1982a: 169). En los ejemplos anteriores que podrían multiplicarse se presentan individuos que normalmente interpretaríamos como sosteniendo creencias conflictivas. M recuerda que ha estado tomando y cree que, si su bolsillo se lo permitiera, habría bebido más; pero confiesa que sólo bebe “De vez en cuando”. E declara que no cree que A vuelva a engañarla (pues, dice, “No es esa clase de persona”); pero toma precauciones para conocer las actividades de su pareja. X compra lo que identifica como champú contra la calvicie (aunque sabe que es más costoso), pero le dice a su amigo “¿Cómo crees [que me estoy quedando calvo]?” Las creencias contradictorias son lo que el teórico motivacionista se proponía evitar pues su estrategia está diseñada precisamente para eliminar las paradojas, pero en estos casos adoptar esta estrategia nos invitarían a dudar de la sinceridad de nuestros personajes. El engaño está ahí, la cuestión es: ¿a quién va dirigido? No parecería razonable asumir que, en este tipo de situaciones, nuestros interlocutores siempre nos engañan. A menudo preferimos, más bien, describir su conducta como un caso de irracionalidad que involucra un conflicto entre creencias contradictorias. Pero, si éste fuera el caso y no deseáramos atribuirles una creencia absurda (*i.e.*, la creencia en la contradicción), parecería que debemos decir que están equivocados acerca de cuáles son sus pensamientos.

Algunos partidarios de la línea intencionalista han señalado que se podrían admitir creencias paradójicas si, *e.g.*, consideramos que hay divisiones temporales en la mente que aíslan al autoengañado de su intención de engañarse (Bermúdez,

2000). Esto supondría que, en un sentido importante, quien se autoengaña se asemeja a alguien que padece amnesia anterógrada (que no puede generar recuerdos a corto plazo). Si uno decidiera intencionalmente ocultar la evidencia que respalda sus creencias podría, confiando en sus fallas de memoria, esperar que en el futuro se generen las creencias opuestas. Esto desplazaría el problema a la memoria. Si el autoengaño se restringiera a estos casos, no afectaría a la autoridad de la primera persona (pues la hemos restringido a las actitudes proposicionales contemporáneas a su atribución por parte de un hablante). Pero no parece que en todos los casos de autoengaño (ni siquiera en la mayoría) debamos responsabilizar a la memoria de estas fallas de racionalidad. En ocasiones los hablantes se adscriben creencias que resultan inconsistentes con las que les atribuiríamos en ese mismo momento para explicar sus acciones. M sabe de su alcoholismo, por eso destina parte de sus recursos ordinarios a la compra de bebidas embriagantes; pero confiesa sinceramente lo contrario. E desconfía de su pareja; pero es honesta (e ingenua) cuando afirma que A “no lo volverá a hacer”. X busca prevenir la pérdida de su cabellera; pero es franco al decir que no cree que esté quedándose calvo. Si estas situaciones ocurren, ¿en qué sentido se preserva la presunción de que un hablante sabe lo que cree cuando dice que cree algo, la presunción de que su autoadscripción es verdadera?

2.5. Autoengaño y autoridad de la primera persona

Davidson introduce la consideración del autoengaño en relación con su teoría explicativa del conocimiento de la propia mente en varios de sus ensayos, contemporáneos a sus discusiones sobre la autoridad de la primera persona (véase Davidson, 1982a; 1985; 1986 y 1997). En la propuesta davidsoniana, si hemos de explicar los casos de auténtico autoengaño y disipar la supuesta amenaza que presentan para la autoridad de la primera persona, debemos asumir la caracterización de este fenómeno bajo el modelo intencionalista. Esto significa que es preciso

considerar situaciones en las que ocurren creencias conflictivas de manera simultánea. La intención de Davidson es mostrar que el autoengaño es compatible con la autoridad de la primera persona, al considerarlos en el marco más general de una teoría general del pensamiento, el significado y la acción (cuyo esbozo se presenta en su 1980).

2.5.1. Autoengaño y racionalidad

Davidson parte del carácter normativo de la racionalidad. Al indicar que la racionalidad es normativa, se sugiere que en la interpretación de un sujeto debemos procurar hacer su conducta inteligible (no necesariamente inteligente), a partir de nuestros propios estándares. Aunque quizá no pueden describirse exhaustivamente el tipo de consideraciones que intervienen en determinar que un sujeto es racional, debe haber, como mínimo, algunos estándares compartidos. De este modo, ser racional implica (entre otras cosas) seguir las normas básicas de la lógica, el principio de la evidencia total para el razonamiento inductivo y el principio de continencia. Estos principios son compartidos por todos los individuos que tienen actitudes proposicionales o actúan intencionalmente, puesto que es una condición para tener pensamientos, juicios e intenciones que los estándares básicos de la racionalidad tengan aplicación (Davidson, 1985: 195).³⁶ No tiene sentido preguntar, de alguien con actitudes proposicionales, si es generalmente racional, pues esto es lo que se requiere para tenerlas; asimismo, los sujetos no pueden *decidir* si aceptan o no los atributos fundamentales de la racionalidad: si están en

³⁶ Debido precisamente a su concepción del vínculo entre pensamiento, lenguaje y acción aunado al carácter normativo de la racionalidad, Davidson señaló en alguna ocasión: “Lo que he repetido a menudo es que los niños, antes de tener gobierno sobre su habla, o los animales, no tienen de ninguna manera pensamientos. Esto parece molestar a personas que tienen mascotas, ¡y supongo que debe molestar a gente que tiene niños!” (Davidson, 1992: 78). Se exploran algunas de las consecuencias de esta concepción de la racionalidad para el problema escéptico en el capítulo 3.

condiciones de decidir en absoluto, poseen tales atributos.³⁷ Esto expresa una condición de individuación de las actitudes proposicionales que hasta ahora hemos dejado implícita: cuáles son los contenidos de los pensamientos de un sujeto está determinado, no sólo por los objetos en el entorno del sujeto (como sugiere el externalismo), sino también por las relaciones de interdependencia que un estado mental guarda con otros, preservando la racionalidad. Esto se relaciona, para Davidson, con la primacía de la inteligibilidad para la comunicación:

Las relaciones entre creencias juegan un papel constitutivo decisivo; un intérprete no puede aceptar desviaciones grandes u obvias de sus propios estándares de racionalidad sin destruir el fundamento de la inteligibilidad sobre el que descansa la interpretación. La posibilidad de comprender el habla o las acciones de un agente depende de la existencia de un patrón fundamentalmente racional, un patrón que debe, en líneas generales, ser compartido por todas las criaturas racionales. No tenemos más elección, entonces, que proyectar nuestra propia lógica sobre el lenguaje y las creencias de otro. Esto significa que es una restricción sobre las interpretaciones posibles de las oraciones que se consideran como verdaderas que sean (al interior de la razón) lógicamente consistentes unas con otras (Davidson, 1990: 63).

Aunada a la restricción externalista para la determinación del contenido mental, podemos añadir ahora una condición holista, *i.e.*, una condición sobre las interrelaciones (de consecuencia lógica, de apoyo evidencial, de racionalidad práctica) entre los diversos estados mentales del individuo. Aunque lo que nos interesa es

³⁷ A menudo se ha criticado que la caracterización davidsoniana de la racionalidad impone estándares que ningún hablante real satisface. Creo que esta crítica es inapropiada, pues ignora que las condiciones para considerar a un hablante como racional están determinadas, en última instancia, por la situación del intérprete: “No hay una definición de ‘racionalidad perfecta’, sin importar lo que consideremos que esto sea. [...] Podemos explicar y comprender la irracionalidad sólo frente a un trasfondo de racionalidad, *un trasfondo que cada uno de nosotros debe, como intérprete, proporcionar por sí mismo*. No hay una lista fija de estándares, no hay jerarquía fija de valores, a los cuales debamos todos suscribirnos, pero algunas normas son tan básicas para la inteligibilidad que no podemos evitar conformar nuestros pensamientos a sus patrones. [...] Yo defendería que no podríamos comprender a alguien a quien fuésemos forzados a tratar como distanciándose radicalmente y en forma predominante de todas esas normas. Esto no sería un ejemplo de irracionalidad, o un conjunto extraño de estándares; sería una ausencia de racionalidad, algo que no podría ser considerado como pensamiento” (Davidson, 1993: 169-170).

una explicación de los casos de autoengaño intencional contemporáneo, pues son los que amenazan la autoridad de la primera persona, es conveniente que digamos algo sobre el error motivado y las creencias inconsistentes en momentos diversos.

Para explicar el rol causal de la motivación en el autoengaño, Davidson hace, en primer lugar, algunas precisiones sobre la explicación de la acción. Aunque los sucesos mentales en alguna forma *causan* otros sucesos (sean mentales, como las creencias, o físicos, como las acciones), para la explicación de la acción los consideramos, más bien, como *razones* del sujeto. Todo lo que el sujeto *hace* lo hace por una razón, a cuya luz la acción correspondiente es razonable; las razones que el sujeto da de su acción ofrecen la intención con la que actuó, y por tanto la explican. Esa explicación debe existir si lo que hace una persona debe contar como una acción (véase Davidson, 1982a: 173). Sin embargo, hay una manera en que los procesos mentales pueden *causar*, pero no *racionalizar* la conducta: se trata de los casos que el motivacionista considera prototípicos del autoengaño. En tales casos hay una causa mental (*e.g.*, una emoción) de la conducta del sujeto que, sin embargo, no es una razón para el sujeto; si dicho estado mental carece de contenido proposicional no puede intervenir en el proceso de dar razones que justifiquen o hagan apropiada la conducta del sujeto.

Asimismo, Davidson reconoce que en algunos casos de autoengaño la división temporal es un factor explicativo importante (1985: 198); pero ni éstos ni los anteriores son los casos de autoengaño más interesantes, sino aquellos en los que la inconsistencia intrínseca es un factor determinante: cuando coexisten dos creencias contradictorias. Como indicábamos arriba, al dar cuenta de estas situaciones hay que tener en cuenta el carácter holístico de lo mental; *i.e.*, que el significado de las preferencias del hablante, el contenido de sus creencias o deseos, no son elementos que puedan ser atados a ellas en aislamiento de sus compañeros (otras preferencias; otras creencias o deseos). De ahí que estos tipos de autoengaño resulten problemáticos.

2.5.2. La mente dividida

En este punto Davidson hace explícita una distinción. Puede decirse, en primer lugar, que un sujeto tiene, simultáneamente, creencias contradictorias (que S cree que p y S cree que $\neg p$); lo cual parece describir la situación de quien se autoengaña. Pero también puede decirse, en segundo lugar, que un sujeto cree, en un momento dado, una contradicción (que S cree que p y $\neg p$); lo cual resulta paradójico y entra en conflicto con la restricción del holismo de lo mental. Si lo que ocurre en casos de autoengaño es lo primero, lo que se requiere, para evitar la paradoja estática de Mele, es impedir que de ahí se siga lo segundo. El problema que esto suscita es cómo pueden las creencias conflictivas coexistir en un mismo sujeto sin dar lugar a la creencia en la conjunción. Aquí Davidson reconoce, adquiriendo una deuda teórica con Freud, que un recurso explicativo que podría funcionar para esta situación es la partición de la mente (1997: 217). Sin embargo, a diferencia de Freud, Davidson no se vale del carácter metafórico de esta afirmación: las divisiones de la mente pueden definirse en términos funcionales; en última instancia, en términos de los conceptos de razón y de causa (1982a: 185). Si las creencias contrarias se presentan, ha de ser en partes distintas de la mente, de modo que el sujeto pueda tener ambas creencias sin creer en la contradicción.

Ahora bien, si se adopta la hipótesis de la partición de la mente como parte del fenómeno del autoengaño esto parece restarle peso intuitivo a la autoridad de la primera persona, pues ésta consiste en la presunción, esencial para la posibilidad de interpretación y, por ende, para la comunicación misma, de que el hablante es interpretable, *i.e.*, de que sus creencias e intenciones causan y racionalizan su conducta lingüística. No obstante, si la mente del hablante está dividida, de modo que puede actuar a la luz de ciertas creencias y adscribirse creencias contradictorias con aquéllas, tal presunción parece desorientada. De modo que la concepción de Davidson del autoengaño parece entrar en conflicto con la autoridad que él mismo reconoce.

Sin embargo, el autoengaño no debe socavar el fundamento de la autoridad de la primera persona: su incompatibilidad es sólo de grado. Evidentemente, cuando uno descubre en un tercero un caso auténtico de autoengaño (*i.e.*, que se adscribe ciertas creencias, pero su conducta puede ser descrita mejor bajo la suposición de que sostiene creencias opuestas), la presunción *prima facie* de que sabe cuáles son sus pensamientos, de que sus autoadscripciones son correctas, es puesta en entredicho, mas no desaparece. Cuando alguien sostiene creencias contradictorias, un tercero puede corregir sus autoadscripciones, al señalarle la inconsistencia;³⁸ pero esto no quiere decir que el intérprete se encuentre en mejores condiciones para determinar cuáles son, en efecto, las creencias del hablante o cómo determinar el significado de sus palabras. En los casos de autoengaño, como intérpretes, no podemos fijar lo que el hablante piensa o quiere decir, al menos no hasta que la autoridad del hablante sea recuperada, cuando éste se percata de la inconsistencia y el autoengaño se disipa.

2.6. *El conocimiento de la propia mente, la racionalidad y el proyecto epistemológico*

En este capítulo hemos defendido una más de las razones por las que tiene sentido hablar de *conocimiento* de la propia mente. En la reconstrucción davidsoniana, este conocimiento constituye realmente un *logro* cognoscitivo. Como hemos visto, se puede fallar en este logro: un hablante puede no saber lo que significan sus palabras (y, por ende, cuál es el contenido de sus pensamientos), puesto que esto se encuentra en parte determinado por factores del entorno que el hablante puede no seleccionar sistemáticamente al tratar de hacerse interpretable; asimismo, es posible que el hablante, en un momento dado, no sea interpretable debido a la

³⁸ Por supuesto, esto no excluye la posibilidad de autoengaño compartido, de una *folie a deux*, de que el intérprete no sólo sea incapaz de ver la inconsistencia del sujeto sino que la comparta.

inconsistencia entre sus autoadscripciones y las hipótesis sobre sus estados mentales que harían de su conducta una forma de acción racional.

Además, hemos visto razones para considerar que el conocimiento de la propia mente exhibe rasgos que lo hacen *especial* frente a otros logros cognitivos. En el capítulo 1 se argumentó que el conocimiento de la propia mente es especial en, al menos, dos sentidos importantes. Por una parte, las autoadscripciones de estados mentales por parte de un hablante conllevan una presunción de verdad, que hace que, cuando son verdaderas, cuenten como conocimiento; no hay tal presunción en el caso de nuestras creencias verdaderas sobre el mundo externo: aunque sean verdaderas, en ocasiones no constituyen conocimiento. Llamamos a este rasgo la autoridad de la primera persona. Por otra, el conocimiento de la propia mente, en la mayoría de los casos, no requiere el tipo de observación de la conducta e inferencia que empleamos para conocer los pensamientos de otras personas. Denominamos a este rasgo la asimetría epistémica. Davidson ofreció una teoría explicativa de estos aspectos del conocimiento de la propia mente que, dadas las condiciones que le impusimos, nos pareció satisfactoria; se apela en ella a las condiciones de la interpretación. Al echar luz sobre la autoridad de la primera persona y la asimetría epistémica al interior de una teoría de la interpretación podemos dar cuenta de su estatus epistémico.

En el próximo capítulo intentaremos clarificar la situación del conocimiento de la propia mente dentro de la empresa cognitiva en general: trataremos de situarlo al interior del proyecto epistemológico. Como se señaló en la sección 1.3.1, buena parte de una tradición filosófica, de la cual Descartes es un exponente ilustre, situó al conocimiento de la propia mente en el núcleo del proyecto epistemológico, sea buscando hacerlo el fundamento del resto del conocimiento o reducir el resto de nuestros logros cognoscitivos al conocimiento de nuestro mundo interno. Este proyecto buscaba responder a inquietantes fundamentales, planteadas a menudo como retos escépticos. En el capítulo 3 se intentará precisar en qué medida la

concepción davidsoniana del conocimiento de la propia mente, en tanto abandona parte importante de los supuestos que subyacían a la concepción cartesiana, está en condiciones de responder a (o de reformular) estos retos. Nos centraremos, en particular, en los problemas escépticos que gravitan en torno al mundo externo, las otras mentes y a la idea misma de la racionalidad.

Por otra parte, como reacción al fracaso del proyecto epistemológico cartesiano, una contra-tradición ha sugerido que abandonemos por completo la idea de que tenemos conocimiento de nuestra propia mente. Se ha insinuado con ello una nueva forma de escepticismo, esta vez sobre la propia mente. Aunque ya se ofrecieron razones, en la sección 1.4, para evitar este tipo de reacciones, una tarea pendiente es la de clarificar en qué medida las otras formas de conocimiento, y la racionalidad misma, dependen del conocimiento de la propia mente.

En el proceso de atender estas cuestiones se verá también que, pese a tener continuidad con las interrogantes tradicionales, la filosofía de Davidson introduce herramientas y consideraciones revolucionarias. Como él mismo lo expresara:

[Hay quienes...] ve[n] a la historia de la filosofía como una batalla confusa y sin vencedores entre un escepticismo ininteligible y pobres intentos por responderle. La epistemología desde Descartes hasta Quine me parece sólo un capítulo complejo, aunque de ninguna manera poco iluminador, en la empresa filosófica. Si este capítulo está llegando a su fin, será mediante el recurso a modos de análisis y la adherencia a estándares de claridad que siempre han distinguido a la mejor filosofía, y que, con suerte y empeño, seguirán haciéndolo. (Davidson, 1987a: 157).

Capítulo 3

El lugar del conocimiento de la propia mente en la epistemología

That knowledge of the contents of my own mind is special, and basic to all my knowledge, is, of course, part of the Cartesian and empiricist dreams. And this much is correct: such knowledge is basic in the sense that without it I would know nothing (though self-knowledge is not sufficient for the rest), and special in that it is irreducibly different from other sorts of knowledge.

(Davidson, 1998: 87)

En los capítulos anteriores se ha examinado la concepción davidsoniana del conocimiento de la propia mente, confrontándola con teorías explicativas rivales y exigiendo que responda a ciertos retos explicativos. En el capítulo 1 se argumentó que la caracterización que ofrece Donald Davidson del conocimiento de la propia mente satisface las demandas de una teoría explicativa; en especial, da cuenta de los rasgos de autoridad de la primera persona y asimetría epistémica de manera más apropiada que otros enfoques. En el capítulo 2 se ofrecieron razones para considerar que, en la caracterización davidsoniana, el conocimiento de la propia mente cumple con un requisito intuitivo adicional, de carácter conceptual, de los logros epistémicos: es susceptible de error; además, se intentó mostrar que, lejos de ser inconsistente con tesis filosóficas como el externalismo y con fenómenos como el autoengaño, la concepción davidsoniana dispone de recursos para dar cuenta de las fallas en el conocimiento de la propia mente, sin socavar con ello el estatuto epistémico de la autoridad de la primera persona.

En este capítulo se confrontará cierta imagen de la epistemología y del papel que en ésta desempeña el conocimiento de la propia mente con la concepción hasta aquí defendida de esta forma de conocimiento. Asimismo, se reconsiderará su lugar al interior de la empresa epistemológica. Con este propósito, en la sección 3.1 se hace un esbozo de lo que podrían considerarse las tareas centrales de la epistemología como disciplina filosófica, intentando preservar neutralidad frente a la cuestión de cómo haya de efectuarse la investigación filosófica del conocimiento. En la sección 3.2 se examina una forma históricamente influyente, que aquí se denomina ‘epistemología cartesiana’, de emprender este tipo de investigación y se analizan los supuestos en torno al conocimiento de la propia mente que permean esta tradición. En la sección 3.3 se plantean tres de los principales retos, encarnados en tres formas de escepticismo, que enfrenta la epistemología cartesiana, con miras a consumir sus propios objetivos teóricos. La sección 3.4 reconsidera los problemas y objetivos de la epistemología desde otra perspectiva teórica y metodológica: el naturalismo filosófico. En la sección 3.5 se examina el papel del conocimiento de la propia mente en la epistemología davidsoniana, considerada como una forma de epistemología naturalizada.

3.1. La epistemología como proyecto filosófico

Desde los orígenes mismos de la reflexión filosófica, en la Grecia antigua, el conocimiento ha sido constante objeto de escrutinio. No obstante, su importancia alcanzó una magnitud sin precedentes en la época moderna, durante los siglos XVII y XVIII. Fue en esta época que la preocupación filosófica por el conocimiento alcanzó autoconciencia y se constituyó como una disciplina por derecho propio: la epistemología o teoría del conocimiento. Sus problemas centrales gravitan en torno a la naturaleza, posibilidad, justificación, alcance y bases generales del conocimiento.

3.1.1. Las motivaciones de una teoría del conocimiento

¿Por qué, y en qué sentido, el conocimiento parece merecer, e incluso requerir, investigación y escrutinio filosóficos? La siguiente podría ser una respuesta atractiva:

...la preocupación por el conocimiento [...] está [...] profundamente incrustada en nuestra tradición intelectual occidental [...] Esta tradición, que en sus aspectos filosófico y científico tiene sus orígenes en la Grecia clásica, es en su totalidad y en el más amplio sentido una tradición *racionalista* y crítica. La ciencia y la filosofía comienzan cuando las ideas acerca del origen y la naturaleza del universo son tratadas como teorías a favor de las cuales debe argumentarse: esto es, comparadas con (y tal vez sustituidas por) teorías rivales. [...]Esta aproximación ampliamente racionalista para la comprensión del mundo puede pensarse como una clase de tradición de ‘segundo orden’: lo que se transmite no son creencias particulares [...] sino una práctica de examinar críticamente las ideas actuales con el fin de preservar sólo aquellas que logren aprobar la inspección. [...]

La tradición racionalista puede aplicarse a sí misma. Cuando así sucede, obtenemos la epistemología: una tradición meta-crítica, de tercer orden, de reflexión sobre nuestras metas y procedimientos epistémicos (Williams, 2001: 4-5).

Nuestra tradición intelectual le concede una importancia primordial al conocimiento. Esto, se ha sugerido, no es un mero accidente de nuestra cultura: se encuentra incrustado en la naturaleza misma de la racionalidad humana (véase, *e.g.*, Zagzebski, 2009: 9-11). En la medida en que algo nos importa, nos preocupamos por adquirir creencias verdaderas acerca de ello. Esto se debe, entre otras cosas, a que a menudo debemos actuar en los dominios de lo que nos importa y queremos creencias que puedan servirnos como bases para la acción. Sin embargo, si hemos de actuar racionalmente, esto requiere que tengamos no sólo creencias verdaderas, sino confianza en que las creencias particulares involucradas en nuestras intenciones para actuar sean verdaderas. Puesto que actuar involucra esfuerzo, y en ocasiones riesgo o sacrificio, no es racional actuar sin cierto grado de confianza en la verdad

de las creencias sobre las que basamos nuestra acción. De modo que, en tanto algo nos importa y queremos actuar racionalmente en los dominios de lo que nos importa, nos preocupamos por tener creencias verdaderas así como razones a favor de su verdad, *i.e.*, nos interesa tener conocimiento.

A menudo tenemos creencias falsas y, debido a que no queremos tener creencias falsas sobre las cuestiones que nos importan, buscamos mecanismos que nos permitan distinguir las creencias falsas de las verdaderas. Puesto que el valor para la acción de una creencia verdadera disminuye si se encuentra mezclada con muchas creencias falsas, sometemos a escrutinio nuestras creencias. Es por ello que, además de obtener creencias verdaderas, nos importa examinar los procedimientos que empleamos para generar y sostener dichas creencias. Aunque estas preocupaciones surgen al interior de nuestras prácticas ordinarias, se extienden de manera natural a empresas intelectuales más sofisticadas, como la filosofía y la ciencia. En la medida en que el examen de nuestras creencias que efectuamos en estos ámbitos es un tipo de acción en el que estamos involucrados, queremos también tener creencias verdaderas sobre este dominio; *i.e.*, nos importa examinar nuestros procedimientos epistémicos.

La filosofía, como búsqueda de comprensión reflexiva, se ocupa de problemas intelectuales que surgen en el esfuerzo por comprender, en aquellos lugares en los que se producen tensiones o perplejidades en nuestro pensar. A diferencia de la ciencia empírica,¹ la filosofía adopta un enfoque más general, más abstracto y orientado de forma distintiva a esta clase de tensiones o perplejidades. Al parecer,

¹ Decir que hay una diferencia entre la filosofía y la ciencia no equivale a señalar que se trata de empresas intelectuales radicalmente distintas. Puede tratarse de una diferencia de grado y no de naturaleza. Otro tanto vale para la distinción entre la ciencia y el sentido común, como defiende, *e.g.*, Quine: “La ciencia no es un sustituto del sentido común, sino una extensión de éste. La búsqueda de conocimiento es propiamente un esfuerzo simplemente por ampliar y profundizar el conocimiento del que ya disfruta el hombre de la calle, con moderación, en relación con las cosas comunes a su alrededor. Repudiar el núcleo mismo del sentido común, requerir evidencia para aquello que tanto el físico como el hombre de la calle aceptan como obviedades, no es un perfeccionismo laudable; es una pomposa confusión, es un fracaso al observar la fina distinción entre el bebé y el agua en la que se baña” (1957: 216-217).

es la presencia de estos rasgos lo que vuelve a un tema particular objeto de preocupación por parte de los filósofos. Respecto al conocimiento estas tensiones resultan especialmente complicadas, intrigantes y perturbadoras en sus implicaciones.

3.1.2. Cinco problemas de la epistemología

Con el fin de clarificar la discusión, es posible agrupar las principales perplejidades filosóficas en torno al conocimiento en problemas, o grupos de problemas. Michael Williams sugiere la siguiente clasificación que, por su perspicuidad, me permito citar *in extensu*:

1. *El problema analítico.* ¿Qué es el conocimiento? (O, si lo preferimos, ¿qué queremos decir o deberíamos querer decir² con ‘conocimiento’ o ‘saber’?). Por ejemplo, ¿cómo se distingue (o debería distinguirse) el conocimiento de la mera creencia u opinión? Lo que queremos aquí, idealmente, es una elucidación o *análisis* preciso del *concepto* de conocimiento.

2. *El problema de la demarcación.* Éste se divide en dos sub-problemas. (a) El problema *externo*: [...] ¿podemos determinar de manera fundamentada acerca de qué clase de cosas podría esperarse razonablemente que tuviéramos conocimiento? [...], ¿podemos determinar el alcance y los límites del conocimiento humano?, ¿residen

² No habría que asumir que la pregunta “¿Qué es el conocimiento?” equivale a interrogarse por el concepto de ‘conocimiento’. Al preguntarse lo primero, los filósofos a veces están interesados en el concepto, pero en otras ocasiones se interesan por la naturaleza del fenómeno o propiedad. Supóngase que nos preguntamos “¿Qué es el agua?” Podríamos querer comprender el concepto de ‘agua’, cuál es el significado de la palabra “agua” cuando alguien la emplea. Por otra parte, podríamos querer saber acerca de la naturaleza de la propiedad de ser agua, los hechos relevantes acerca del agua, *e.g.*, que es un compuesto formado por dos átomos de hidrógeno y uno de oxígeno. Estos proyectos no necesariamente son excluyentes: mi concepto de agua presumiblemente captura muchos hechos importantes acerca del agua, *e.g.*, que se encuentra en estado líquido a temperaturas superiores a 0°C e inferiores a 100°C, a nivel del mar. Aún así, podría captar el concepto de agua sin saber todos los hechos acerca de su naturaleza. En el caso del conocimiento, la relación entre teorías del concepto y teorías del fenómeno o propiedad es más complicada. A diferencia del caso del agua, no tenemos un acceso empírico independiente a los hechos acerca del conocimiento excepto a través de ese concepto. Así, las disputas sobre el fenómeno o propiedad se efectúan frecuentemente (pero no siempre) en terreno conceptual, sobre cómo podríamos definir mejor el concepto de ‘conocimiento’.

algunas áreas dentro de la provincia del conocimiento mientras otras están destinadas a permanecer en la provincia de la opinión (o de la fe)? [...] El objetivo aquí es trazar una frontera que separe la provincia del conocimiento de otros dominios cognitivos (o quizá lo cognitivo de lo no cognitivo). (b) El problema *interno* se pregunta si hay fronteras importantes *dentro* de la provincia del conocimiento. Por ejemplo, muchos filósofos han sostenido que hay una distinción fundamental entre el conocimiento que es *a posteriori* o ‘empírico’ y el que es *a priori* o ‘no empírico’. [...] No obstante, otros filósofos han negado que tal distinción pueda trazarse.

3. *El problema del método.* Éste tiene que ver con cómo ha de obtenerse o buscarse el conocimiento. [...]Distingamos tres sub-problemas. (a) El problema de la *unidad* [...]: ¿hay sólo una manera de adquirir conocimiento, o hay varias, dependiendo de la clase de conocimiento en cuestión? [...] (b) El problema de la *optimización* [...]: ¿podemos mejorar nuestras maneras de buscar conocimiento? [...] (c) Por último, si bien no menos importante, se encuentra el problema de la *razón* o *racionalidad*. La preocupación aquí es si hay métodos de investigación, o de fijación de la creencia, que sean distintivamente racionales y, de ser así, cuáles son.

4. *El problema del escepticismo.* ¿Es posible tener conocimiento *en absoluto*? Este problema es apremiante porque hay poderosos argumentos, algunos muy antiguos, a favor de la conclusión de que no lo es. Por ejemplo, aunque el conocimiento no puede depender de meras suposiciones, todos los argumentos deben llegar a un fin en algún momento. Parece seguirse que, en última instancia, los puntos de vista de todos dependen de suposiciones que no pueden justificar y, por ende, no constituyen conocimiento genuino. [...] Puesto que hay buenas razones para conectar al conocimiento con la justificación, el problema del escepticismo está íntimamente ligado con el *problema de la justificación*.

5. *El problema del valor.* [...]¿Tiene valor el conocimiento], y si es así, por qué? Suponiendo que lo queremos, ¿para qué lo queremos? ¿Lo queremos de manera incondicional, o sólo para algunos propósitos y en algunas situaciones? ¿Es el conocimiento la única meta de la investigación, o hay otras de igual (o mayor) importancia? (Williams, 2001: 1-2).

Aunque se presentan en cierto orden, no hay una jerarquía definida que indique cuál de estos problemas tiene prioridad sobre el resto o en qué secuencia hayan de ser enfrentados. Esto se debe a que los problemas se encuentran íntimamente relacionados entre sí: la respuesta que se dé a cada uno de ellos impondrá severas restricciones sobre qué respuestas son razonables para los restantes. Se puede comenzar, *e.g.*, por responder al problema analítico, ofreciendo una definición del conocimiento que sea neutral frente a la cuestión de si el conocimiento es posible o si tiene algún valor. También es admisible enfrentar primero el problema del escepticismo, sin partir de una definición precisa de ‘conocimiento’ ni haber tomado posición ante la incógnita de si el conocimiento es deseable o valioso. Asimismo, es viable comenzar respondiendo al problema del valor, con la esperanza de que esto eche luz sobre los otros problemas.³

3.1.3. Objetivos y estrategias

Las decisiones sobre la importancia relativa de, o sobre cómo establecer las relaciones de prioridad explicativa entre, los diversos problemas de la epistemología están, asimismo, asociadas al carácter de diversos proyectos filosóficos. Éstos se manifiestan en la predilección por ciertos objetivos teóricos que se plantea una explicación sustantiva del conocimiento. Siguiendo a Sosa (1982), puede considerarse que, tradicionalmente, el estudio filosófico del conocimiento se ha orientado a la prosecución de tres objetivos distintos pero interrelacionados:

- (a) *Objetivo descriptivo.* Se quiere comprender de manera adecuada lo que es el conocimiento y cómo llegamos a obtenerlo. Se busca una explicación simple, general y persuasiva de las condiciones requeridas por los logros epistémicos.
- (b) *Objetivo normativo.* Se desea establecer preceptos o reglas generales para asegurarnos de que nuestras creencias verdaderas sean conocimiento. Se propone

³ A menudo se sugiere que ésta es la vía recientemente emprendida por los epistemólogos de la virtud (véase, *e.g.*, *Normas, virtudes y valores epistémicos*, 2011: 11-12).

modificar nuestra situación epistémica, para llegar a satisfacer las condiciones necesarias para el conocimiento.

- (c) *Objetivo de validación.* Se pretende responder a los cuestionamientos radicales que amenazan con desacreditar nuestro presunto conocimiento. Se plantea evaluar nuestra condición epistémica y, si es preciso, revisar nuestros juicios sobre lo que consideramos que sabemos.

Estos objetivos no son, por sí mismos, mutuamente excluyentes. A menudo se tiene la intención satisfacerlos todos. Sin embargo, es frecuente que alguno de ellos predomine sobre los restantes. Cuando así sucede, el tipo de investigación emprendida para afrontar los problemas centrales de la epistemología adquiere un carácter peculiar, que se refleja en la clase de presuposiciones que se consideran legítimas. Considérense, *e.g.*, las siguientes estrategias teóricas, claramente orientadas a la consecución de uno de los objetivos anteriores:

- (Ea) *Estrategia descriptivista.* Asume que hay maneras de obtener conocimiento y que, en buena medida, nuestras atribuciones de conocimiento son correctas. Se propone encontrar una explicación satisfactoria de estos hechos, *i.e.*, de por qué los procedimientos que empleamos para obtener conocimiento efectivamente nos lo proporcionan, de un modo tal que nuestras atribuciones de conocimiento sean, en su mayoría, correctas.
- (Eb) *Estrategia metodista.* Supone que hay una explicación de por qué ciertos procedimientos nos proporcionarían conocimiento y de bajo qué condiciones nuestras atribuciones de conocimiento serían correctas, así como que nuestras atribuciones de conocimiento son, en buena medida, correctas. Busca formular cánones mediante los cuales se nos permita conocer más y entender mejor.
- (Ec) *Estrategia validacionista.* Admite que hay una explicación de por qué ciertos medios proporcionarían conocimiento y de bajo qué condiciones nuestras atribuciones de conocimiento serían correctas, así como que hay preceptos para obtenerlo. Considera que la tarea de una teoría filosófica del conocimiento con-

siste en investigar si en efecto tenemos, o si podemos tener, conocimiento de acuerdo a dicha explicación y a tales preceptos.

A la luz de un estrategia descriptivista, se presupone una solución al problema escéptico y a los problemas concernientes al método y al valor; la cuestión radica en encontrar una respuesta a los problemas restantes que haga justicia, de manera simple y general, a tales suposiciones. Un partidario de la línea metodista asumirá una respuesta a los problemas analítico, de la demarcación y del escepticismo, concentrando sus esfuerzos en responder apropiadamente a los problemas del método y del valor. Quien adopta la estrategia validacionista supondrá una respuesta a los primeros tres problemas y al último, dedicando su atención a la resolución del problema del escepticismo.Cuál de estas estrategias deba emprenderse es una cuestión polémica e intrincada, que involucra temas concernientes a la metodología filosófica. No obstante, sería razonable suponer que algunas estrategias son preferibles, en buena medida, en términos de su *productividad teórica* (*i.e.*, que sus resultados sea comparativamente mejores, más simples y más generales que los productos de las estrategias alternativas) y de la *razonabilidad de sus supuestos* (*i.e.*, que sea, en la medida de lo posible, conservadora o requiera una mutilación mínima de nuestra concepción del mundo), así como de su *desempeño en la resolución de problemas* (*i.e.*, que maximice la cantidad de problemas resueltos y minimice la de problemas no resueltos o anómalos).

Dos acotaciones son requeridas. En primer lugar, al menos una parte del problema del valor parece tener un estatuto especial, que lo hace ubicuo a las tres estrategias. En efecto la importancia del proyecto epistemológico mismo está supeditada a que el conocimiento sea, en cierto sentido, valioso. Negar esto socavaría la importancia de la empresa epistemológica en general. En segundo lugar, mientras las estrategias (Ea) y (Ec) son preponderantemente teóricas, el proyecto que se desprende de la estrategia (Eb) involucra aspectos relativos a consideraciones de racionalidad práctica, que involucran *desiderata* epistémicos y extra-epistémicos.

La caracterización de estas estrategias es, por supuesto, una idealización: difícilmente se presentan en sus formas puras. Hay al menos dos razones para esto. La primera es que, contra lo que pudiera sugerir la discusión previa, los cinco problemas arriba descritos se entrelazan en varios puntos; sus fronteras son difusas. La segunda se conecta con la naturaleza de la investigación filosófica: las suposiciones de cada estrategia están continuamente sujetas a examen, por lo que un proyecto que comienza orientado por una de ellas puede, posteriormente, dirigirse más enfáticamente a la prosecución de alguna de las otras.

Haciendo uso de estas distinciones, en la siguiente sección se examinará una concepción históricamente influyente de cómo entender la tarea de ofrecer una teoría filosófica del conocimiento. Asimismo, se harán explícitas algunas de las suposiciones en torno al conocimiento de la propia mente que permean esta concepción del proyecto epistemológico. En la sección 3.3 se expondrán tres de los retos que enfrenta esta tradición filosófica. En las secciones restantes se argumenta a favor de una concepción distinta de la empresa epistemológica y del lugar que el conocimiento de la propia mente ocupa en ésta.

3.2. El lugar del conocimiento de la propia mente en la epistemología cartesiana

Dentro de toda su diversidad, la epistemología moderna osciló de manera predominante entre dos estrategias teóricas: la metodista y la validacionista. Aunque con importantes excepciones (quizá Kant pueda contarse entre ellas), la estrategia descriptivista no fue muy socorrida por los filósofos modernos.⁴ Esta tendencia encuentra ilustrada, de manera ejemplar, en la filosofía de Descartes. Éste, por una

⁴ Muestra de ello se encuentra en el hecho de que el problema analítico recibió, entre los siglos XVII y XIX, considerablemente poca atención. Este problema, que ya había sido enfrentado por Platón, fue en gran medida ignorado hasta inicios del siglo XX. Uno de los hitos en su tratamiento es el artículo de Gettier (1963) y las múltiples respuestas que le sucedieron (véase el apartado 1.1.1).

parte, se propuso responder a los embates escépticos y determinar la verdadera extensión de nuestro supuesto conocimiento, siguiendo la estrategia validacionista; por otra parte, quería formular un método, como lo consigna el subtítulo de su (1637), “para conducir bien la propia razón y encontrar la verdad en las ciencias”, perseverando en la estrategia metodista. En lo sucesivo, nos referiremos a la influyente tradición filosófica que se inspira en estos aspectos del proyecto de Descartes con la designación ‘epistemología cartesiana’.

3.2.1. Objetivos y supuestos de la epistemología cartesiana

Con frecuencia se considera que los objetivos de legitimar el conocimiento y reglamentar su búsqueda, que se encuentran en el núcleo de la epistemología cartesiana, fueron motivados principalmente por dos sucesos históricos. En primer lugar, las profundas revoluciones intelectuales de la naciente ciencia moderna frente a la aún persistente influencia de las ideas aristotélicas reclamaban alguna clase de vindicación y fundamento. En segundo lugar, en el siglo XVI se produjo una importante recuperación de las ideas de los filósofos escépticos de la antigüedad. Estos sucesos propiciaron que se considerara necesaria una evaluación minuciosa de las opiniones heredadas de la antigüedad y el Medioevo, así como que se buscaran métodos para garantizar la obtención de conocimiento. No obstante, es preciso advertir que, sin importar cuáles fuesen las motivaciones subyacentes para este proyecto filosófico, no debería suponerse que la estratagema cartesiana es la única, ni la mejor, manera de afrontar filosóficamente los problemas en torno al conocimiento humano.⁵

⁵ Algunos han considerado, no obstante, que adoptar la estrategia validacionista es el rasgo característico de la epistemología en general: “El proyecto de justificar o legitimar lo que consideramos que es nuestro conocimiento, o al menos exhibir una posible justificación para él, era la meta de lo que llegó a conocerse como ‘epistemología’” (Stroud, 2008: 514). Contra esta clase de interpretaciones, aquí se defiende que ésa ha sido sólo una, sin duda influyente, de las varias estrategias que pueden emprenderse en la tarea de ofrecer una teoría filosófica del conocimiento.

Si somos consecuentes con la caracterización de las estrategias identificadas en la sección precedente, el compromiso del epistemólogo cartesiano con los proyectos validacionista y metodista involucra la suposición de cierta explicación sustantiva de la naturaleza del conocimiento y, en buena medida, de la cuestión de la demarcación. Estos supuestos constituyen el trasfondo no problemático sobre cuya base se propone ofrecer una respuesta satisfactoria a los problemas del escepticismo, del valor y del método. Dadas las intrincadas y minuciosas distinciones entre los diversos proyectos que se inscriben en esta tradición epistemológica, resultaría deshonesto tratar de hacerles justicia en unas cuantas páginas. De modo que lo que aquí se presenta es, en el mejor de los casos, esquemático.

¿Qué suposiciones en torno al conocimiento se encuentran en el trasfondo del proyecto cartesiano? Podemos asumir que la epistemología cartesiana adopta, en líneas generales, la caracterización tripartita del conocimiento; *i.e.*, considera que creencia, verdad y justificación proporcionan un análisis satisfactorio del conocimiento, entendido como saber proposicional.⁶ De acuerdo con este análisis, un enunciado de la forma “*S* sabe que *p*” puede ser representado de manera más perspicua señalando que “*S* cree que *p*, *p* es verdadera y la creencia de *S* de que *p* está justificada”. Dado el énfasis de esta tradición filosófica en los objetivos normativo y de validación, la más crucial de las condiciones de este análisis concierne a la justificación. En efecto, desde el extremo validacionista, se vuelven apremiantes cuestiones como las siguientes: ¿cuál es la conexión apropiada entre una creencia y la verdad de su contenido para que resista los cuestionamientos escépticos?, ¿qué clase de apoyo requiere una creencia verdadera para constituir conocimiento?; en el flanco metodológico, se presentan también interrogantes en torno a la justificación: ¿cómo debe procederse para maximizar nuestras creencias

⁶ Es oportuno señalar que buena parte de la epistemología moderna le concedió mayor importancia al conocimiento objetual que al saber proposicional. No obstante, dejaremos al margen este asunto para los propósitos de esta discusión.

verdaderas, evitando las falsas?, ¿qué principios o preceptos nos aseguran que lo que hemos obtenido es conocimiento y no mera opinión (fortuitamente correcta)?

3.2.2. Justificación: fuerza y estructura

En este punto, el énfasis en el proyecto validacionista tiene una consecuencia peculiar. Enfrentando radicalmente el problema escéptico, el epistemólogo cartesiano asume que la justificación de una creencia no debería dejar lugar para la duda sobre su verdad; *i.e.*, que el auténtico conocimiento (*scientia*) requiere justificación de carácter infalible. A esto le corresponde un precepto metodológico sumamente demandante, *i.e.*, una regla de (auto)evidencia (*e.g.*, “procura adquirir o conservar sólo aquellas creencias que no puedan ser falsas, o que se deriven de creencias de esta índole mediante procedimientos que preserven necesariamente verdad”).⁷ Pero, ¿es éste un análisis razonable de nuestra concepción cotidiana del conocimiento? Hay un argumento atractivo que proporciona apoyo intuitivo al infalibilismo sobre la justificación (véase Sturgeon, 1998: 14-15). Comienza al notar la plausibilidad *prima facie* de la siguiente afirmación:

(*) Si *S* sabe que *p*, entonces *S* no puede estar equivocado acerca de que *p*.

El argumento procede de la siguiente manera: puesto que *S* no puede estar equivocado acerca de que *p* cuando *S* sabe que *p*, esto significa que la justificación de *S* para *p* debe garantizar que *p* es verdadera. De modo que (*) implica que la justificación garantiza verdad, *i.e.*, implica el infalibilismo. Sin embargo, la aparente

⁷ En el *Discurso del método*, la regla de evidencia se formula en los siguientes términos: “[n]o aceptar no admitir cosa alguna como verdadera si no se la había conocido evidentemente como tal. Es decir, con todo cuidado debía evitar la precipitación y la prevención, admitiendo exclusivamente en mis juicios aquello que se presentara tan clara y distintamente a mi espíritu que no tuviera motivo alguno para ponerlo en duda” (Descartes, 1937: 15). Aunque esta regla en ocasiones no se formula explícitamente, parece atravesar la tradición cartesiana; pueden, *e.g.*, ubicarse al interior de ésta los recelos humeanos en torno al razonamiento causal o inductivo.

plausibilidad de este argumento reposa en una falacia.⁸ Ésta consiste en asumir que (*) afirma que para contar como conocimiento una creencia debe estar apoyada de una manera que implique necesariamente su verdad; no obstante, en una interpretación más generosa, lo que (*) afirma es simplemente que la verdad de la creencia es una condición necesaria para el conocimiento (*i.e.*, que no puede ser el caso que S sepa que p y que p no sea verdadero).

La concepción infalibilista de la justificación sólo parece desprenderse de nuestra concepción intuitiva del conocimiento; es, más bien, producto de la estrategia validacionista asumida por la epistemología cartesiana. Esta concepción tiene también consecuencias en lo que respecta a la estructura de la justificación. Supóngase que S cree que p , y que se le pregunta por qué lo cree. Si la pregunta puede formularse, es razonable suponer que existen dudas en torno a la verdad de p . S podría responder ofreciendo una razón, q , para creer que p . Dada la concepción infalibilista de la justificación, q constituirá una buena razón para creer que p si y sólo si q es autoevidente o se deriva apropiadamente de una creencia autoevidente. Luego podría preguntársele a S por qué cree que q . A esto S podría responder ofreciendo una razón, r , para creer que q . Este proceso podría continuar durante algún tiempo. Si en algún momento S admite que una de las razones que ha dado no tiene justificación, entonces la cadena entera parecería colapsarse, dejando a p injustificada. Pero si S continúa ofreciendo razones para sus creencias, hay una amenaza de que se produzca un regreso infinito. Este problema se

⁸ Una forma de caracterizar esta falacia es considerarla como el producto de una ambigüedad entre una lectura *de dicto* del enunciado (*) y una lectura *de re*. Se trata de una ambigüedad en el alcance de la expresión modal “no es posible”. A manera de ilustración, considérese el siguiente ejemplo: Si alguien es pobre, entonces esa persona no puede poseer abundantes recursos económicos. Una lectura *de re* del enunciado anterior hace de la miseria un *fiat* de las clases menos favorecidas, al indicar que quien de hecho es pobre no puede modificar su situación económica. Por contraparte, una lectura *de dicto* se compromete con la menos polémica afirmación de que poseer escasos recursos económicos es una condición necesaria para que pueda afirmarse que alguien es pobre, *i.e.*, que si alguien que dispone de una copiosa fortuna no es unja persona pobre. La ambigüedad entre las lecturas *de re* y *de dicto* respecto a las actitudes proposicionales se discute en la sección 2.1.

conoce como el problema del regreso epistémico. Plantea la siguiente pregunta: cuando se tiene una creencia justificada, ¿cuál es la estructura completa de las razones que uno tiene en apoyo a esa creencia? Hay cuatro posibles respuestas:

- (1) *Fundacionismo*. Una creencia está justificada si y sólo si está justificada de manera básica (*i.e.*, es una creencia fundacional) o está basada, directa o indirectamente, en una o más creencias justificadas de manera básica.
- (2) *Coherentismo*. Una creencia está justificada en virtud de la manera en que se ajusta con otras creencias que constituyen un sistema o conjunto.
- (3) *Infinatismo*. Cuando una creencia está justificada hay una cadena infinita y no repetitiva de razones que la apoyan.
- (4) *Escepticismo*. Puesto que ninguna de las estructuras anteriores puede realmente respaldar (justificar) las creencias, no hay creencias justificadas.

Al menos una de las opciones (1) a (4) debe ser correcta. Si la justificación es una condición necesaria para el conocimiento, la postura (4) ofrece una solución definitiva, aunque indeseable, al problema del escepticismo: equivale a negar la posibilidad del conocimiento. También haría espurios los intentos por formular preceptos para obtener conocimiento. Incluso si la justificación no fuese necesaria para el conocimiento, (4) tendría consecuencias para el problema del método: *e.g.*, implicaría que no hay creencias racionales en virtud de su justificación. Considerando inaceptable esta respuesta,⁹ la epistemología cartesiana se orienta a mostrar que una de las opciones (1) a (3) puede caracterizar la estructura de la justificación epistémica; éstas son compatibles con el optimismo epistémico, aunque no lo implican.

La exigencia de infalibilidad para la justificación parece comprometerse con la demanda de fundamentos últimos para la creencia. En efecto, si lo que se requiere para responder al reto escéptico es que la justificación de una creencia

⁹ Esta respuesta es inaceptable no sólo porque constituye una forma de pesimismo epistémico (*i.e.*, la tesis de que de hecho no tenemos conocimiento, de que nuestra situación epistémica fáctica es deplorable), sino porque tiene como consecuencia el escepticismo radical. Véase la sección 3.3.1.

implique su verdad, entonces las únicas las creencias que pueden alcanzar el estatus de conocimiento son aquellas cuya verdad es evidente o las que se derivan a partir de éstas mediante inferencias que preservan necesariamente verdad. El coherentismo y el infinitismo no parecen satisfacer este requisito, pues no admiten que haya creencias fundacionales (*i.e.*, creencias cuya justificación no sea parasitaria de la de otras creencias). Luego, de acuerdo con la epistemología cartesiana, el fundacionismo es la respuesta al problema del regreso epistémico. En consecuencia, la justificación sería una condición bifronte: una de sus caras sería fundacional, *i.e.*, básica, autoevidente; la otra, derivada, *i.e.*, inferencial. Así, el epistemólogo cartesiano propone que se acepten como conocimiento sólo aquellas creencias que se presentan como autoevidentes o las que pueden inferirse a partir de ellas, de manera evidente.

3.2.3. El fundacionismo y el conocimiento de la propia mente

Pero, ¿qué creencias tienen estos rasgos?, y ¿qué relación debe darse y cuáles son los términos de dicha relación para que haya justificación y, por ende, conocimiento? Al afrontar estas cuestiones, el epistemólogo cartesiano debe introducir una parte de la historia descriptiva, al interior de la cual es posible formular, y responder a, los cuestionamientos escépticos y metodológicos. Intuitivamente, parece haber ejemplos claros de creencias justificadas de manera básica. Un caso paradigmático se encuentra en las creencias acerca de nuestros estados mentales conscientes en el momento presente. Si se me pregunta: “¿Por qué crees que sientes dolor?” Una respuesta sensata sería: “Porque siento dolor”. Ante esta réplica, parecería absurdo repetir la pregunta. Si, en su lugar, se inquiriera: “¿Por qué sientes dolor?”, ésta ya no sería una pregunta por la justificación de una creencia, sino por la causa del estado mental. De modo que, de acuerdo con el epistemólogo cartesiano,

el conocimiento de los propios estados mentales parece estar en la base de nuestra cadena de justificaciones sin requerir, a su vez, de justificación adicional.¹⁰

En la época moderna, los epistemólogos de la tradición cartesiana suministraron, en líneas generales, una teoría de las ideas como parte de las presuposiciones descriptivas sobre el conocimiento humano. De acuerdo con ésta, toda actividad mental es una forma de pensamiento. Pensar siempre involucra ‘ideas’, las cuales se presentan ante la mente y son los objetos inmediatos del pensamiento. Las ideas son, por así decirlo, la sustancia del pensamiento y son organizadas por las diversas actividades de la mente.¹¹ Al presentarse ante la mente, las ideas son autoevidentes, pues no puede dudarse de ellas — llevan una marca intrínseca de su propia verdad. Cuando las organizamos conceptualmente estas ideas producen representaciones (*e.g.*, creencias), que tienen condiciones de verdad objetivas (*i.e.*, son acerca de algo, pueden ser verdaderas o falsas). Dichas representaciones también se presentan a la mente de manera evidente. Si el fundacionismo caracteriza correctamente la estructura de la justificación y ésta debe ser infalible, para que algunas nuestras creencias con contenido objetivo sean conocimiento debe ser posible mostrar que son verdaderas únicamente en virtud de las ideas que suministran su contenido. La teoría de las ideas cumple dos funciones: por un lado, constituye un fundamento a partir del cual puede evaluarse nuestra condición epistémica respecto a otros logros cognoscitivos; por otro lado, es lo que proporciona unidad metódica a la obtención de conocimiento.

¹⁰ Continuando la tradición epistemológica cartesiana en lo que concierne al papel de lo mental en la fundamentación del conocimiento, Roderick Chisholm señalaba: “[a] la pregunta ‘¿Qué justificación tengo para pensar que sé o para contar como evidente que ahora algo me parece rojo o me sabe agrio?’ Podría responder sólo reiterando que ahora algo parece rojo o sabe agrio” (Chisholm, 1977: 29).

¹¹ Es probable que Descartes fuese el primer filósofo en emplear la palabra ‘idea’ de esta manera, pero su uso se volvió frecuente y fue común tanto a Locke como a Berkeley (véase Scruton, 1998: 446-447); otro tanto podría decirse acerca de Hume (pero véase Stroud, 1977: chap. 2). Entre los epítetos para los sucesores contemporáneos de las ideas pueden contarse los datos sensoriales, las sensaciones no interpretadas, lo dado en la experiencia.

Podemos ilustrar la función evaluativa de la teoría las ideas en la epistemología cartesiana por medio de una analogía. John Heil sugiere que, desde la perspectiva de lo que aquí denominamos estrategia validacionista, el reto de ofrecer una evaluación de nuestras atribuciones de conocimiento puede compararse la tarea de calibrar un instrumento:

Los sujetos cognoscentes son instrumentos que registran cómo son las cosas. [...] Antes de que podamos depender confiablemente de un instrumento, debemos ver que esté calibrado. Podemos calibrar un instrumento sólo si captamos cómo funciona [...] Al calibrar un instrumento, no se depende de ese mismo instrumento. El proceso de calibrar un instrumento requiere que se tenga un acceso independiente a información de cómo son las cosas en el dominio al interior del cual el instrumento debe funcionar. De este modo se está en posición de usar esa información para evaluar la exactitud de los resultados del instrumento. [...] Parece que si hemos de calibrarnos como conocedores, debemos hacerlo de alguna manera “desde dentro”. [...] Descartes es famoso por presentar este punto de manera prominente. El compromiso con un mundo independiente de la mente trae consigo una pesada carga epistemológica. Si nos proponemos establecer que de hecho poseemos conocimiento del mundo, debemos hacerlo de forma tal que no involucre que tengamos conocimiento independiente de ese mundo. Esto equivale a un requisito de que los intentos por establecer que tenemos conocimiento no sean epistémicamente circulares (Heil, 2002: 316-317).

El requisito de no circularidad epistémica en un argumento equivale a que su conclusión no sea presupuesta en la suposición de que sus premisas están justificadas. De modo que la tarea de la estrategia validacionista se propone responder a la amenaza de que no seamos capaces de producir una *demonstración* satisfactoria de la hipótesis de que tenemos conocimiento, *i.e.*, de que algunas de nuestras creencias con contenido objetivo son verdaderas. Como señala Stroud, la consumación del objetivo validacionista bajo estas restricciones requiere el previo reconocimiento de cierta independencia epistémica entre los dominios a investigar:

La exigencia de una comprensión completamente general del conocimiento en cierto dominio requiere que nos veamos a nosotros mismos al comienzo como no sabiendo nada en ese dominio y luego llegando a tener tal conocimiento sobre la base de algún conocimiento o experiencia independiente y, en ese sentido, previo. Y esto nos lleva a buscar un punto de vista a partir del cual podamos vernos a nosotros mismos sin dar por sentado nada de ese conocimiento que queremos comprender (Stroud, 2000: 120).

Una importante consecuencia de esta concepción de la mente, que hemos asociado con el objetivo de validación de la epistemología cartesiana, es que la investigación filosófica rigurosa debe proceder mediante una maniobra que va desde el interior (*i.e.*, desde la propia mente) hacia lo externo (*i.e.*, el mundo y las otras mentes). En este sentido, la epistemología cartesiana es una aproximación a los problemas del conocimiento desde la perspectiva de la primera persona. Esta táctica es asiduamente practicada por el autor de la *Meditaciones metafísicas*,¹² pero perdura como el sello distintivo de muchos filósofos posteriores. En última instancia, para la tradición epistemológica cartesiana todo conocimiento se fundamenta en la inspección de las ideas de la mente. En esta concepción, algunos estados mentales (en la tradición empirista, las sensaciones) operan como intermediarios epistémicos entre nuestras creencias y aquello que determina su verdad. Los contenidos de los estados mentales no son, pues, determinados por el mundo externo sino que se constituyen en el dominio autónomo, subjetivo, de lo mental.

En la siguiente sección examinaremos tres retos que se desprenden, en el marco de estos supuestos, de la agenda teórica de la epistemología cartesiana. Se sugerirá en concordancia con buena parte del diagnóstico contemporáneo que ésta no se encuentra en condición de responder a dichos retos. En la sección 3.4 se considera la adopción de otro proyecto filosófico: la epistemología naturaliza-

¹² Un diagnóstico similar puede encontrarse en la lectura de Bernard Williams de la filosofía cartesiana: “Un último punto debe señalarse sobre el proyecto de Descartes: que es radicalmente de primera persona. Algunos filósofos han supuesto o presupuesto que la pregunta más básica de la teoría del conocimiento debe tomar la forma ‘¿qué *puedo* saber?’, y Descartes está entre ellos, es quizá el primero de ellos” (Williams, 1978: 52).

da, que, como sugeriré, se desprende de otra estrategia teórica para abordar los problemas de la epistemología. En la sección 3.5 se discute la función del conocimiento de la propia mente en una forma de epistemología naturalizada.

3.3. Tres retos de la epistemología cartesiana

La epistemología cartesiana enfrenta varios retos, que se desprenden de las estrategias teóricas que suscribe. Por un parte, el objetivo normativo determina la búsqueda de principios o preceptos para la obtención de conocimiento; por otra, el objetivo de validación exige la respuesta a ciertos planteamientos que amenazan la corrección de nuestras atribuciones de conocimiento. En esta sección se examinarán tres de los problemas que se desprenden de estas estrategias. En lo que sigue analizaremos los desafíos de la epistemología cartesiana desde la perspectiva de la estrategia validacionista, por lo que adoptarán la forma de cuestionamientos escépticos. Aunque hay correlato natural de estos retos para la estrategia metodista que ubica al proyecto cartesiano en una situación igualmente desesperada, para fines expositivos será conveniente que nos centremos en el escepticismo. Se sugerirá que, en el trasfondo de las suposiciones antes expuestas en torno al conocimiento de la propia mente, estos retos socavan el proyecto epistemológico cartesiano. Como respuesta, en las siguientes secciones se examinará una aproximación distinta a los problemas epistemológicos.

3.3.1. El problema del escepticismo

El epistemólogo cartesiano se propone hacer una evaluación general de nuestro putativo conocimiento. Para ello asume, como indicamos en la sección precedente, un análisis del conocimiento: éste consiste en una creencia verdadera que se encuentra justificada de tal manera que su falsedad no es posible. Supone, asimismo,

que los únicos recursos disponibles para tal evaluación se limitan a creencias fundacionales acerca del ‘mundo interno’, de las ideas de la mente y las representaciones generadas a partir de ellas. Una pregunta apremiante para este programa de investigación es: ¿cómo podemos tener conocimiento fuera de este restringido dominio?, ¿cómo podemos determinar si el contenido de nuestras representaciones es verdadero? Lo que se busca es una demostración de que tal conocimiento es posible; una demostración de esta índole nos permitiría evaluar nuestras atribuciones de conocimiento (y, tal vez, mejorar nuestra situación epistémica). Una restricción importante a este proyecto es que dicha demostración no debe ser epistémicamente circular, no debe presuponer que tengamos conocimiento del dominio que deseamos investigar.

El escepticismo filosófico es la tesis de que no podemos tener conocimiento o de que nuestra situación epistémica es más precaria de lo que ordinariamente creemos. En sus formas filosóficamente interesantes, depende siempre de un argumento y, por ende, parte de ciertas suposiciones.¹³ En el marco de esta discusión, nos centraremos en formas de escepticismo que se desprenden de las suposiciones adoptadas por la epistemología cartesiana. De este modo, si el conocimiento en algún dominio es posible, debe estar justificado (de manera infalible) sobre la base del conocimiento de nuestros propios estados mentales. De lo contrario, debemos claudicar ante la conclusión escéptica.

¹³ Podemos ilustrar este rasgo del escepticismo filosófico considerando dos criterios que a menudo se emplean para denominar a alguien escéptico:

- (i) cuando, ante cada afirmación, replica con la pregunta “¿Cómo lo sabes?” hasta que las respuestas se agotan;
- (ii) cuando tiene estándares de evidencia muy elevados, tanto que ni siquiera pueden satisfacerse.

Las posiciones (i) y (ii) no son, por sí mismas, filosóficamente interesantes. A menos que, quien se comporta de acuerdo a (i) crea que para saber algo es necesario saber *cómo* se sabe, su conducta no manifiesta una posición de la que podamos aprender mucho. Y (ii) es filosóficamente interesante sólo si podemos ofrecer razones para sostener que los estándares más elevados de evidencia son mejores que los ordinarios. En resumen, las formas filosóficamente interesantes de escepticismo dependen siempre de un argumento (véase Dancy, 1985: 7-8).

Es preciso distinguir dos grupos de cuestiones en torno al problema escéptico. El primer grupo concierne a nuestra situación epistémica fáctica, *i.e.*, a si de hecho tenemos o podemos tener conocimiento, si nuestras atribuciones epistémicas son correctas. Este tipo de cuestiones constituye una de las fuerzas motrices que guían al epistemólogo cartesiano. El segundo grupo refiere a asuntos de carácter trascendental, *i.e.*, a si el conocimiento es, en general, posible. La respuesta a este tipo de cuestiones no es opcional, ni siquiera para el epistemólogo cartesiano; esto se debe a que si las preguntas en torno al conocimiento tienen siquiera sentido, deberían tener condiciones de aplicación (*i.e.*, condiciones de verdad), incluso si éstas son sólo contrafácticas. Emplearemos diversos epítetos para las formas de escepticismo que se ubican en el primer grupo; constituyen formas de pesimismo epistémico. La denominación ‘escepticismo radical’ se reservará a la negación de que el conocimiento sea (fáctica o contrafácticamente) posible. Puesto que el escepticismo radical no es opcional, no puede considerarse una conclusión razonable de un proyecto epistemológico, sino una reducción al absurdo de sus supuestos.

3.3.2. Tres formas de escepticismo

Los objetivos y supuestos del proyecto cartesiano permiten precisar con claridad la naturaleza de los problemas escépticos que enfrenta. La cuestión es: ¿podemos mostrar que algunas de nuestras creencias tienen que ser verdaderas y son, por ende, inmunes al error?⁹ En lugar de basarse en un examen exhaustivo de las fuentes del error, como lo hicieran los escépticos antiguos, el problema cartesiano se gesta cuando la duda llega a un germen irreducible: nuestra experiencia del mundo. De este modo, la noción de idea, de impresión, de dato sensorial o de lo dado en la experiencia desempeña su función en la arquitectura fundacionista. Quizá nuestras creencias sobre el mundo sean falsas, pero es claro que los tene-

mos; y las tenemos en virtud de que cierto contenido nos es dado en la experiencia. La duda y la posibilidad del error llegan a su límite en el reconocimiento de la existencia de ciertos estados mentales que presuntamente tratan acerca del mundo y se presentan con un contenido determinado. En el proyecto epistemológico cartesiano, si tenemos conocimiento de un mundo independiente de nuestra mente, debe haber una manera de mostrar que el contenido de nuestras creencias, suministrado por lo dado en la experiencia, garantiza su verdad objetiva, independientemente de cualquier representación. El escepticismo del mundo externo asevera que no podemos tener conocimiento sobre el contenido objetivo de nuestras creencias, puesto que la experiencia no puede justificar dicho contenido, no puede implicar su verdad.

Podemos emplear la analogía de la calibración para ilustrar este punto. Si la verdad de una creencia depende de su adecuación con el mundo, parecería que la única manera en que podríamos saber que una creencia es verdadera sería comparándola con las porciones relevantes del mundo. El problema es que, puesto que se ha adoptado la estrategia validacionista y ésta exige una demostración que no sea epistémicamente circular, al efectuar tal comparación sólo podemos depender de otras creencias. La calibración de nuestro aparato cognitivo, de la relación entre el mundo y una creencia, consiste en la adquisición de más creencias. Surge entonces la cuestión de si estas creencias son ellas mismas verdaderas. El epistemólogo cartesiano postula una clase especial de creencias, cuya verdad podemos determinar sin tener que confrontarlas con la realidad: creencias acerca de las propias sensaciones. Si la verdad de algunas nuestras creencias sobre el mundo externo fuera implicada por estas creencias privilegiadas, entonces tendríamos una demostración de la hipótesis de que tenemos conocimiento. Pero, para hacer esto, las creencias acerca de las propias sensaciones deberían desempeñar un papel epistémico extraño. Por una parte, deberían ser representaciones del mundo, presentar una unidad de información, ser una pequeña porción de conocimiento

capaz suministrar justificación; de lo contrario, no podría implicar la verdad de otras creencias. Por otro lado, deberían ser previas a toda interpretación, tener el contenido que tienen debe ser un hecho bruto, no el resultado de ninguna actividad de razonamiento por nuestra parte; si no fuese así, su verdad no estaría garantizada y, por ende, no podrían ocupar el papel de creencias fundacionales, con justificación básica, sino que requerirían apoyo suplementario.

La objeción del escéptico es que ningún estado mental puede satisfacer esta doble función que se le asigna a las creencias sobre las propias sensaciones. Una vez que se ha abierto la brecha entre el contenido de nuestras creencias y el mundo objetivo, el daño es irremediable. Las sensaciones, como intermediarios epistémicos, no pueden cerrarla. Si el carácter representacional de las creencias se gesta con independencia del mundo, cualquiera de nuestras creencias podría ser objetivamente falsa. Podríamos encontrarnos en un estado subjetivamente indiscernible (*i.e.*, tener las mismas experiencias, las mismas sensaciones), incluso si el mundo objetivo fuese completamente distinto a como nos parece, de modo que todas nuestras creencias sobre el mundo externo fuesen falsas.

Pese a que el escepticismo del mundo externo parecería conducir a una posición extremadamente pesimista acerca del conocimiento humano, no tiene como consecuencia, de manera inmediata, el escepticismo radical (*i.e.*, la conclusión de que el conocimiento no, en absoluto, es posible). Al parecer, una manera de evitar esta conclusión sería reduciendo el dominio del conocimiento del mundo objetivo a lo intersubjetivo. Así, incluso si no tenemos una garantía de que el contenido objetivo de nuestras creencias representa adecuadamente el mundo externo, independiente de cualquier representación, quizá sea posible extender nuestro conocimiento más allá del limitado dominio de las propias ideas. Este proyecto —bajo el que podrían suscribirse diversas formas de idealismo, pragmatismo, verificacionismo y realismo interno— consiste en mostrar que si ciertas creencias son el resultado de la convergencia de cualquier sujeto racional, enton-

ces quizá su situación epistémica sea tan buena como el estatus (¿inalcanzable?) de ser objetivas — completamente independientes de la subjetividad.¹⁴ La intersubjetividad garantizaría la verdad de una creencia al establecer su validez con independencia de quien la sustenta. No obstante, el problema radica ahora en identificar aquellas creencias sobre las que deba haber acuerdo intersubjetivo, y hacerlo de una manera que no sea epistémicamente circular. El escepticismo de otras mentes sugiere que esto no es posible.

Para el epistemólogo cartesiano, nuestras creencias sobre el mundo externo son resultado de ciertas ideas. No obstante, su carácter representacional es producto de nuestra operación conceptual sobre la experiencia; esto le da al pensamiento su carácter representacional. El contenido de nuestras creencias representacionales excede lo dado en la experiencia, reflejando nuestra contribución inferencial e interpretativa. El objetivo del fundacionismo es desmantelar los estratos de la inferencia y la interpretación que se presentan en nuestras creencias (donde impera “la soberanía conceptual del hombre”), y llegar a algo previo a éstas, a lo dado en bruto (donde nuestros juicios se confrontan directamente con “el tribunal de la experiencia”);¹⁵ luego, a partir de ello, se propone volver a construir nuestra imagen del mundo, verificando todos los vínculos de la cadena interpretativa. Si tenemos conocimiento de un mundo independiente de *nuestra* mente, debe haber una manera de mostrar que el contenido de nuestras creencias, suministrado por lo dado en la experiencia, garantiza su verdad intersubjetiva, para cualquier representación posible. El escepticismo de otras mentes asevera que no podemos tener conocimiento sobre el contenido objetivo de nuestras creencias, pues puede haber maneras radicalmente distintas de organizar la experiencia.

¹⁴ Esta respuesta puede robustecerse si se acompaña de la tesis metafísica de que no hay una brecha entre experiencia y realidad; no obstante, si hemos de persistir en la investigación de nuestros logros cognitivos, esto no debe tener como consecuencia que todas nuestras creencias sean conocimiento (de ahí la importancia de la intersubjetividad como restricción).

¹⁵ La expresiones entrecomilladas son de Quine (1960) y (1951), respectivamente.

La situación en este punto puede considerarse incluso más desesperada que la del escepticismo del mundo externo, pues para saber si otros sujetos racionales comparten nuestras creencias debemos estar en condiciones de saber si hay otros sujetos racionales y cuáles son sus creencias. Para hacer esto, parece que debemos servirnos de más recursos que aquellos de los que disponíamos originalmente, pues dar cuenta del acuerdo intersubjetivo requiere un mundo compartido en el que varios sujetos interactúan y en el cual la comunicación es posible. Dada la concepción de lo mental defendida por el epistemólogo cartesiano, esto parece problemático incluso si contásemos con conocimiento del mundo externo. En efecto, para determinar si algo es capaz de pensamiento y cuáles son sus pensamientos requerimos disponer de criterios que nos permitan hacer una inferencia a partir de la observación de su conducta; pero si existe una brecha entre lo mental y el mundo objetivo, esto no es suficiente para zanjar la cuestión:

La independencia lógica de lo mental opera igualmente en la otra dirección: ninguna cantidad de conocimiento del mundo externo implica la verdad acerca del funcionamiento de una mente. Si hay una barrera lógica o epistémica entre la mente y la naturaleza, no sólo nos previene de ver hacia afuera; también bloquea una visión desde fuera hacia adentro (Davidson, 1991: 207).

Si el mundo externo y el pensamiento de otros sujetos racionales nos están vedados en el intento de investigar la justificación de nuestras creencias, queda abierta la cuestión de si podemos emprender tal investigación a partir de nuestro propio caso. Tal investigación consistiría en dirimir si somos capaces de establecer, con los limitados recursos puestos a nuestra disposición por el tipo fundacionismo avalado por la epistemología cartesiana, principios de racionalidad cognitiva. El escepticismo acerca de la racionalidad niega que esto sea posible.

Lo que el programa cartesiano se propone es mostrar que algunas creencias con contenido objetivo constituyen conocimiento, *i.e.*, que puede mostrarse que son necesariamente verdaderas a partir de ciertas experiencias. No obstante, en

esta inferencia no puede intervenir información sobre cómo son las cosas independientemente de cómo se nos presentan; tampoco puede interferir información acerca de lo que otros agentes racionales consideran que es el caso. Tales intrusiones implicarían que se produce cierta circularidad epistémica, con lo que colapsaría el proyecto de ofrecer una demostración de la hipótesis de que el conocimiento es posible. De modo que, si algunas de nuestras creencias objetivas verdaderas han de constituir conocimiento, necesitamos mostrar que tales creencias se siguen a partir de sensaciones, mediante ciertas reglas o preceptos. Puesto que se han asumido tanto el contenido de dichas creencias como lo dado en la experiencia, lo problemático es identificar cánones de racionalidad epistémica mediante lo que hayamos de adoptar ciertas creencias sobre la base de sensaciones. Si pueden identificarse procedimientos o procesos de formación de creencias de esta índole que sean intrínsecamente confiables, entonces tendremos una demostración no circular de la hipótesis de que el conocimiento es posible. No obstante, el escéptico objetará que la identificación de tales preceptos requeriría, a su vez, garantías independientes. Esto apunta a un regreso indeseable en la justificación. Si intentamos detener dicho regreso, habremos renunciado a la exigencia de no circularidad epistémica. *Ergo*, el conocimiento objetivo, en el trasfondo de los supuestos cartesianos, no es posible. Dado el carácter general de los argumentos escépticos, esta conclusión no sólo se aplica a nuestra precaria situación epistémica, sino que, partiendo de los supuestos del proyecto cartesiano, es la condición de cualquier sujeto que busque fundamentar sus creencias a partir de sensaciones.

3.3.3. El fracaso de la epistemología cartesiana

Al intentar justificar todo nuestro conocimiento, como lo hace el epistemólogo cartesiano, se pretende demostrar que aquello que creemos en la ciencia y en la vida cotidiana está conectado con y está garantizado por las bases o fundamentos sobre los cuales llegamos a creerlo. En esta concepción del proyecto epistemoló-

gico, toda la posible evidencia que poseemos se restringe, en última instancia, al contenido de nuestra propia mente. Pero, al reconocer la autonomía de lo mental, cualquier curso particular de experiencia podría fracasar en darnos información confiable acerca de cómo son las cosas: por lo que sabemos, el mundo puede ser distinto de la manera en que lo percibimos; otros sujetos podrían organizar conceptualmente la experiencia de maneras divergentes; nuestros estándares de racionalidad podrían ser deficientes. Al interior del proyecto epistemológico cartesiano, esto tiene consecuencias desastrosas.

Puesto que todo nuestro conocimiento está siendo examinado a la vez, no contamos con información independientemente confiable que nos permita dirimir la cuestión de si nuestro curso de experiencia presente es o no una guía confiable sobre cómo son las cosas. Una vez que hemos concedido que los fundamentos o bases de todas nuestras creencias se restringen al flujo de la experiencia, y ya que hemos distinguido entre cómo son las cosas (*i.e.*, el contenido objetivo de las creencias) y cómo nos parece que son (*i.e.*, lo dado en la experiencia), no puede eliminarse la posibilidad de que el mundo sea completamente distinto a como lo percibimos y creemos que es. Si tal posibilidad ha de ser eliminada para que el conocimiento posible, y no puede ser eliminada sobre la base del contenido de nuestra propia mente por sí sólo, entonces el conocimiento es imposible (véase Stroud, 209-210). Sugiero que esta *coda*, que constituye una forma de escepticismo radical, debería considerarse una reducción al absurdo del proyecto epistemológico cartesiano.

El epistemólogo cartesiano se propone considerar con seriedad los cuestionamientos escépticos debido al interés por comprender y clarificar la naturaleza de la nuestro conocimiento. Esto requiere ser considerado con cierto recelo. Lo que el escéptico solicita es que pongamos en entredicho la valoración que atribuimos a nuestros métodos racionales y desconfiemos (al menos provisionalmente) de ellos. Lo inquietante es si, respecto a esta cuestión particular, considerar las dudas del

escéptico no es, en cierto sentido, abandonar la empresa epistémica más general, de la cual el problema del escepticismo constituye sólo un capítulo. El escéptico radical que he bosquejado pretende tomar parte en una empresa epistémica y a la vez no disponer de ninguna metodología ni criterio asegurado para llevarla a cabo. Pero no puede hacer esto. No puede poner en duda todos los métodos y estándares racionales, y, simultáneamente, buscar una respuesta racional a los cuestionamientos que plantea. Podemos ver ahora que el escepticismo radical no es una postura que pueda seriamente adoptarse dentro de la empresa epistemológica. Suspender nuestro juicio sobre toda cuestión epistémica equivale, de igual manera, a tirar por la borda nuestra tarea cognitiva. Aunque se puede desistir del proyecto de ofrecer una teoría filosófica del conocimiento, hacer esto no equivale a dar una respuesta a los problemas que tal teoría plantea.

En la medida en que sostengamos pretensiones cognoscitivas, no podemos suspender, sin más, todas nuestras atribuciones de conocimiento. Esto no significa que debamos, de entrada, avalarlas todas y ser dogmáticos al respecto. Tampoco significa que debamos renunciar a nuestro objetivo de dar cuenta del conocimiento en general. En todo caso, abdicar ante la tarea de dar cuenta de nuestros procedimientos epistémicos no significa dejar de depender de ellos. En estricto sentido, no podemos abandonarlos todos y seguir siendo racionales, pues son parte de lo que constituye ser racional. No tiene sentido preguntar, de alguien que se formula dudas escépticas (y que, por ende, piensa), si es generalmente racional, ya que ser racional es lo que se requiere para tenerlas. Asimismo, las personas no pueden decidir si aceptan o no los atributos fundamentales de la racionalidad: si están en condiciones de decidir en absoluto, es debido a que poseen tales atributos.

El problema del escepticismo radical no es como parece diagnosticar Williams que su punto de partida no sea intuitivo, ni que deba serlo para que pueda considerársele como un problema serio (véase Williams, 2001: 147). El

problema se ubica en que parece conducir a la investigación epistemológica cartesiana una especie de *reductio ad absurdum*. Si el proyecto epistémico del que toma parte el epistemólogo cartesiano desemboca en la conclusión de que no podemos asegurar métodos para la obtención y justificación de conocimiento, entonces ni si quiera los supuestos del escéptico están garantizados. Una vez llegados a este punto, resulta más iluminador encontrar una mejor manera de plantear el asunto. Ésta es la motivación tras el naturalismo filosófico.

3.4. La naturalización de la epistemología

La expresión lingüística “naturalismo” designa, en la filosofía contemporánea, a un grupo más o menos heterogéneo de puntos de vista en diversas áreas de la filosofía; la expresión tiene diversas acepciones para gran variedad de pensadores, tanto para aquellos que suscriben alguna de sus variantes como para quienes lo critican. Asimilar el naturalismo a un conjunto de tesis específicas (sean epistemológicas, metafísicas o éticas) representaría una distorsión de su contenido poco propicia para la comprensión de su influencia. Alternativamente, el naturalismo contemporáneo en filosofía puede ser (toscamente) caracterizado como una *actitud* ante la filosofía, sus métodos, objetivos y alcance, a saber: una actitud que enfatiza las similitudes, más que las diferencias, entre la filosofía y la ciencia natural. Esto, no obstante, resulta poco informativo.

3.4.1. El naturalismo filosófico

Una manera más esclarecedora de representar la posición naturalista es por medio de la disyunción de varias afirmaciones, algunas de las cuales, sin ser exclusivas de esta postura, enérgicamente sostienen sus defensores:

- (i) Rechazo a lo sobrenatural (divinidades, demonios, almas, fantasmas; la extensión de “sobrenatural” está sujeta a amplia controversia) en la explicación de cualquier ámbito de lo real.
- (ii) Alta valoración de la ciencia (ya sea de sus métodos y/o de sus resultados), de ahí que el naturalismo sea a veces denominado “cientificismo” o “naturalismo científicista”.
- (iii) Introducción de consideraciones empíricas en la investigación filosófica y, por ende, oposición a o, en el mejor de los casos, sospecha ante la introducción de consideraciones *a priori* en la argumentación filosófica.
- (iv) Falibilismo, *i.e.*, reconocimiento de que cualquiera de nuestras creencias puede ser, en principio, errónea; no hay verdades incontestables o fragmentos del saber humano que estén exentos de revisabilidad.

A esta todavía parca presentación del naturalismo puede añadirse (con el riesgo de hacer una presentación un tanto comprometida) una descripción de dos temas generales de suma importancia para el naturalismo: el tema ontológico y el metodológico (véase De Caro & MacArthur, 2004: 3-8; y Papineau, 2007: *passim*):

- (A) *El tema ontológico.* El naturalismo parece estar comprometido con una concepción exclusivamente científica de la naturaleza. Esto implica no sólo que la ciencia proporciona *una* imagen verdadera de la naturaleza, sino que es *la única* imagen verdadera.¹⁶ En apoyo de esta idea se cita frecuentemente el gran éxito de la ciencia moderna, el cual se asume como implicando la posibilidad de que sus teorías sean probablemente verdaderos y de que su concepción de la naturaleza sea la única *bona fide* o no problemática. Esto ha guiado en gran medida la investigación en metafísica, bajo la premisa de que los conceptos no naturales que empleamos en la explicación del mundo deben ser: 1) *reducidos* o reconstruidos en términos de términos de tesis naturalistas respetables; o 2) tratados de manera instrumental (como *ficciones útiles*); o 3) *construidos sin*

¹⁶ En este tenor declaraba Sellars: “la ciencia es la medida de todas las cosas; de lo que existe, que existe, y de lo que no existe, que no existe” (Sellars, 1956: §41).

un papel referencial o fáctico; o bien, 4) *eliminados* como manifestaciones de pensamiento precientífico. Estas posturas con respecto a la ontología están en continuidad, principalmente, con dos tesis más generales: la unidad de la ciencia y el fisicalismo.

- (B) *El tema metodológico*. El naturalismo parece involucrar un cambio importante en la concepción tradicional del proceder de la filosofía. Mientras la filosofía tradicional intentaba establecer de manera definitiva las presuposiciones, alcances y límites del conocimiento a partir de consideraciones *a priori*, *i.e.*, buscaba una *fundamentación* del conocimiento científico desde la filosofía (ejemplos de ello son los proyectos cartesiano y kantiano en epistemología), el naturalismo filosófico se concibe, como lo expresara Quine famosa y vagamente,¹⁷ como una empresa que es *continua* con la ciencia y, por ende, que abandona de la meta de una filosofía primera. Esto puede ser traducido en: 1) la negación de la autoridad tradicional de la filosofía, que no ha de ser considerada como un tribunal por encima del conocimiento científico; o 2) la negación de que los resultados de la actividad filosófica tengan un papel de fundamentación con respecto a la ciencia.

Aunque el naturalismo filosófico se propone proporcionar un enfoque con amplias consecuencias para diversas áreas de la filosofía, en lo que sigue nos centraremos primordialmente en una de las formas que puede adoptar cuando se lo aplica a la epistemología.

3.4.2. La epistemología naturalizada y el objetivo descriptivo

Puede verse a la epistemología naturalizada es como un conjunto de puntos de vista de acuerdo con los cuales existe una íntima relación entre la investigación filosófica del conocimiento y la ciencia natural.¹⁸ En este sentido, la epistemología

¹⁷ Esto puede apreciarse especialmente en Quine (1951), (1957), (1975), (1975a) y (1990).

¹⁸ La naturaleza y alcance de esta relación varía de manera importante entre los pensadores que se caracterizan a sí mismos como epistemólogos naturalistas. Oscila entre la reducción de la episte-

naturalizada adopta en buena medida los lineamientos de la estrategia descriptivista. Desde esta perspectiva, la epistemología se propone estudiar al conocimiento como un fenómeno natural (véase, *e.g.*, Quine, 1969: 82-83) y examinar explicaciones sustantivas del mismo como teorías en competencia:

Es mejor considerar a la epistemología, entonces, como una empresa al interior de la ciencia natural. La duda cartesiana no es la forma de comenzar. Reteniendo nuestras creencias actuales acerca de la naturaleza, podemos aún preguntarnos cómo podemos llegar a ellas. La ciencia nos dice que nuestra única fuente de información acerca del mundo externo es por medio del impacto de rayos de luz y moléculas sobre nuestras superficies sensoriales. Estimulados de esta manera, de algún modo desarrollamos una ciencia elaborada y útil. ¿Cómo hacemos esto?, y ¿por qué la ciencia resultante funciona tan bien? Éstas son preguntas genuinas, y ninguna simulación de duda se requiere para apreciarlas. Son preguntas científicas acerca de una especie de primates, y están abiertas a investigación en la ciencia natural, la ciencia misma cuya adquisición está siendo investigada (Quine, 1975: 258).

Algunos defensores de la epistemología naturalizada enfatizan cuestiones metodológicas, argumentando que los epistemólogos deben emplear los resultados y procedimientos de las ciencias que estudian el razonamiento humano para dedicarse a responder preguntas epistemológicas. El punto de vista más radical en esta línea, que parece ser suscrito por Quine (1969: 75, 82; 1975a: 72), recomienda reemplazar la epistemología tradicional por el estudio psicológico de cómo razonamos. Una perspectiva más modesta recomienda que los filósofos empleen los resultados de las ciencias para resolver cuestiones epistemológicas. Una forma más bien distinta de epistemología naturalizada trata de asimilar el contenido de enunciados paradigmáticamente epistemológicos a un vocabulario científico. Haciendo eco del tema ontológico, los defensores de esta clase de epistemología

mología al estudio científico de la cognición (la cual equivale, virtualmente, a su eliminación) y posiciones más moderadas, que se limitan a reconocer cierta continuidad entre ciencia y filosofía. El enfoque que aquí privilegiamos pertenece a este último grupo.

naturalizada proponen explicaciones de estos enunciados enteramente en términos de objetos y propiedades científicamente respetables (véase Feldman, 2001).

La importancia de las afirmaciones de los defensores de la epistemología naturalizada puede ser apreciada mejor al verlas como una reacción frente a los métodos y concepciones que predominaron bajo la hegemonía del proyecto epistemológico cartesiano. Así como éste proporcionaba una concepción unificada de los problemas del conocimiento y de las formas de enfrentarlos, al privilegiar el objetivo descriptivo el proyecto naturalista ofrece una concepción de las tareas de una teoría filosófica del conocimiento. A diferencia del cartesiano, el epistemólogo naturalista no se propone proporcionar un fundamento al resto de nuestro conocimiento, validar o justificar nuestras creencias objetivas al derivarlas, mediante pasos autoevidentes, a partir de fundamentos autoevidentes; tampoco considera que esto sea requerido para ofrecer la clase de orientación epistémica que demanda el objetivo normativo. Su propósito es, más bien, ofrecer una explicación general y simple de cómo nos hemos formado la imagen del mundo que tenemos, a partir de la mejor teoría de la que disponemos: la ciencia actual.

A menudo se critica a la epistemología naturalizada por abandonar los objetivos normativo y de validación que caracterizaban a la epistemología cartesiana. El rechazo de estos objetivos, se arguye, parece constituir un ‘cambio de tema’ más una solución alternativa a los problemas de la teoría del conocimiento. Esta objeción no es del todo injustificada, ya que, al privilegiar la estrategia descriptiva, la epistemología naturalizada reformula, o llanamente rechaza, buena parte de los supuestos y problemas que el epistemólogo cartesiano consideraba prioritarios. Sin embargo, no debería suponerse, sin más, que el abandono del proyecto cartesiano constituye una renuncia a responder a los cuestionamientos y objetivos que lo motivaban, y por ende un rechazo de la tarea de ofrecer una teoría filosófica del conocimiento. Más que un cambio de tema, se trata de un cambio en el énfasis. Los problemas y objetivos tradicionales persisten, con estatus clari-

ficado. Al discutir la polémica en torno al externalismo sobre la intencionalidad, Robert Stalnaker ha capturado, de manera ejemplar, la naturaleza del desacuerdo. Se trata del contraste

...entre dos clases de proyecto filosófico [...] un contraste entre decisiones acerca de dónde comenzar, entre diferentes suposiciones sobre lo que es no problemático y sobre cómo caracterizar los problemas filosóficos centrales. El internalista cartesiano comienza con los contenidos de su mente con lo que encuentra por introspección y reflexión. Esto es lo no problemático; éstas son las cosas y los hechos que conocemos directamente. El problema del internalista es, entonces, cómo avanzar más allá de esto para formar una concepción de un mundo externo, y cómo somos capaces de saber que el mundo más allá de nosotros responde a las concepciones que nos formamos. El externalista, en contraste, propone que comencemos con el mundo en el que nos encontramos, y con lo que el sentido común o nuestras mejores teorías científicas nos dicen sobre él. Entre las cosas que encontramos están los seres humanos nosotros mismos que son cosas que (parece) pueden saber acerca del mundo, puede tener experiencias de él, tener un punto de vista en él. Nuestro problema es explicar cómo nuestra concepción objetiva del mundo puede ser una concepción de un mundo que contiene cosas como nosotros, que son capaces de pensar acerca del mundo y tener experiencias de él de la manera en que nosotros lo hacemos (Stalnaker, 2008: 2-3).

El epistemólogo naturalista rechaza, de entrada, el escepticismo radical. Asume también que, dada nuestra situación, el conocimiento es posible, al reconocerlo como un fenómeno natural. Si estamos en condiciones de emprender una investigación epistémica, es en virtud de que tenemos conocimiento; la tarea de una teoría epistemológica es ofrecer una explicación de este fenómeno. Con este fin, el epistemólogo naturalista desplaza el problema de cómo justificar nuestras creencias objetivas a partir de algo cierto e inmediato a un problema más básico, que subyace en su trasfondo: ¿cómo llegamos a tener el concepto de objetividad en

primer lugar?⁹ El problema no es cómo podemos determinar si nuestras creencias son verdaderas, sino cómo es posible el pensamiento objetivo, sea verdadero o falso:

Puede ser que la pregunta epistemológica sea resuelta si podemos responder a la aparentemente más simple pregunta de cómo es posible el pensamiento. Si podemos comprender qué hace posible al error, puede que entonces veamos cómo, dada la existencia del pensamiento, debe ser el caso que muchas de nuestras creencias sean verdaderas y estén justificadas, y por ende constituyan conocimiento (Davidson, 1995: 4).

Al enfrentar el problema desde esta perspectiva, el epistemólogo naturalista comienza *in medias res*, asumiendo que tenemos, a grandes rasgos, una concepción correcta de nuestro entorno y de otras personas, con pensamientos propios; luego se propone ofrecer una descripción de la naturaleza y los orígenes de tal conocimiento (véase, *e.g.*, Quine, 1960: 4-5 y Davidson, 1995: 5). A diferencia del epistemólogo cartesiano, el naturalista adopta la perspectiva de la tercera persona; *i.e.*, la concepción de una perspectiva compartida en un mundo público común. No se busca validar ‘desde dentro’ las creencias que cada uno de nosotros posee, sino dar cuenta, con los recursos que nos ofrece nuestra previa comprensión del mundo, de cómo llegamos a tenerlas y bajo qué condiciones se consideran un logro cognitivo.¹⁹ No obstante, esta transformación no necesariamente implica el abandono de los otros objetivos de la epistemología tradicional.

3.4.3. Falibilismo, holismo, externalismo

¿Qué queda entonces de los objetivos normativo y de validación, y de los problemas que bajo ellos se agrupan? Si lo que guiaba dichos objetivos era dar cuenta

¹⁹ Al respecto Davidson ha señalado: “No acepto la explicación de Quine de la naturaleza del conocimiento, que esencialmente de primera persona y cartesiana. Pero encuentro afinidad con la aproximación resueltamente de tercera persona de Quine a la epistemología, y en la medida que la naturalización de la epistemología alienta o adopta tal aproximación, me da gusto contarme a mí mismo como un epistemólogo naturalista.” (Davidson, 1990: 194)

de nuestras formas de investigación epistémica, comprender la racionalidad en general, entonces los problemas persisten y siguen exigiendo respuesta. No obstante, una vez que han sido subordinados al objetivo descriptivo, su estatuto puede ser reivindicado. Para el epistemólogo naturalista sigue vigente la cuestión de cómo nuestra evidencia se relaciona con nuestra concepción objetiva del mundo; pero no concibe esta tarea como una legitimación de nuestras creencias ante un tribunal supracientífico, sino como una explicación, al interior de nuestra concepción del mundo, de cómo llegamos a tenerlas. La epistemología naturalizada busca también proporcionarnos consejos sobre cómo revisar nuestros métodos de adquisición y retención de creencias, pero asume que para esto hemos de emplear el mejor conocimiento del que disponemos. El problema escéptico no ha desaparecido, pero sólo adquiere claridad al formularlo al interior de la ciencia (véase Quine, 1975: 257-258).

Quizá visitar a Descartes ayude a ilustrar este punto. En el proyecto cartesiano, para explicar cómo es posible el conocimiento es ilegítimo hacer uso de los recursos de la ciencia, de nuestras creencias objetivas acerca del mundo, pues esto involucraría circularidad epistémica. En consecuencia, el intento de responder al escéptico comienza rechazando todas aquellas creencias sobre las que es posible dudar, en virtud de su contenido objetivo. En su punto de partida, el epistemólogo cartesiano no puede disponer de creencias con contenido objetivas. El epistemólogo naturalista concibe la tarea de explicar la posibilidad del conocimiento de manera distinta. Considera que la formulación misma de los problemas escépticos se requiere una representación objetiva de la realidad, una en la que pueda trazarse la diferencia entre apariencia y realidad:

El escepticismo es un vástago de la ciencia. La base para el escepticismo es el percatarse de la ilusión, el descubrimiento de que no siempre debemos creerle a nuestros ojos. El escepticismo se atrinchera en espejismos, en varas aparentemente dobladas en el agua, en arcoíris, imágenes residuales, imágenes dobles, sueños. Pero ¿en qué

sentido son éstas ilusiones? En el sentido de que parecen ser objetos materiales, lo que de hecho no son. Las ilusiones son ilusiones sólo en relación con la aceptación previa de cuerpos genuinos con los cuales contrastarlas. En un mundo de datos de los sentidos inmediatos sin cuerpos postulados ni preguntas formuladas, una distinción entre realidad e ilusión no tendría lugar. La postulación de cuerpos es ya ciencia física rudimentaria; y es sólo después de esa etapa que la distinción desigual del escéptico puede tener sentido. Los cuerpos deben ser postulados antes de que pueda haber un motivo, sin importar cuán tenue, para consentir un mundo que no se restrinja a lo inmediatamente dado (Quine, 1975: 257).

La formulación misma del problema del escepticismo hace una apelación a nuestra representación objetiva del mundo, para después expulsarla del proyecto epistémico. Si el escrúpulo en contra de la circularidad epistémica es consecuentemente asimilado, entonces la duda no puede constituir el punto de partida. Visto de esta manera, el problema del escepticismo resulta espurio, pues ni siquiera puede formularse. Si se abandona el intento de fundamentar nuestro conocimiento sobre una base más firme, el objetivo de validación puede adquirir una nueva forma. La pregunta de cómo es posible el conocimiento se vuelve entonces una pregunta empírica, que reclama una explicación. Para investigarla, podemos emplear legítimamente los recursos de la ciencia. Como afirmara Quine en su «*Epistemology Naturalized*»:

Si el objetivo del epistemólogo es la validación de los fundamentos de la ciencia empírica [de nuestra representación objetiva del mundo], el uso de la psicología o de otra ciencia empírica en esa validación traicionaría su propósito. Sin embargo, estos escrúpulos contra la circularidad tienen escasa importancia una vez que hemos dejado de soñar en deducir la ciencia a partir de observaciones. Si lo que buscamos es simplemente comprender el vínculo entre observación y ciencia, es recomendable que usemos cualquier información disponible, incluyendo aquella proporcionada por la ciencia misma cuyo vínculo con la observación estamos tratando de comprender (Quine, 1969: 75-76).

De este modo, en la investigación de nuestros logros epistémicos hemos recurrir al resto del arsenal metodológico y epistémico del que disponemos. Es así como se conecta la investigación epistemológica con el resto de nuestro saber: no en una jerarquía, que pueda poner en suspenso nuestras creencias buscando su fundamento, sino como una tarea que puede ser emprendida dado el hecho de que tenemos dichas creencias.

El rechazo de la tentativa de validar universalmente nuestra concepción objetiva del mundo tiene consecuencias para la naturaleza y la estructura de la justificación. Una vez que abandonamos la idea de que el conocimiento debe resistir cualquier duda (pues es el conocimiento lo que la hace posible), podemos reconocer que el conocimiento es posible incluso si se apoya en razones falibles para la creencia. El falibilismo sobre la justificación tiene mucho que lo recomiende. Como vimos en la sección 3.2.2, el atractivo intuitivo del infalibilismo sobre la justificación es producto de un razonamiento falaz. El requisito de que el conocimiento involucre creencias verdaderas no es equivalente a la afirmación de que sólo constituyen conocimiento aquellas creencias que sean demostrablemente verdaderas. El falibilismo nos permite preservar la idea de que la verdad es objetiva, *i.e.*, la idea de que “...la verdad de una creencia u oración es independiente de si está justificada por toda nuestra evidencia, es creída por nuestros prójimos o es buena para conducirnos” (Davidson, 1997a: 7). Asimismo, tiene consecuencias para el objetivo normativo. Si la verdad es objetiva, entonces el hecho de que una creencia sea verdadera no está determinado por la clase de consideraciones en las que nos apoyamos para determinar lo que debemos creer; la verdad no es una norma de la investigación (el precepto: “Procura tener creencias verdaderas” es vacuo).

Podría pensarse que el falibilismo tiene como consecuencia la conclusión escéptica. Después de todo, si cualquiera de nuestras creencias con contenido objetivo, considerada en aislamiento, puede ser falsa, entonces no parecería

haber una razón que evitara la posibilidad de que todas las creencias de esta índole puedan ser, en conjunto, falsas. Pero este razonamiento es defectuoso por las mismas razones que invitaban al infalibilismo acerca de la justificación. Supone que el contenido objetivo de una creencia está determinado individualmente, en aislamiento de otros estados mentales. Si la justificación de una creencia debe mostrar que, en caso de que sea verdadera, su verdad no es fortuita, parecería que de algún modo debe implicar su verdad. Pero ni las sensaciones ni las creencias pueden implicar la verdad de otras creencias, pues las sensaciones no pueden entrar en relaciones lógicas y las creencias pueden ser falsas. Nos encontramos ante la Escila y la Caribdis del infalibilismo y del escepticismo.

El epistemólogo naturalista sugiere que resolvamos este falso dilema “...mostrando que alguien con un conjunto (más o menos) coherente de creencias tiene una razón para suponer que sus creencias no están equivocadas en su mayoría” (Davidson, 1983: 146); por ende, tiene una razón para creer que la verdad de sus creencias, en caso de que sean verdaderas, no es fortuita. Esto puede apreciarse con más claridad si adoptamos la perspectiva de la tercera persona, como sugiere el naturalista. Si alguien hace una aserción de la oración “Hay un ratón en mi cocina”, la creencia que con ella expresa puede ser verdadera o falsa dependiendo de dos factores: lo que significan sus palabras y cómo sea el mundo (véase, *e.g.*, Quine, 1951: 36). Podría pensarse que su creencia sería falsa si no hubiese un roedor en su cocina; pero esto sólo sería el caso si su preferencia de “ratón” significase “roedor”, *i.e.*, si creyera que los ratones son roedores. Sólo podríamos determinar la verdad o la falsedad de sus creencias si podemos determinar su contenido y éste está, en parte, determinado por las relaciones inferenciales con otras creencias. Ésta es una forma de decir que el contenido de las creencias es holístico: una creencia con contenido objetivo, sea verdadera o falsa, depende para su identificación de un trasfondo de creencias verdaderas; los conceptos que caracterizan el contenido objetivo de una creencia no pueden ser los mismos sin importar de cuáles

creencias formen parte. Debido al carácter holístico de la creencia, es imposible que todas nuestras creencias con contenido objetivo sean falsas.

Adoptar la perspectiva de la tercera persona tiene otra consecuencia importante. Al tratar de darle sentido a las preferencias de otra persona, los únicos recursos de los que disponemos son la observación de su conducta y del mundo al que observamos que reacciona. Para investigar el significado de sus preferencias y, por ende, el contenido de sus creencias debemos prestar atención a las interacciones causales entre esa persona y su entorno, *i.e.*, a los factores externos en su entorno que causan su conducta (en especial, su conducta lingüística). Si asumimos que estos factores determinan el contenido de las creencias del hablante, la última trinchera del escepticismo ha sido superada.²⁰ Incluso las creencias falsas deben ser formuladas en términos de conceptos cuyo contenido se establezca mediante sus conexiones con el mundo, en el contexto de otras creencias que sean verdaderas (*cf.* Nagel, 1999: 197). De este modo, si tenemos creencias coherentes con contenido un objetivo, suministrado por el mundo externo, esto proporciona una explicación de cómo es posible el conocimiento.

Visto desde la perspectiva naturalista, el problema del regreso epistémico cobra un nuevo significado; pero ahora es positivo. Se ha renunciado a ofrecer una fundamentación última de nuestras creencias; en la medida en que su contenido es objetivo, cualquiera de ellas puede ser falsa y es, por ende, susceptible de

²⁰ Se ha insistido con frecuencia en que Davidson, quien emplea un argumento a favor del externalismo como respuesta al problema del escepticismo sobre el mundo externo, proporciona un argumento trascendental (véase, *e.g.*, Sosa, 2003 y Carpenter, 2003). Sugiero que esto es producto de una profunda incompreensión de la metodología naturalista. No se debería considerar que el argumento a partir del externalismo es *a priori* y suministra, en línea con el objetivo de validación, una *demonstración* de la hipótesis de que tenemos conocimiento. Para empezar, una de las premisas del argumento es empírica: parte de la observación de pensamientos. Asimismo, he sugerido en la sección 2.3 que no se debería considerar que Davidson procede ofreciendo un argumento trascendental, sino una inferencia a la mejor explicación. Esta interpretación parece ser respaldada por la réplica que Davidson ofrece a Nagel: “Es así como el argumento, en su forma más directa, procede. Comienzo con la *observación empírica* de que el pensamiento, el pensamiento proposicional, ocurre, y formulo dos preguntas relacionadas: ¿*qué explica* el hecho de que los pensamientos sean objetivos, esto es, que su verdad (generalmente) sea independiente de que sean creídos?; ¿y *qué explica* los contenidos de estos pensamientos?” (Davidson, 1999: 207. Cursivas mías).

revisión. El falibilismo es la contraparte naturalista del infinitismo sobre la justificación: ninguna cantidad de apoyo inferencial es suficiente para demostrar la verdad de una creencia particular, cada una de ellas puede ser examinada indefinidamente, pues nuestra condición epistémica es perfectible. No obstante, la posibilidad de que alguna de nuestras creencias tenga contenido que tiene depende de su relación con otras creencias, que deben ser, en su mayoría, verídicas. Las credenciales epistémicas de una creencia están determinadas por sus relaciones con otras creencias, que pueden entrar en relaciones de inferencia con ella. El holismo es la contraparte naturalista del coherentismo sobre la justificación. Finalmente, una intuición de fundacionismo pervive en el externalismo. Nuestras creencias están conectadas con el mundo a través de nuestra percepción y nuestras sensaciones, pero la conexión es causal, no epistémica (lo que no equivale a decir que no tenga consecuencias epistémicas).

Se ha sugerido abandonar el proyecto epistemológico cartesiano, y su intento de fundamentar nuestro conocimiento en el contenido de la propia mente. Pero, si ha de abandonarse la perspectiva cartesiana, ¿queda aún lugar para el conocimiento de la propia mente en la empresa epistemológica? En la siguiente sección se examina una manera de ubicar la concepción del conocimiento de la propia mente defendida en los primeros capítulos de esta disertación al interior de la empresa epistemológica.

3.5. El conocimiento de la propia mente en la epistemología naturalizada

Desde la época de Descartes, gran parte de la discusión epistemológica ha adoptado la perspectiva de la primera persona, ubicando los fundamentos del conocimiento en el contenido de la propia mente. Desde esta perspectiva, el conocimiento de las propias sensaciones y creencias se ha considerado como básico y como capaz de

validar aquellos de nuestros logros cognitivos que no disfrutaban de su relativa certeza. La epistemología naturalizada sugiere una reconsideración de esta imagen. No se pregunta ya cómo podemos justificar, hacer inmunes ante la duda, nuestras creencias objetivas sobre el mundo y las otras mentes, sino, de manera más general, cómo podemos tener pensamientos con contenido objetivo en absoluto.

Adoptando una actitud afín a la del epistemólogo naturalista, Davidson sugiere enfrentar el problema del conocimiento desde la perspectiva de la tercera persona. Su punto de partida no es, pues, el dominio interno, privado, del pensamiento, sino nuestra concepción de una perspectiva compartida en un mundo público común. Dicha perspectiva presupone la posibilidad de la comunicación humana. Lo que se busca es mejorar nuestra comprensión de cómo es posible la comunicación, el mutuo entendimiento, e investigar sus consecuencias para el conocimiento.

3.5.1. Tres variedades de conocimiento

Para Davidson, el propósito central del lenguaje es la comunicación, es transmitir a un oyente lo que uno piensa por medio de palabras que el oyente comprende como uno quiere.²¹ La comunicación es exitosa cuando el oyente interpreta las preferencias del hablante como éste lo desea, en un contexto particular. ¿Cómo es posible este resultado? Davidson argumenta que lo que explica la posibilidad de la comunicación es la convergencia de tres clases de conocimiento íntimamente interrelacionadas: conocimiento de otras mentes, conocimiento del mundo externo y conocimiento de la propia mente. La comunicación requiere que seamos capaces de atribuirles correctamente pensamientos a otras personas, que tengamos una imagen del mundo, en gran medida, verdadera y que sepamos, la mayor parte del tiempo, cuáles son nuestros pensamientos. La contribución de cada una de estas formas de conocimiento es lo que mejor explica la posibilidad de la comunicación

²¹ “El lenguaje tiene un sinfín de propósitos distintos, pero ninguno subyace a éste” (Davidson, 1994: 120).

y, por ende, de una perspectiva compartida. Para ver esto, podemos tratar de aislar las contribuciones relativas de estas formas de conocimiento.

Que la comunicación sea exitosa depende que seamos capaces de identificar lo que significan las palabras del hablante. Habitualmente se asume que sabemos esto, si somos hablantes competentes de la misma lengua. No obstante, esto sólo desplaza el problema a determinar cómo sabemos que somos hablantes del mismo lenguaje, que las mismas expresiones deben ser interpretadas de la misma manera. Davidson sugiere considerar la situación de dos sujetos que interactúan lingüísticamente en un entorno compartido, sin previo conocimiento de los significados de las preferencias de sus interlocutores. Se trata de una situación en la que la comunicación requiere interpretación radical.²² En esta situación, determinar lo que el hablante quiere decir con sus preferencias equivale a saber cuáles son sus pensamientos:

Existen muchas razones para sostener [...] que establecer la corrección de una atribución de creencias no es más fácil que interpretar el habla de un hombre. Pero creo que podemos ir aún más lejos y decir que los problemas son idénticos. Las creencias no pueden descubrirse, en general, si no se domina el lenguaje de una persona; y no podemos dominar el lenguaje de alguien sin conocer mucho de lo que cree (1974: 238).

Una vez que se ha adoptado la actitud naturalista, y con ella la perspectiva de la tercera persona, se ha abandonado la búsqueda de una validación última de la posibilidad del conocimiento, y éste se ha situado en el dominio común de lo públicamente compartido. Por ende, “[l]o que un intérprete informado puede aprender acerca de lo que un hablante quiere decir es todo lo que hay que saber; lo mismo se aplica a lo que el hablante cree” (Davidson, 1983: 148).

²² Aunque esta situación parece, en principio, peculiar, Davidson sugiere que es ubicua al asumir la primacía del idiolecto del hablante sobre los lenguajes compartidos: “[u]oda comprensión del habla de otro involucra interpretación radical” (Davidson, 1973a: 125).

Pero, ¿de qué recursos dispone el intérprete para emprender el proyecto de interpretar radicalmente al hablante? Lo mejor que el intérprete puede hacer es adoptar el recurso interpretativo que Davidson denomina ‘principio de caridad’. Éste combina dos rasgos.²³ Por una parte, involucra una suposición externalista acerca del contenido de las creencias del hablante: la suposición de que, en los casos más básicos, los objetos en el entorno del hablante y a los que éste reacciona proporcionan el contenido de sus creencias. Por otra parte, el principio de caridad introduce una suposición holista sobre racionalidad en las creencias del hablante: la suposición de que sus preferencias y acciones son suficientemente racionales para hacerlas inteligibles.

Sobre la base de estas suposiciones, el intérprete puede ofrecer una teoría empírica del significado de las preferencias del hablante y del contenido de sus creencias: “...cuando el intérprete encuentra una oración del hablante a la que éste asiente con regularidad, bajo condiciones que el intérprete reconoce, el intérprete considera a esas condiciones como las condiciones de verdad de la oración del hablante” (Davidson, 1983: 149). Asimismo, cuando el intérprete identifica una estructura recursiva en las preferencias del hablante, los rasgos finitos a partir de los cuales el hablante puede combinar un número ilimitado de preferencias significativas en su lenguaje, formula hipótesis de la contribución semántica de estos rasgos y sus relaciones de inferencia. Las atribuciones de creencias al hablante, así como las hipótesis empíricas sobre el significado de sus preferencias, que se realizan de acuerdo con el principio de caridad son, por supuesto, falibles; no obstante, posibilitan la interpretación radical en la medida en que maximizan el grado de comprensión, de acuerdo, que es requerido como trasfondo para la atribución de creencias falsas o conductas irracionales.

²³ Davidson los denomina ‘Principio de Correspondencia’ y ‘Principio de Coherencia’ respectivamente (véase 1991: 211). He preferido otras denominaciones para hacer explícita la continuidad con el enfoque naturalista presentado en la sección 3.4.

Al examinar la situación del intérprete radical podemos percatarnos de las formas de conocimiento involucradas en la atribución de pensamientos a otras personas. La suposición externalista del principio de caridad asume que el intérprete tiene conocimiento del mundo externo, *i.e.*, que sabe que se encuentra en un mundo, que cohabita con el hablante y al que éste reacciona de diversas maneras, incluida su conducta lingüística. Por su parte, la suposición holista presupone que el intérprete tiene conocimiento de la propia mente, *i.e.*, sabe el significado de los recursos de su lenguaje, el cual cuenta con sofisticados recursos conceptuales, le permite hacer distinciones finas sobre el mundo y calibrar las respuestas lingüísticas del hablante. Estas formas de conocimiento le proporcionan al intérprete una guía metodológica para dar cuenta de las proferencias lingüísticas del hablante y del contenido de sus pensamientos. Podría pensarse que esto equivale a una reducción del conocimiento de otras mentes a conocimiento del mundo externo y de la propia mente. Pero esta conclusión sólo sería correcta si estas últimas formas de conocimiento no dependieran a su vez del conocimiento de otras mentes.

Pero nuestro conocimiento del mundo y de la propia mente depende también del conocimiento de otras mentes. El conocimiento de otras mentes se encuentra en la base del pensamiento objetivo y, por ende, de la posibilidad misma del conocimiento. Esto se hace patente si prestamos atención a las situaciones básicas de triangulación: “[s]e requieren dos para triangular” (Davidson, 1991: 213). Es la interacción con otras personas y el reconocimiento de la conducta de otros como manifestación de sus pensamientos lo que posibilita la noción de objetividad, al introducir algo que cuente como error, y lo que permite determinar el contenido empírico de los pensamientos, al situar las causas que suministran su contenido en los objetos externos donde convergen los interlocutores. El punto crucial de Davidson es que una explicación satisfactoria de la comunicación humana y del pensamiento objetivo requiere el reconocimiento de que estas tres

variedades de conocimiento no son reducibles unas a otras,²⁴ sino que en su interdependencia hacen posible nuestra concepción compartida en un mundo público común:

...nuestro saber proposicional tiene su base no en lo impersonal sino en lo interpersonal. Así, cuando miramos al mundo natural que compartimos con otros no perdemos contacto con nosotros mismos, sino que más bien reconocemos nuestra pertenencia a una sociedad de mentes. Si yo no supiera lo que otros piensan, no tendría pensamientos propios y por ende no sabría lo que pienso. Si yo no supiera lo que pienso, carecería de la habilidad de calibrar el pensamiento de otros. La calibración del pensamiento de otros requiere que viva en el mismo mundo con ellos, compartiendo muchas reacciones a sus rasgos más importantes, incluyendo sus valores. De modo que no hay peligro de que al ver el mundo objetivamente perdamos contacto con nosotros mismos. Las tres clases de conocimiento forman un trípode: si alguna de sus patas se perdiera, ninguna parte permanecería en pie (Davidson, 1991: 219-220).

3.5.2. La indeterminación de lo mental

Aunado al rechazo del proyecto de validación universal de nuestro conocimiento, el cambio de enfoque hacia la epistemología naturalizada tiene consecuencias importantes para la concepción de lo mental. Si el proyecto de la interpretación radical ha de resultar exitoso, será en la medida en que lo que las palabras de un hablante significan, y el contenido de sus pensamientos, sean públicamente determinables. No obstante, afirmar esto no equivale a señalar que el significado de las palabras de un hablante, y el contenido de sus pensamientos, de hecho estén

²⁴ “Hay razones convincentes para aceptar la concepción de que ninguna de las tres formas de conocimiento es reducible a una de las otras o a ambas. [...] Considero que la desesperanza de encontrar modos efectivos de reducción es evidente a partir del rechazo casi universal de los programas reduccionistas estándar. El escepticismo en varias de sus familiares apariencias es nuestro reticente tributo a la evidente imposibilidad de unificar las tres variedades de conocimiento: una forma de escepticismo surge de la dificultad de dar cuenta de nuestro conocimiento del mundo externo sobre la base del nuestro conocimiento de nuestras propias mentes; otra reconoce que nuestro conocimiento de otras mentes no puede consistir sólo en lo que podemos observar desde el exterior. La inextricabilidad del problema mente cuerpo es otro de tales tributos” (Davidson, 1991: 206).

determinados. La discusión tradicional en epistemología asume que el contenido de nuestros pensamientos es fijo, que está determinado; sin este supuesto, no tendría sentido afirmar que podemos tener conocimiento de la propia mente prescindiendo de otras formas de conocimiento. A partir de esta suposición, el epistemólogo cartesiano luego se pregunta cómo podemos extender nuestro conocimiento más allá de este reducido dominio. El diagnóstico escéptico es que esto no es posible. Al intentar preservar la castidad del conocimiento subjetivo, el proyecto cartesiano crea una brecha entre la mente y el mundo que ninguna reconstrucción puede zanjar.

Sin embargo, al reconocer la interdependencia entre las formas de conocimiento, el dominio de lo mental pierde su inocencia. Nuevamente, podemos ilustrar la situación en epistemología si partimos del examen de la situación con respecto al lenguaje. Si asumimos que el significado de las preferencias lingüísticas está determinado de antemano (*e.g.*, a través de reglas o convenciones), surge la cuestión de explicar cómo se establecieron dichas convenciones entre los hablantes primer lugar, sin presuponer entre ellos intercambio lingüístico. Es por ello que Davidson prefiere situar el problema partiendo de la mutua comprensión de los interlocutores en situaciones de interpretación radical. Al hacerlo invierte la dirección habitual de la explicación: considera como básicos la comunicación y la mutua comprensión *vis à vis* las nociones de significado y un lenguaje compartido, si éstas son definidas a través de reglas o convenciones. Así, las nociones de lenguaje y significado son nociones teóricas que dependen de la comunicación exitosa, en lugar de explicarla (véase Röska-Hardy, 2011: 383). Si concebimos a las preferencias lingüísticas como acciones intencionales, el requisito de que sean interpretables implica que deben satisfacer ciertas restricciones racionales y sociales. Dichas restricciones se ubican en la tríada entre hablante, oyente y objetos a los que responden. Para que la comunicación sea posible, el hablante debe hacerse interpretable y esto introduce un factor irreduciblemente

social en sus preferencias, que controla lo que puede querer decir con sus palabras. Muestra por qué un hablante no puede querer decir con sus palabras algo que no pueda ser correctamente descifrado por alguien más. Debe haber otra persona que comprenda lo que el hablante dice para que sus preferencias tengan un significado o contenido específico (véase, *e.g.*, Davidson, 1994). De este modo, el significado de una preferencia lingüística no es privado o subjetivo, sino que surge en situaciones sociales en un entorno compartido.

Si, como sugiere Davidson, el problema de establecer la corrección de una atribución de creencias a un hablante es idéntico al de interpretar el significado de sus preferencias, entonces el contenido del pensamiento depende también situaciones sociales de intercambio lingüístico. Hasta que la comunicación con una segunda persona es establecida, los pensamientos de un individuo no tienen ningún contenido específico *i.e.*, no tienen contenido en absoluto. Sólo en el intercambio lingüístico que da lugar a una comunicación, en gran medida, exitosa es posible determinar el contenido del pensamiento de un hablante. Pero esto puede hacerse de diversas maneras:

Cada uno de nosotros puede pensar en sus propias oraciones (o en sus contenidos) como en los números; tienen múltiples relaciones unas con otras y con el mundo. Manteniendo iguales estas relaciones, podemos equiparar nuestras oraciones con aquellas de un hablante, y con las actitudes [proposicionales] de ese hablante, de diferentes maneras sin cambiar de parecer acerca de lo que el hablante piensa y quiere decir. Así como el sinfín de conjuntos de números nos permite registrar las mismas estructuras complejas en el mundo, así nuestras oraciones pueden usarse en un sinfín de maneras distintas para registrar las actitudes de otros, y el significado de sus oraciones. [...] Esto puede sonar como una tesis negativa, y lo es; es un ataque a la idea de que los significados pueden ser capturados exactamente de una manera, al fijar significados platónicos sobre expresiones (Davidson, 1998a: 317).

El ataque se extiende también a la idea de que hay objetos ‘ante la mente’, objetos del pensamiento que el sujeto conoce o con los que establece familiaridad o está

en contacto.²⁵ En ocasiones se ha sugerido que esto pone en apuros a la perspectiva naturalista, pues parece hacer a lo mental misterioso o irreal,²⁶ siendo que el conocimiento de la propia mente nos sigue pareciendo especialmente seguro. Davidson ha respondido a estas inquietudes señalando que la indeterminación de lo mental, relativa a un lenguaje, no implica la irrealidad de lo mental,²⁷ ni que no haya hechos relevantes acerca de lo que una persona piensa.²⁸ En efecto, la indeterminación tampoco es lo que distingue a lo mental de lo físico: “...la indeterminación [...] es endémica a todas las disciplinas” (Davidson, 1998a: 316). ¿Qué es, entonces, lo que hace especial al conocimiento de los propios pensamientos?

3.5.3. La indispensabilidad del conocimiento de la propia mente

Al final del capítulo 1 se expuso, a grandes rasgos, la explicación davidsoniana del conocimiento de la propia mente. Sobre la base de su desempeño frente a explicaciones sustantivas rivales, se argumentó que esta explicación tiene mucho que la recomiende. No obstante, si el enfoque naturalista en epistemología defendido en este capítulo es viable, la explicación davidsoniana se ve robustecida. Por una parte, la indeterminación de lo mental proporciona apoyo suplementario al rechazo de las concepciones cartesianas del conocimiento de la propia mente, que involucran una forma especial de acceso epistémico, las cuales examinamos en la sección 1.3. Por otra parte, la interdependencia entre el conocimiento del mundo, de otras mentes y de la propia mente respalda el repudio de las explicaciones no epistémicas de la autoridad de la primera persona, contra las que se argumentó en la sección 1.4.

²⁵ Que la indeterminación de la interpretación constituye también un ataque a esta idea es parte de lo que se argumenta en Davidson (1988) y (1989).

²⁶ Esto es lo que parecen sugerir, *e.g.*, Searle, (1987) y Rorty (1998: 55-61).

²⁷ “¿Cómo podría haber para mí una cuestión acerca de la *ontología* de las entidades mentales si, como sostengo, son idénticas con entidades que también describimos y explicamos, en diferentes términos, en las ciencias naturales?” (Davidson, 1998: 316). La identidad ontológica, sin reducción explicativa, de lo mental con lo físico se defiende en Davidson (1970), (1973), (1974) y (1993a).

²⁸ “...[L]os hechos son las relaciones empíricas entre un hablante, sus oraciones y su entorno. Este patrón es invariante.” (Davidson, 1998a: 317).

En la epistemología davidsoniana, el conocimiento de la propia mente se encuentra en la base del resto de nuestro conocimiento. Pero esto no quiere decir que sea su fundamento, pues también requiere y recibe justificación desde fuera de la perspectiva de la primera persona:

Hay por su puesto algunas creencias que poseen un alto grado de certeza, y en algunos casos su contenido crea una presuposición a favor de su verdad. Éstas son las creencias acerca de nuestras propias actitudes proposicionales presentes. Pero la relativa certeza de estas creencias no las hace apropiadas para ser el fundamento del conocimiento empírico. Surge, más bien, de la naturaleza de la interpretación. Como intérpretes tenemos que tratar las autoadscripciones de creencia, duda, deseo, y actitudes similares como privilegiadas; éste es un paso esencial en la interpretación del resto de lo que la persona dice y piensa. Los fundamentos de la interpretación no son los fundamentos del conocimiento, aunque una apreciación de la naturaleza de la interpretación puede llevar a una apreciación de la naturaleza esencialmente verídica de la creencia (Davidson, 1982: 175).

¿Cuál es la contribución distintiva del conocimiento de la propia mente a la empresa epistemológica? La respuesta a esta pregunta se desprende de la doble contribución del conocimiento de la propia mente en la tarea de la interpretación radical. Por una parte, el conocimiento de la propia mente se hace manifiesto en el hablante, a quien hemos de interpretar como intencionalmente proporcionando pistas para que el intérprete descifre su conducta, *i.e.*, como usando "...un repertorio finito de sonidos distinguibles aplicados consistentemente a objetos y situaciones que él cree que son evidentes para su oyente" (Davidson, 1984: 13). Por otra parte, el conocimiento de la propia mente también desempeña un papel desde la perspectiva del intérprete, quien dispone del conocimiento de cómo su lenguaje se relaciona con el mundo e impone sus estándares de racionalidad como restricciones a los pensamientos del hablante.

Como ya hemos señalado, la indispensabilidad del conocimiento de la propia mente puede apreciarse al notar que la interpretación requiere la suposición

de que el hablante sabe lo que quiere decir cuando dice que cree algo. Esto es importante porque pone de relieve lo que distingue al conocimiento de la propia mente de las otras formas de conocimiento, *i.e.*, la autoridad de la primera persona. Sin embargo, visto de esta manera “[e]l conocimiento de los contenidos de nuestras propias mentes debe, en la mayoría de los casos, ser trivial. La razón es que, aparte de casos especiales, el problema de la interpretación no puede surgir [para el hablante]” (Davidson, 1991: 217). Una contribución más sustantiva se hace manifiesta si consideramos el conocimiento de la propia mente desde la perspectiva del intérprete. Su indispensabilidad puede ser apreciada al reflexionar sobre lo que caracteriza a la indeterminación de lo mental frente a otras formas de indeterminación. Que las actitudes proposicionales están indeterminadas (que podemos describirlas correctamente de muchas maneras distintas) no es peculiar de lo mental; lo que distingue a las actitudes proposicionales es una indeterminación adicional: que involucran estándares normativos de racionalidad, *i.e.*, lo que Davidson denominó “el ideal constitutivo de la racionalidad” (véase 1970: 223 y 1998a: 318-319). En este punto el conocimiento de la propia mente, esta vez, por parte del intérprete, desempeña también un papel importante. Al interpretar a un hablante como actuando racionalmente (*e.g.*, en la comunicación lingüística), lo interpretamos bajo nuestras propias luces: calibramos su pensamiento de sobre la base de nuestras propias creencias sobre el mundo objetivo y en el trasfondo de nuestros propios estándares de racionalidad. En esta sutil calibración, nuestras habilidades interpretativas son puestas a prueba: “[e]s aquí que la contribución única del punto de vista de la primera persona [...] se vuelve más evidente” (Davidson, 1998: 89). Se trata de “...un hecho tan obvio que puede pasar desapercibido: [...] los estándares de racionalidad y realidad de los que dependo al comprender a otros son los míos, y [...] no puede haber una apelación más allá de ellos” (Davidson, 1998: 91).

Conclusiones

It should not be assumed that [...] all the problems of epistemology will evaporate. But the problems that seem salient will change. [...] Plenty of questions of equal or greater interest [than those of traditional epistemology] will remain, or be generated by the new stance. [...] They are questions that, like the epistemological questions they replace, have no final answer; but unlike the questions they replace, they are worth trying to answer.

(Davidson, 1988: 47)

En esta tesis me he propuesto defender la concepción davidsoniana del conocimiento de la propia mente, mostrando que ofrece una explicación robusta de la intuición de que cada uno de nosotros sabe, de manera especial, cuáles son sus pensamientos en el momento que se los adscribe.

Con este fin, en el capítulo 1 intenté precisar qué aspectos del conocimiento de la propia mente resultan más relevantes al intentar explicarlo. Éstos se reducen a dos. Primero, la autoridad de la primera persona, *i.e.*, la presunción de que cuando alguien se adscribe estados mentales, su atribución es verdadera y justifica la creencia de que está en lo correcto. Segundo, la asimetría epistémica, *i.e.*, el carácter no mediado por la observación de la conducta e inferencia que exhiben las atribuciones de estados mentales en primera persona. La autoridad de la primera persona y la asimetría epistémica son rasgos característicos del conocimiento de la propia mente, que no se manifiestan en ningún otro de nuestros logros cognitivos. Resultan inquietantes y ameritan examen.

Me di a la tarea de mostrar que las explicaciones del conocimiento de la propia mente que tienen mayor influencia en la discusión actual no dan cuenta adecuadamente de estos rasgos. La apelación a una condición distintiva de acceso

epistémico privilegiado o bien no explica cómo las autoadscripciones de estados mentales justifican la atribución correspondiente o bien no aclara por qué dichas adscripciones tienen un estatus especial al realizarse en primera persona. Aunado a esto, el acceso privilegiado como recurso explicativo puede resultar un caso de *obscurum per obscurius*. Las explicaciones no epistémicas, por otra parte, resultan insuficientes para dar cuenta de los rasgos que se proponen explicar. Además, ponen en tela de juicio la univocidad de los conceptos psicológicos.

Con este trasfondo, la explicación davidsoniana tiene mucho que la recomiende. Muestra cómo los rasgos distintivos del conocimiento de la propia mente pueden ser elucidados al reflexionar sobre la naturaleza de la interpretación. Esto preserva sus rasgos epistémicos y da cuenta de por qué son propios de esta forma de conocimiento.

En el capítulo 2 intenté mostrar que la concepción davidsoniana del conocimiento de la propia mente puede enfrentar otros retos teóricos, que se amenazan la condición epistémica de la autoridad de la primera persona. Por una parte, el externalismo sobre la intencionalidad parece poner en entredicho la suposición de que un hablante sabe cuál es el significado de sus palabras y el contenido de sus pensamientos. Por otra parte, las situaciones en las que un individuo manifiesta irracionalidad, como el autoengaño, parecen desacreditar la presunción de que un hablante está en lo correcto cuando se adscribe un estado psicológico.

Argumenté que el externalismo, en la formulación davidsoniana, es compatible con la autoridad de la primera persona. Asimismo, ofrecí razones para preferir el externalismo que Davidson suscribe a las formulaciones ortodoxas de esta tesis. En lo que concierne al autoengaño, defendí que la presencia de creencias conflictivas en un individuo no socava la autoridad de la primera persona. Vistos desde otra perspectiva, el externalismo y el autoengaño robustecen la conclusión de que el conocimiento de la propia mente es un logro epistémico, al presentar situaciones en las que se falla en su consecución.

En el capítulo 3 se exploraron las consecuencias para la epistemología del abandono de la concepción tradicional del conocimiento de la propia mente. Traté de mostrar que esta concepción se encuentra en el núcleo de lo que denomino ‘epistemología cartesiana’; asimismo, se ofrecieron razones para diagnosticar el fracaso de este proyecto filosófico. Pero la epistemología cartesiana es sólo una de las muchas formas que puede adoptar la teoría del conocimiento.

Davidson sugiere que emprendamos otra vía: una forma de epistemología naturalizada. Este proyecto da cabida a una concepción reformulada del conocimiento de la propia mente, y nos permite investigar sus relaciones de interdependencia con otras formas de conocimiento. Si Davidson está en lo correcto, la concepción cartesiana del conocimiento de la propia mente es el producto de una confusión, y da origen a muchas otras. Su propuesta resulta atractiva, pues promete poner fin a los que, quizá, han sido los problemas centrales de la tradición epistemológica predominante.

No obstante, con esto no se pretende declarar los obituarios de la investigación filosófica del conocimiento. Nuevos problemas surgen, o bien los antiguos persisten, con un estatus clarificado. Quizá tampoco tengan una solución definitiva. Pero esto no debería resultar demasiado inquietante; después de todo, que había respuestas concluyentes, perentorias, era uno de los supuestos de la tradición cartesiana. La epistemología naturalizada nos pone en la posición más ventajosa de mejorar nuestra comprensión del mundo y del ser humano, evitando la catatonia filosófica.

El cartesianismo puede aún persistir, modificando parte de su agenda original. Su prolija descendencia satura, en nuestros días, el panorama filosófico. Pueden reclamarle al naturalista que, al abandonar la perspectiva de la primera persona, niegue los hechos que, para cada uno de nosotros, resultan obvios. Así, *e.g.*, John Searle ha señalado: “[e]n todas las discusiones en la filosofía del lenguaje y la filosofía de la mente, es absolutamente esencial en algún punto recordarse a sí mismo

del caso de la primera persona. Nadie puede, por ejemplo, convencernos por medio de argumentos, sin importar cuán ingeniosos, de que los dolores no existen si de hecho los tenemos” (Searle, 1987: 126). Esta protesta, de ser correcta, socavaría el proyecto naturalista. No obstante, quizá sea más prudente sugerir que la predilección de los filósofos naturalistas por la perspectiva de la tercera persona no es producto de una negación de la subjetividad, sino de un intento por situarla como objeto de investigación.

No sugiero que haya argumentos demostrativos a favor de la posición davidsoniana, ni que sea intocable o definitiva. Pero tampoco creo que el interés y la importancia de la obra filosófica de Donald Davidson dependan de la existencia de argumentos de esa índole. Podría ser que, sobre asuntos tan fundamentales como los que trata, no haya argumentos concluyentes; quizá la elección entre las diversas actitudes sobre cómo afrontar los problemas filosóficos deba medirse en términos de qué tanto se ajustan a la totalidad de nuestras intuiciones prefilosóficas. En todo caso, me parece que, aunque perfectible, la obra davidsoniana en sus intrincadas interconexiones ofrece un horizonte apropiado para formular y afrontar los problemas filosóficos. Considerada en su conjunto, proporciona elementos que propician la comprensión de nuestra posición epistémica en el mundo.

Referencias

Advertencia: Con el fin de preservar la cronología de las fuentes, se consignan las obras por el año de su primera aparición. No obstante, a menudo el texto citado se retoma de compilaciones más recientes. Se indica la fuente de la que se extraen los pasajes con el símbolo ‘*’ a la izquierda del año de publicación.

- A Companion to Descartes.* (2008). J. Broughton & J. Carriero (eds.). Malden: Blackwell Publishing Ltd.
- A Companion to the Philosophy of Mind.* (1994). Samuel Guttenplan (ed.). Malden: Blackwell Publishing Ltd.
- A. J. Ayer: Memorial Essays.* (1991). A. Phillips Griffiths (ed.). Cambridge: Cambridge University Press.
- Argumentos trascendentales.* (1999). Isabel Cabrera Villoro (comp.). México: UNAM-IIF.
- Armstrong, David M. (1968). *A Materialist Theory of the Mind.* London: Routledge.
- Bach, Kent. (2009). Self-Deception. In *The Oxford Handbook of Philosophy of Mind* (2009: 781-796).
- Barnes, Annette. (1997). *Seeing Through Self-Deception.* New York: Cambridge University Press.
- Bar-On, Dorit. (2004). *Speaking My Mind. Expression and Self-Knowledge.* New York: Oxford University Press.
- Bermúdez, Jose Luis. (2000). Self-Deception, Intentions and Contradictory Beliefs. *Analysis*, 60(4): 309-319.
- Bilgrami, Akeel. (1992). *Belief and Meaning. The Unity and Locality of Mental Content.* Oxford: Blackwell Publishers Ltd.
- Bilgrami, Akeel. (2006). *Self-Knowledge and Resentment.* Cambridge: Harvard University Press.
- Boghossian, Paul A. (1989). Content and Self-Knowledge. *Philosophical Topics*, 17: 5-26. Reprinted in *Externalism and Self-Knowledge* (1998: 149-173).*
- Bonjour, Laurence & Sosa, Ernest. (2003). *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues.* Malden: Blackwell.
- Bonjour, Laurence. (1985). *The Structure of Empirical Knowledge.* Cambridge: Harvard University Press.
- Bonjour, Laurence. (2003). Back to Foundationalism. In Bonjour & Sosa (2003: 60-76).
- Brandom, Robert B. (2000). *Articulating Reasons: An Introduction to Inferentialism.* Cambridge: Harvard University Press.

- Brentano, Franz Clemens. (1874). *Psychologie vom empirischen Standpunkt*. Leipzig: Duncker & Humbolt. [Versión castellana: Brentano, Franz. (1935)*. *Psicología desde un punto de vista empírico*. Trad. de José Gaos. Madrid: Revista de Occidente.]
- Broughton, Janet. (2008). Self-Knowledge. In *A Companion to Descartes* (2008: 179-195).
- Brown, Jessica. (1995). The Incompatibility of Anti-Individualism and Privileged Access. *Analysis*, 55: 149-156. Reprinted in *Externalism and Self-Knowledge* (1998: 185-193).*
- Brown, Jessica. (2004). *Anti-Individualism and Knowledge*. Cambridge: The MIT Press.
- Brown, Jessica. (2009). Semantic Externalism and Self-Knowledge. In *The Oxford Handbook of Philosophy of Mind* (2009: 767-780).
- Brueckner, Anthony. (1997). Is Skepticism About Self-Knowledge Incoherent? *Analysis*, 57: 287-290.
- Brueckner, Anthony. (2007). Externalism and Privileged Access are Inconsistent. In *Contemporary Debates in the Philosophy of Mind* (2007: 37-52).
- Brueckner, Anthony. (2012). Skepticism and Content Externalism. *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/skepticism-content-externalism/>]
- Burge, Tyler. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, 4: 73-122. Reprinted in *Externalism and Self-Knowledge* (1998: 21-83) and in Burge (2007: 100-181)*. [Versión castellana: Burge, Tyler. (1996). El individualismo y lo mental (selección). Trad. de Salma Saab. En *Pensamiento y lenguaje* (1996: 311-382).]
- Burge, Tyler. (1986). Individualism and Psychology. *Philosophical Review*, 95: 3-45. Reprinted in Burge (2007: 221-253).*
- Burge, Tyler. (1988). Individualism and Self-Knowledge. *The Journal of Philosophy*, 85 (11): 649-663. Reprinted in *Externalism and Self-Knowledge* (1998: 111-127).*
- Burge, Tyler. (1996). Our Entitlement to Self-Knowledge. *Proceedings of the Aristotelian Society*, 96: 91-116.
- Burge, Tyler. (2007). *Foundations of Mind. Philosophical Essays 2*. New York: Oxford University Press.
- Carpenter, Andrew N. (2003). Davidson's Transcendental Argumentation. In *From Kant to Davidson* (2003: 219-237).
- Chalmers, David John. (1996). *The Conscious Mind. In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chalmers, David John. (2003). The Content and Epistemology of Phenomenal Belief. In *Consciousness: New Philosophical Perspectives* (2003: 220-272).
- Chisholm, Roderick. (1977). *Theory of Knowledge*. [2nd ed.]. New Jersey: Prentice-Hall.
- Churchland, Paul M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78 (2): 67-90.
- Churchland, Paul M. (1988). *Matter and Consciousness. An Contemporary Introduction to the Philosophy of Mind*. Cambridge: The MIT Press.

- Conee, Earl. (1994). Phenomenal Knowledge. *Australasian Journal of Philosophy*, 72: 136-150.
- Consciousness: New Philosophical Perspectives* (2003). Q. Smith & A. Jolic (eds.). Oxford: Oxford University Press.
- Contemporary Debates in the Philosophy of Mind*. (2007). Brian McLaughlin and John Cohen (eds.). Oxford: Blackwell.
- Crane, Tim. (2003). *The Mechanical Mind. A Philosophical Introduction to Minds, Machines and Mental Representation*. [2nd ed.] New York: Routledge. [Versión castellana: Crane, Tim. (2008).* *La mente mecánica. Introducción filosófica a mentes, máquinas y representación mental*. Trad. Juan Almela. México: Fondo de Cultura Económica.]
- Cuarto Simposio Internacional de Filosofía*. (1989). Vol. II. Enrique Villanueva (comp.). México: UNAM-IIF.
- Cutrofello, Andrew. (1999). On the Transcendental Pretensions of the Principle of Charity. In *The Philosophy of Donald Davidson* (1999: 333-341).
- Dancy, Jonathan. (1985).* *An Introduction to Contemporary Epistemology*. Malden: Blackwell Publishers. [Versión castellana: Dancy, Jonathan. (1993). *Introducción a la epistemología contemporánea*. Trad. de José Luis Prades Celma. Madrid: Tecnos.]
- Davidson, Donald. (1968) On Saying That. *Synthese*, 19: 130-146. Reprinted in Davidson (1984a: 93-108).* [Versión castellana: Davidson, Donald. (1990b). Al decir *that*. En Davidson (1990b: 108-122).]
- Davidson, Donald. (1970). Mental Events. In *Experience and Theory* (1970). Reprinted in Davidson (2001a: 207-225).* [Versión castellana: Davidson, Donald. (1995a). Sucesos mentales. En Davidson (1995a: 263-285)].
- Davidson, Donald. (1973). The Material Mind. In *Proceedings of Fourth International Congress for Logic, Methodology, and Philosophy of Science*. (1973). P. Suppes, L. Henkin, G.C. Moisil & A. Joja (eds.). North-Holland. Reprinted in Davidson (2001a: 245-259).* [Versión castellana: Davidson, Donald. (1995a). La mente material. En Davidson (1995a: 309-326).]
- Davidson, Donald. (1973a). Radical Interpretation. *Dialectica*, 27: 313-328. Reprinted in Davidson (1984a: 125-139).* [Versión castellana: Davidson, Donald. (1990b). Interpretación radical. En Davidson (1990b: 137-150).]
- Davidson, Donald. (1974). Psychology as Philosophy. In *Philosophy of Psychology*. (1974). S. C. Brown (ed.). London: The Macmillan Press and Barnes, Noble. Reprinted in Davidson (2001a: 229-244).* [Versión castellana: Davidson, Donald. (1995a). Psicología como filosofía. En Davidson (1995a: 289-307).]
- Davidson, Donald. (1974a). On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, 47: 5-20. Reprinted in Davidson (1984a: 183-198).* [Versión castellana: Davidson, Donald. (1990b). Sobre la idea misma de un esquema conceptual. En Davidson (1990b: 189-203).]

- Davidson, Donald. (1975) Thought and Talk. In *Mind and Language* (1975: 7-23). Reprinted in Davidson (1984a: 155-170).* [Versión castellana: Davidson, Donald. (1990b). Al decir *that*. En Davidson (1990b: 164-178).]
- Davidson, Donald. (1980). Towards a Unified Theory of Thought, Meaning, and Action. In *Grazer Philosophische Studien, 11*: 1-12 Reprinted as: A Unified Theory of Thought, Meaning, and Action. In Davidson (2004: 151-166).*
- Davidson, Donald. (1982). Empirical Content. *Grazer Philosophische Studien, 16-17*: 471-489. Reprinted in Davidson (2001b: 159-176).* [Versión castellana: Davidson, Donald. (2003a). El contenido empírico. En Davidson (2003a: 222-244).]
- Davidson, Donald. (1982a). Paradoxes of Irrationality. In *Philosophical Essays on Freud* (1982: 189-305). Reprinted in Davidson (2004: 169-187).*
- Davidson, Donald. (1982b). Rational Animals. *Dialectica, 36* (4): 317-27. Reprinted in Davidson (2001b: 95-105).* [Versión castellana: Davidson, Donald. (2003a). Animales racionales. En Davidson (2003a: 141-155).]
- Davidson, Donald. (1983). A Coherence Theory of Truth and Knowledge. In *Kant oder Hegel?* (1983). Henrich Klett-Cotta (ed.). Stuttgart. Reprinted in Davidson (2001b: 137-154). [Versiones castellanas: Davidson, Donald. (1992a/2003a). Verdad y conocimiento: una teoría de la coherencia. En Davidson (1992a: 73-97) y en Davidson (2003a: 193-216) como Una teoría coherentista de la verdad y el conocimiento.]
- Davidson, Donald. (1984). First Person Authority. *Dialectica, 38* (2-3): 101-112. Reprinted in Davidson (2001b: 3-14).* [Versión castellana: Davidson, Donald. (1989/2003a). La autoridad de la primera persona. En *Cuarto Simposio Internacional de Filosofía II* (1989: 9-23) y en Davidson (2003: 25-40).]
- Davidson, Donald. (1984a). *Inquiries into Truth and Interpretation*. New York: Oxford University Press.
- Davidson, Donald. (1985). Incoherence and Irrationality. *Dialectica, 39*: 345-354. Reprinted in Davidson (2004: 189-198).*
- Davidson, Donald. (1986). Deception and Division. In *The Multiple Self*. (1986): 79-92. Reprinted in Davidson (2004: 199-212).* [Versión castellana: Davidson, Donald. (1992a). Engaño y división. En Davidson (1992a: 99-117).]
- Davidson, Donald. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association, 60* (3): 441-458. Reprinted in *Externalism and Self-Knowledge* (1998: 87-110) and in Davidson (2001b: 15-38).* [Versión castellana: Davidson, Donald. (1992a/2003a). El conocimiento de la propia mente. En Davidson (1992a: 119-152) y en Davidson (2003a: 41-71) como Conocer la propia mente.]
- Davidson, Donald. (1987a). Afterthoughts. In *Reading Rorty* (1990). Reprinted in Davidson (2001b: 154-157).*
- Davidson, Donald. (1988). The Myth of the Subjective. In *Relativism: Interpretation and Confrontation* (1989). Reprinted in Davidson (2001: 39-52).* [Versión castellana: Davidson, Donald. (1992a/2003a). El mito de lo subjetivo. En Davidson (1992a: 51-71) y en Davidson (2003a: 72-90).]

- Davidson, Donald. (1989) What is Present to the Mind. *Grazer Philosophische Studien*, 36: 3-18. Reprinted in Davidson (2001b: 53-67).* [Versión castellana: Davidson, Donald (2003a). ¿Qué está presente ante la mente? En Davidson (2003a: 91-109).]
- Davidson, Donald. (1990). Epistemology Externalized. *Dialectica*, 45 (2-3): 191-202. Reprinted in Davidson (2001: 193-204).* [Versión castellana: Davidson, Donald. (2003a). La externalización de la epistemología. En Davidson (2003a: 265-279).]
- Davidson, Donald. (1990a). The Structure and Content of Truth. *The Journal of Philosophy*, 87 (6): 279-328. [Reprinted in Davidson (2005a: 7-75).*]
- Davidson, Donald. (1990b). *De la verdad y la interpretación*. Trad. De Guido Filippi. Barcelona: Gedisa.
- Davidson, Donald. (1991). Three Varieties of Knowledge. In *A. J. Ayer: Memorial Essays* (1991). Reprinted in Davidson (2001: 205-220).* [Versión castellana: Davidson, Donald. (2003). Tres variedades de conocimiento. En Davidson (2003a: 280-300).]
- Davidson, Donald. (1992). The Second Person. Published as Jusqu'où va le caractère public d'une langue? In *Wittgenstein et la philosophie d'aujourd'hui*. (1992: 241-259). Reprinted in Davidson (2001b: 107-121).* [Versión castellana: Davidson, Donald. (2003). La segunda persona. En Davidson (2003a: 156-175).]
- Davidson, Donald. (1992a). *Mente, mundo y acción. Claves para una interpretación*. Trad. de Carlos Moya Espí. Barcelona: Paidós-UAB-ICE.
- Davidson, Donald. (1993). Locating Literary Language. In *Literary Theory after Davidson* (1993). Reprinted in (2005a: 167-181).*
- Davidson, Donald. (1993a). Thinking Causes. In *Mental Causation* (1993: 3-17). Reprinted in Davidson (2005a: 185-200).*
- Davidson, Donald. (1994). The Social Aspect of Language. In *The Philosophy of Michael Dummett* (1994: 1-16). Reprinted in Davidson (2005: 109-125).*
- Davidson, Donald. (1995). The Problem of Objectivity. *Tijdschrift voor Filosofie*: 203-220. Reprinted in Davidson (2004: 3-18).*
- Davidson, Donald. (1995a). *Ensayos sobre acciones y sucesos*. Trad. de Olbeth Hansberg, José Antonio Robles y Margarita Valdés. Barcelona: UNAM-IIF-Crítica.
- Davidson, Donald. (1997). Who is fooled? In *Self-Deception and Paradoxes of Rationality* (1997: 15-27). Reprinted in Davidson (2004: 213-230).*
- Davidson, Donald. (1997a). Truth Rehabilitated. Originally Published in *Rorty and his Critics* (2000: 65-73). Reprinted in Davidson (2005: 3-17).*
- Davidson, Donald. (1998). Replies to Rorty, Stroud, McDowell, and Pereda. *Crítica*, 90 (XXX): 89-96. Reprinted in Davidson (2005a: 315-327).*
- Davidson, Donald. (1999). Reply to Thomas Nagel. In *The Philosophy of Donald Davidson* (1999: 342-344).
- Davidson, Donald. (1999a). Reply to Andrew Cutrofello. In *The Philosophy of Donald Davidson* (1999: 207-209).

- Davidson, Donald. (1999b). The Emergence of Thought. *Erkenntnis*, 51: 7-17. Reprinted in Davidson (2001: 123-134).* [Versión castellana: Davidson, Donald. (2003a). La aparición del pensamiento. En Davidson (2003a: 176-190).]
- Davidson, Donald. (2001). Externalisms. In *Interpreting Davidson* (2001: 1-16).
- Davidson, Donald. (2001a). *Essays on Actions and Events*. New York: Oxford University Press.
- Davidson, Donald. (2001b). *Subjective, Intersubjective, Objective*. New York: Oxford University Press.
- Davidson, Donald. (2003). Quine's Externalism. *Grazer Philosophische Studien. Fifty Years of Quine's Two Dogmas*. Hans-Johann Glock, Kathrin Glür, & Geert Keil (eds.), 17: 281-297.
- Davidson, Donald. (2003a). *Subjetivo, intersubjetivo, objetivo*. Trad. de Olga Fernández Prat. Madrid: Cátedra.
- Davidson, Donald. (2004). *Problems of Rationality*. New York: Oxford University Press.
- Davidson, Donald. (2005). *Truth and Predication*. Cambridge: Harvard University Press.
- Davidson, Donald. (2005a). *Truth, Language, and History*. New York: Oxford University Press.
- Davidson, Donald. (2006). The Perils and Pleasures of Interpretation. In *The Oxford Handbook of Philosophy of Language* (2006: 1056-1068).
- Davies, Martin. (1998). The Philosophy of Mind. In *Philosophy 1: A Guide through the Subject* (1998: 250-335).
- Davies, Martin. (1998a). Externalism, Architecturalism, and Epistemic Warrant. In *Knowing Our Own Minds* (1998: 321-361).
- De Caro, Mario & MacArthur, David. (2004). Introduction. In *Naturalism in Question* (2004: 1-17).
- Dennett, Daniel C. (1991). *Consciousness Explained*. New York: Back Bay Books. [Versión castellana: Dennett, Daniel. (1995). *La conciencia explicada*. Trad. de Sergio Balari Ravera. Barcelona: Paidós.]
- Dennett, Daniel C. (1996). Cow-Sharks, Magnets and Swampman. *Mind & Language*, 11 (1): 76-77.
- Dennett, Daniel C. (2005). *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*. Cambridge: The MIT Press. [Versión castellana: Dennett, Daniel. (2006). *Dulces sueños. Obstáculos filosóficos para una teoría de la conciencia*. Trad. de Julieta Barba y Silvia Jawerbaum. Buenos Aires: Katz Editores.]
- Descartes, Renati. (1636). *Discours de la méthode. Pour bien conduire sa raison, et chercher la vérité dans les sciences*. [Versión castellana: Descartes, René. (1981).* *Discurso del método, Dióptrica, Meteoros y Geometría*. Trad. de Vidal Peña. Madrid: Alfaguara.]

- Descartes, Renati. (1641). *Meditationes de prima philosophia*. [Versión castellana: Descartes, René. (1977).* *Meditaciones metafísicas con objeciones y respuestas*. Trad. de Vidal Peña. Madrid: Alfaguara.]
- Deweese-Boyd, Ian. (2012). Self-Deception. *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/self-deception/>]
- Dialogues with Davidson. Acting, Interpreting, Understanding*. (2011). Jeff Malpas (ed.). Cambridge: The MIT Press.
- Donald Davidson*. (2003). Kirk Ludwig (ed.). Cambridge: Cambridge University Press.
- Dummett, Michael. (1975). Can Analytic Philosophy be Systematic, and Ought it to Be? In *Ist systematische Philosophie möglich? Hegel-Kongress Stuttgart 1975*. (1977). Dieter Henrich (ed.). *Hegel-Studien*, 17: 305-326. Reprinted in Dummett (1978: 437-458).* [Versión castellana: Dummett, Michael. (1990). ¿Puede y debe ser sistemática la filosofía analítica? En Dummett (1990: 534- 556).]
- Dummett, Michael. (1978). *Truth and Other Enigmas*. Cambridge: Harvard University Press.
- Dummett, Michael. (1990). *La verdad y otros enigmas*. Trad. de Alfonso Herrera Patiño. México: Fondo de cultura Económica.
- Ebbs, Gary. (2001). Is Skepticism About Self-Knowledge Coherent? *Philosophical Studies*, 105: 43 58.
- Ebbs, Gary. (2005). Why Skepticism About Self-Knowledge is Self-Undermining. *Analysis*, 65 (3): 237-244.
- Ernest Sosa and His Critics*. (2004). J. Greco (ed.). Malden: Blackwell.
- Experience and Theory*. (1970). Lawrence Foster and J. W. Swanson (eds.). Cambridge: The University of Massachusetts Press-Duckworth.
- Externalism and Self-Knowledge*. (1998). Peter Ludlow & Norah Martin. (eds.). Stanford: CSLI Publications.
- Fales, Evan. (1996). *A Defense of the Given*. Lanham: Rowman & Littlefield.
- Farkas, Katlin. (2003). Does Twin Earth Arguments Rest on a Mistake? *Croatian Journal of Philosophy*, III (8): 155-69.
- Feldman, Richard. (2004). Foundational Justification. In *Ernest Sosa and His Critics*. (2004: 42-58).
- Finkelstein, David H. (2003). *Expression and the Inner*. Cambridge: Harvard University Press.
- Finkelstein, David H. (2010). Expression and Avowal. In *Wittgenstein: Key Concepts* (2010: 185-198).
- From Kant to Davidson. Philosophy and the Idea of the Transcendental*. (2003). Jeff Malpas (ed.). New York: Routledge.
- Fumerton, Richard. (1995). *Metaepistemology and Skepticism*. Lanham: Rowman & Littlefield.

- Funkhouser, Eric. (2005). Do the Self-Deceived get what they Want? *Pacific Philosophical Quarterly*, 86(3): 295-312
- Gallois, André. (1996). *The Mind Within, the World Without*. Cambridge: Cambridge University Press.
- Gertler, Brie. (2001). Introspecting Phenomenal States. *Philosophy and Phenomenological Research*, 63: 305-328.
- Gertler, Brie. (2008). Self-Knowledge. *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/selfknowledge/>]
- Gertler, Brie. (2011). *Self-Knowledge*. New York: Routledge.
- Gettier, Edmund L. (1963). Is Justified True Belief Knowledge? *Analysis*, 23: 121-123.
- Goldberg, Sanford. (2011). Self-Knowledge. In *The Routledge Companion to Epistemology* (2011: 305-315).
- Goldman, Alvin. (1976). Discrimination and Perceptual Knowledge. *The Journal of Philosophy*, 73: 771-791.
- Goldman, Alvin. (2006). *Simulating Minds*. New York: Oxford University Press.
- Guttenplan, Samuel. (1994). An Essay on Mind. In *A Companion to the Philosophy of Mind* (1994: 1-107).
- Guttenplan, Samuel. (1994a). First-Person Authority. In *A Companion to the Philosophy of Mind* (1994: 291).
- Hacker, P. M. S. (1997). Davidson on First-Person Authority. *The Philosophical Quarterly*, 47(188): 265-304.
- Harman, Gilbert. (1973). *Thought*. Princeton: Princeton University Press.
- Heil, John. (2002). Mind and Knowledge. In *The Oxford Handbook of Epistemology* (2002: 316-335).
- Hume, David. (1739). *A Treatise on Human Nature*. Reprinted in (1985)* by Penguin Books. [Versión castellana: Hume, David. (1988). *Tratado sobre la naturaleza humana*. Trad. de C. Mellizo. Madrid: Tecnos.]
- Internalism and Externalism in Semantics and Epistemology*. (2007). Sanford Golberg (ed.). New York: Oxford University Press.
- Interpreting Davidson*. P. Kotatko, P. Pagin & G. Segal (eds.). Stanford: CSLI Publications.
- Jackson, Frank. (1982)*. Epiphenomena Qualia. *Philosophical Quarterly*, 32 (127): 127-136. [Versión castellana: Jackson, Frank. (2003). Qualia epifenoménicos. Trad. de Laura E. Manríquez. En *La naturaleza de la experiencia* (2003: 95-110).]
- Jackson, Frank. (1986)*. What Mary Didn't Know. *The Journal of Philosophy*, 83 (105): 291-295. [Versión castellana: Jackson, Frank. (2003). Lo que María no sabía. Trad. de Laura E. Manríquez. En *La naturaleza de la experiencia* (2003: 111-118).]
- Jackson, Frank. (1998). *From Metaphysics to Ethics*. New York: Oxford University Press.

- Kant, Immanuel. (1787). *Kritik der Reinen Verunft*. [Versión castellana: Kant, Immanuel. (2008).* *Crítica de la razón pura*. Trad. de Pedro Ribas. México: Taurus.]
- Kim, Jaegwon. (2005). *Physicalism, or Something Near Enough*. New Jersey: Princeton University Press.
- Knowing Our Own Minds: Essays in Self-Knowledge*. (1998). Smith McDonald & Crispin Wright (eds.). New York: Oxford University Press.
- Kripke, Saul A. (1972). Naming and Necessity. In *Semantics of Natural Language* (1972). Published as Kripke, Saul A. (1980).* *Naming and Necessity*. Cambridge: Harvard University Press. [Versión castellana: Kripke, Saul. (2005). *El nombrar y la necesidad*. [2ª ed.] Trad. de Margarita M. Valdés. México: UNAM-IIF.]
- La concepción analítica de la filosofía*. (1981). Javier Muguerza (comp.). Madrid: Alianza.
- La naturaleza de la experiencia*. Vol. I *Sensaciones*. (2003). Maite Ezcurdia y Olbeth Hansberg (comps.). México: UNAM-IIF.
- Language, Mind, and Knowledge*. (1975). K. Gunderson (ed.). Minneapolis: Minnesota University Press.
- Lau, Joe & Deutsch, Max. (2010). Externalism About Mental Content. *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/content-externalism/>]
- Levine, Joseph. (2007). Phenomenal Concepts and the Materialist Constraint. In *Phenomenal Concepts and Phenomenal Knowledge* (2007: 145-166).
- Lewis, David K. (1988). What Experience Teaches. *Proceedings of the Russellian Society*, 13: 29-57.
- Literary Theory after Davidson*. (1993). R. W. Dasenbrock (ed.). Pennsylvania: Pennsylvania State University Press.
- Locke, John. (1690). *An Essay Concerning Human Understanding*. Reprinted in (1975)* by Clarendon Press. [Versión castellana: Locke, John. (2002). *Ensayo sobre el entendimiento humano*. Trad. de Antonio Ziriñ. México: Fondo de Cultura Económica.]
- Lormand, Eric. (1996). Inner Sense until Proven Guilty. [Online resource: http://www-personal.umich.edu/~lormand/phil/cons/inner_sense.htm]
- Ludlow, Peter & Martin, Norah. (1998). Introduction. In *Externalism and Self-Knowledge* (1998: 1-15).
- Lycan, William G. (1996). *Consciousness and Experience*. Cambridge: The MIT Press.
- Maker, William. (1991). Davidson's Transcendental Arguments. *Philosophy and Phenomenological Research*, 51 (2): 345-360.
- Marcus, Eric. (2004). Why Zombies are Inconceivable. *Australasian Journal of Philosophy*, 82: 477-490.
- McKay, Thomas & Nelson, Michael. (2010). The De Re/De Dicto Distinction. (Supplement to Propositional Attitude Reports). *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/prop-attitude-reports/dere.html>]

- McKinsey, Michael. (1991). Anti-Individualism and Privileged Access. *Analysis*, 51: 9-16.
Reprinted in *Externalism and Self-Knowledge*. (1998: 175-184).
- McKinsey, Michael. (2002a). The Semantic Basis of Externalism. In *Meaning and Truth. Investigations in Philosophical Semantics* (2002: 34-52).
- McKinsey, Michael. (2002b). On *Knowing Our Own Minds*. *The Philosophical Quarterly*, 52 (206): 107-116.
- McKinsey, Michael. (2007). Externalism and Privileged Access are Inconsistent. In *Contemporary Debates in the Philosophy of Mind* (2007: 53-65).
- McLaughlin, Brian & Tye, Michael. (1998). The Brown-McKinsey Charge of Inconsistency. In *Knowing Our Own Minds: Essays in Self-Knowledge* (1998). Reprinted in *Externalism and Self-Knowledge* (1998: 207-214).*
- Meaning and Truth. Investigations in Philosophical Semantics*. (2002). J. K. Campbell, M. O'Rourke & D. Shier (eds.). New York: Seven Bridges Press, LLC.
- Mele, Alfred. (1997). Real Self-Deception. *Behavioral and Brain Sciences*, 20(1): 91-102.
- Mental Causation*. (1993). J. Heil & A. Mele (eds.). New York: Oxford University Press.
- Mind and Language. Wolfson College Lectures*. (1975). Samuel Guttenplan (ed.). Oxford: Clarendon Press.
- Minnesota Studies in the Philosophy of Science. Vol. I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. (1956). H. Feigl & M. Scriven (eds.). Minneapolis: University of Minnesota Press.
- Moran, Richard. (2001). *Authority and Strangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Nagel, Thomas. (1974).* What is it Like to be a Bat? *Philosophical Review*, 83: 435-450.
[Versión castellana: Nagel, Thomas. (2003). ¿Cómo es ser un murciélago? Traducción de Héctor Islas. En *La naturaleza de la experiencia* (2003: 45-61).]
- Nagel, Thomas. (1987). *What Does It All Mean? A Very Short Introduction to Philosophy*. New York: Oxford University Press. [Versión castellana: Nagel, Thomas. (2003). ¿Qué significa todo esto? Una brevísima introducción a la filosofía. Trad. de Alfonso Montelongo. México: Fondo de Cultura Económica.]
- Nagel, Thomas. (1999). Davidson's New *Cogito*. In *The Philosophy of Donald Davidson* (1999: 195-206).
- Naturalism in Question*. (2004). Mario de Caro & David MacArthur. Cambridge: Harvard University Press.
- New Essays on Externalism and Self-Knowledge*. (2003). Susana Nuccetelli (ed.). Cambridge: The MIT Press.
- Nichols, Shaun & Stich, Stephen. (2003). *Mindreading*. Oxford: Oxford University Press.
- Normas, virtudes y valores epistémicos. Ensayos de epistemología contemporánea*. (2011). Margarita M. Valdés y Miguel Ángel Fernández (comps.). México: UNAM-IIF.
- Nozick, Robert. (1981). *Philosophical Explanations*. Cambridge: Harvard University Press.

- Papineau, David. (1993). *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau, David. (2007). Naturalism. In *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/naturalism/>]
- Parent, Ted. (2007). Infallibilism about Self-Knowledge. *Philosophical Studies*, 133: 411-424.
- Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. (2007). T. Alter & S. Walter (eds.). New York: Oxford University Press.
- Philosophical Essays on Freud*. (1982). R. Wollheim & J. Hopkins (eds.). Cambridge: Cambridge University Press.
- Philosophy 1: A Guide through the Subject*. (1998). A. C. Grayling (ed.). Oxford: Oxford University Press.
- Pitt, David. (2004). The Phenomenology of Cognition, or, What is it Like to Think that *P*? *Philosophy and Phenomenological Research*, 69: 1-36.
- Plantinga, Alvin. (1993). *Warrant: The Current Debate*. Oxford: Oxford University Press.
- Platón. (1986). *Diálogos*. 9 vols. Trad., intro y notas de C. García Gual, M. Martínez Hernández y E. Lledo Íñigo. Madrid: Gredos.
- Putnam, Hilary. (1981).* *Reason, Truth & History*. New York: Cambridge University Press. [Versión castellana: Putnam, Hilary. (1988). *Razón, verdad e historia*. Madrid: Tecnos.]
- Putnam, Hilary. (1975). The Meaning of 'Meaning'. In *Language, Mind and Knowledge* (1975). Reprinted in Putnam (1975a).* [Versión castellana: Putnam, Hilary. (1984). *El significado de "significado"*. Traducción de Jorge G. Flematti. México: UNAM.]
- Putnam, Hilary. (1975a). *Mind, Language and Reality. Philosophical Papers*. Vol. 2. Cambridge: Cambridge University Press.
- Quine, Willard Van Orman. (1951). Two Dogmas of Empiricism. *Philosophical Review*, 60: 20-43. Reprinted in Quine (1980: 20-46).* [Quine, Willard Van Orman. (2002). Dos dogmas del empirismo. En Quine (2002: 61-91).]
- Quine, Willard Van Orman. (1956). Quantifiers and Propositional Attitudes. *The Journal of Philosophy*, 53: 177-187.
- Quine, Willard Van Orman. (1957). The Scope and Language of Science. *British Journal for the Philosophy of Science*, 8(29): 1-17. Reprinted in Quine (1966: 215-232).*
- Quine, Willard Van Orman. (1960).* *Word and Object*. Cambridge: The MIT Press. [Versión castellana: Quine, Willard Van Orman. (2001). *Palabra y objeto*. Trad. de Manuel Sacristán. Barcelona: Herder.]
- Quine, Willard Van Orman. (1966). *The Ways of Paradox and Other Essays*. Cambridge: Harvard University Press.
- Quine, Willard Van Orman. (1968). Epistemology Naturalized. In Quine (1969: 69-90).* [Versión castellana: Quine, Willard Van Orman. (1974). Naturalización de la epistemología. En Quine (1974: 93-119).]
- Quine, Willard Van Orman. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

- Quine, Willard Van Orman. (1974). *La relatividad ontológica y otros ensayos*. Trad. de Manuel Garrido y Josep Ll. Blasco. Madrid: Tecnos.
- Quine, Willard Van Orman. (1975). The Nature of Natural Knowledge. In *Mind and Language* (1975: 67-81). Reprinted in Quine (2008: 257-270).*
- Quine, Willard Van Orman. (1975a). Five Milestones of Empiricism. Reprinted in Quine (1981: 67-72).* [Versión castellana: Quine, Willard Van Orman. (1986). Cinco hitos del empirismo. En Quine (1986: 87-93).]
- Quine, Willard Van Orman. (1980). *From a Logical Point of View*. [2nd ed.]. Cambridge: Harvard University Press. [Versión castellana: Quine, Willard Van Orman. (2002) *Desde un punto de vista lógico*. Trad. de Manuel Sacristán. Barcelona: Paidós.]
- Quine, Willard Van Orman. (1981). *Theories and Things*. Cambridge: Harvard University Press.
- Quine, Willard Van Orman. (1986). *Teorías y cosas*. Trad. de Antonio Ziri6n. Mexico: UNAM-IIF.
- Quine, Willard Van Orman. (1990).* *The Pursuit of Truth*. Cambridge: Harvard University Press. [Versión castellana: *La búsqueda de la verdad*. Trad. de Javier Rodríguez Alcázar. Barcelona: Crítica.]
- Quine, Willard Van Orman. (2008). *Confessions of a Confirmed Externsionalist and Other Essays*. Dagfin Føllesdal & Douglas B. Quine (eds.). Cambridge: Harvard University Press.
- Ramachandran, V. S. (1996). The Evolutionary Biology of Self-Deception, Laughter, Dreaming and Depression: Some Clues from Anosognosia. *Medical Hypotheses*, 47: 347-362.
- Reading Rorty*. (1990). Alan Malachowski (ed.). Cambridge: Blackwell.
- Relativism: Interpretation and Confrontation*. (1989). M. Krauz (ed.). Indiana: Notre Dame University Press.
- Rorty and his Critics* (2000). Robert B. Brandom (ed.). Oxford: Blackwell.
- Rorty, Richard. (1970). Incommensurability as the Mark of the Mental. *Journal of Philosophy*, 67: 339-424.
- Rorty, Richard. (1998). Davidson between Wittgenstein and Tarski. *Crítica*, 80(XXX): 49-71.
- Röska-Hardy, Louise. (2011). Davidson and the Source of Self-Knowledge. In *Dialogues with Davidson* (2011: 371-403).
- Russell, Bertrand. (1918-1919). The Philosophy of Logical Atomism. *The Monist*, 28: 495-527; & 29: 31-63, 190-222, 345-380. [Versión castellana: Russell, Bertrand. (1981). La filosofía del atomismo lógico. En *La concepción analítica de la filosofía*. (1981: 139-251).]
- Ryle, Gilbert. (1949).* *The Concept of Mind*. United States of America: Barnes & Noble [Versión castellana: Ryle, Gilbert. (2005). *El concepto de lo mental*. Trad. de Eduardo Rabossi. Barcelona: Paidós.]

- Sawyer, Sarah. (2007). There Is No Viable Notion of Narrow Content. In *Contemporary debates in Philosophy of Mind* (2007: 20-34).
- Schiffer, Stephen. (1996). Contextualist Solutions to Skepticism. *Proceedings of the Aristotelian Society*, 96: 317-333.
- Schiffer, Stephen. (2004). Skepticism and the Vagaries of Justified Belief. *Philosophical Studies*, 103: 161-84.
- Scruton, Roger. (1998). Modern Philosophy I: The Rationalists and Kant. In *Philosophy 1: A Guide through the Subject* (1998: 440-483).
- Searle, John R. (1987). Indeterminacy, Empiricism, and the First-Person. *The Journal of Philosophy*, 84 (3): 123-146.
- Segal, Gabriel. (2007). Cognitive Content and Propositional Attitude Attributions. In *Contemporary debates in Philosophy of Mind* (2007: 5-19)
- Segundo Simposio Internacional de Filosofía*. Vol. I. (1988). Enrique Villanueva (comp.). México: UNAM.
- Self-Deception and Paradoxes of Rationality*. (1997). J.-P. Dupuy (ed.). Stanford: CSLI-Stanford University Press.
- Sellars, Wilfrid. (1956). Empiricism and the Philosophy of Mind. In *Minnesota Studies in the Philosophy of Science* (1956: 253-329). Reprinted in Sellars (1963: 127-196).*
- Sellars, Wilfrid. (1963). *Science, Perception and Reality*. London: Routledge & Kegan Paul.
- Semantics of Natural Language*. (1972). D. Davidson & G. Harman (eds.). Cambridge: Reidel.
- Shoemaker, Sydney. (1994). Self-Knowledge and 'Inner Sense'. *Philosophy and Phenomenological Research*, 54: 249-314.
- Sosa, Ernest. (1980). The Raft and the Pyramid. In *Midwest Studies in Philosophy*, vol. 5 *Studies in Epistemology*: 3-25.
- Sosa, Ernest. (2003). Knowledge of Self, Others, and World. In *Donald Davidson* (2003: 163-182).
- Sosa, Ernest. (2003). Privileged Access. In *Consciousness* (2003: 273-292).
- Sosa, Ernesto. (1982). Rostros y destinos de la epistemología. En *Segundo Simposio Internacional de Filosofía* (1988: 57-87).
- Stanford Encyclopedia of Philosophy*. (2012). Edward N. Zalta (ed.). Stanford: The Metaphysics Research Lab/CSLI/Stanford University. [Online in: <http://plato.stanford.edu/>]
- Steup, Matthias. (2006). The Analysis of Knowledge. *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/knowledge-analysis/>]
- Stroud, Barry. (1977).* *Hume*. Routledge & Kegan Paul. [Versión castellana: Stroud, Barry. (1995). *Hume*. Trad. de Antonio Ziri6n. México: UNAM-IIF.]

- Stroud, Barry. (1989). Understanding Human Knowledge in General. In *Knowledge and Skepticism*. M. Clay & K. Lehrer (eds.). Boulder: Westview. Reprinted in Stroud (2000: 99-120).*
- Stroud, Barry. (2000). *Understanding Human Knowledge*. New York: Oxford University Press.
- Stroud, Barry. (2008). Our Debt to Descartes. In *A Companion to Descartes* (2008: 513-525).
- Sturgeon, Scott. (1998). Knowledge. In *Philosophy 1: A Guide through the Subject* (1998: 10-26).
- Tagle Marroquín, Jorge R. (2007). La teoría unificada de Davidson, el problema residual y la elucidación reflexiva del concepto de verdad. *Sapere aude!, II*: 7-22.
- Tagle Marroquín, Jorge R. (2009). Naturalismo metodológico, escepticismo filosófico e intuición. Conferencia inédita presentada en el coloquio *Escepticismo filosófico*.
- The Multiple Self*. (1986). J. Elster (ed.). Cambridge: Cambridge University Press
- The Oxford Handbook of Epistemology*. (2002). Paul K. Moser (ed.). New York: Oxford University Press.
- The Oxford Handbook of Philosophy of Language*. (2006). Ernest Lepore & Barry C. Smith (eds.). New York: Oxford University Press.
- The Oxford Handbook of Philosophy of Mind*. (2009). B. P. McLaughlin, A. Beckermann & S. Walter (eds.). New York: Oxford University Press.
- The Philosophy of Donald Davidson*. (1999). Lewis Edwin Hahn (ed.). Illinois: Open Court Publishing Company.
- The Philosophy of Michael Dummett*. (1994). B. McGuinness (ed.). Dordrecht: Kluwer.
- The Routledge Companion to Epistemology*. (2011). Sven Bercker & Duncan Pritchard. New York: Routledge.
- Treanor, Nick. (2006). The *Cogito* and the Metaphysics of the Mind. *Philosophical Studies*, 130 (2): 247-271.
- Wheatherson, Brian. (2006). Intrinsic vs. extrinsic properties. *Stanford Encyclopedia of Philosophy*. [Online resource: <http://plato.stanford.edu/entries/intrinsic-extrinsic/>]
- Williams, Bernard. (1977). *Descartes: The Problem of Pure Enquiry*. Sussex: The Harvester Press. Also published in (2005)* by Taylor & Francis. [Versión castellana: Williams, Bernard. (1995). *Descartes: el proyecto de la investigación pura*. Trad. de Laura Benítez. México: UNAM-IIF].
- Williams, Michael. (2001). *Problems of Knowledge. A Critical Introduction to Epistemology*. New York: Oxford University Press.
- Williamson, Timothy. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy. (2011). First Knowledge Epistemology. In *The Routledge Companion to Epistemology* (2011: 208-218).

- Wittgenstein et la philosophie d'aujourd'hui.* (1992). J. Sebestik et A. Soulez (dir.). Paris: Méridiens Klincksieck.
- Wittgenstein, Ludwig. (1953).* *Philosophische Untersuchungen/Philosophical Investigations*. G. E. M. Anscombe & R. Rhees (eds.). Oxford: Basil Blackwell. [Versión castellana: Wittgenstein, Ludwig. (2003). *Investigaciones filosóficas*. Trad. de Alfonso García Suárez y Ulises Moulines. México: UNAM-IIF.]
- Wittgenstein, Ludwig. (1967).* *Zettel*. G. E. M. Anscombe & G. H. von Wright (eds.). Oxford: Basil Blackwell. [Versión castellana: Wittgenstein, Ludwig. (1979). *Zettel*. Trad. de Octavio Castro y Carlos Ulises Moulines. México: UNAM-IIF.]
- Wittgenstein, Ludwig. (1969).* *Über Gewissheit/On Certainty*. G. E. M. Anscombe & G. H. von Wright (eds.). Oxford: Basil Blackwell. [Versión castellana: Wittgenstein, Ludwig. (1988). *Sobre la Certeza*. Trad. de Josep Lluís Prades y Vicent Raga. Barcelona: Gedisa.]
- Wittgenstein: Key Concepts*. Kelly Jolley (ed.). Durham: Acumen Press.
- Wright, Crispin. (1998). Self-Knowledge: The Wittgensteinian Legacy. In *Knowing Our Own Minds* (1998).
- Yablo, Stephen. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, 53:1-42. Reprinted in Yablo (2008: 39-78).
- Yablo, Stephen. (2008). *Thoughts. Papers on Mind, Meaning, and Modality*. Vol. I. New York: Oxford University Press.
- Zagzebski, Linda. (2009). *On Epistemology*. California: Wadsworth.

