

El siguiente material se reproduce con fines estrictamente académicos y es para uso exclusivo de los estudiantes de la materia **Juegos de Identidad en los Mundos Virtuales** de la facultad de **Derecho y Ciencias Sociales** de la Universidad ICESI, de acuerdo con el Artículo 32 de la Ley 23 de 1982. Y con el Artículo 22 de la Decisión 351 de la Comisión del Acuerdo de Cartagena.

ARTÍCULO 32:

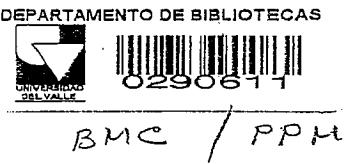
“Es permitido utilizar obras literarias o artísticas o parte de ellas, a título de ilustración en obras destinadas a la enseñanza, por medio de publicaciones, emisiones o radiodifusiones o grabaciones sonoras o visuales, dentro de los límites justificados por el fin propuesto o comunicar con propósito de enseñanza la obra radiodifundida para fines escolares educativos, universitarios y de formación personal sin fines de lucro, con la obligación de mencionar el nombre del autor y el título de las así utilizadas”.

Artículo 22 de la Decisión 351 de la Comisión del Acuerdo Cartagena.

ARTÍCULO 22:

Sin prejuicio de lo dispuesto en el Capítulo V y en el Artículo anterior, será lícito realizar, sin la autorización del autor y sin el pago de remuneración alguna, los siguientes actos:

- b) Reproducir por medio reprográficos para la enseñanza o para la realización de exámenes en instituciones educativas, en la medida justificada por el fin que se persiga, artículos lícitamente publicados en periódicos o colecciones periódicas, o breves extractos de obras lícitamente publicadas, a condición que tal utilización se haga conforme a los usos honrados y que la misma no sea objeto de venta o transacción a título oneroso, ni tenga directa o indirectamente fines de lucro;...”.



Filosofía de la mente

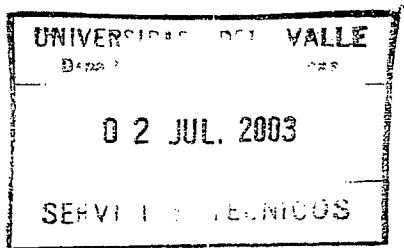
E. J. Lowe



UNIVERSIDAD DEL VALLE

CENTRO DE DOCUMENTACION
DEPARTAMENTO DE FILOSOFIA
Facultad de Humanidades


IDEA BOOKS



*Identidad personal
y conocimiento de uno mismo*

En el capítulo 1 describí la filosofía de la mente como el estudio filosófico de los seres dotados de mente en la medida en que están dotados de mente. El término general que introduce para referir a un ser que tiene mente era el de *sujeto de experiencias* —interpretando aquí 'experiencias' en un sentido amplio que incluyera cualquier clase de sensación, percepción o pensamiento. Supongo que el término 'sujeto de experiencias' es más amplio que el término 'persona', es decir que, si bien todas las personas son (al menos potencialmente) sujetos de experiencias, no todos los sujetos de experiencias son personas. Ello se debe a que pienso que al menos algunas criaturas no humanas, como los chimpancés, son realmente sujetos de experiencias aunque puedan no ser personas. Es perfectamente concebible que haya personas no humanas, pero que existan en realidad es algo que puede cuestionarse. ¿Qué es entonces lo que es característico de las personas por contraposición a otros sujetos de experiencias? Sugiero que lo siguiente: una persona es *un yo*, es decir, es un sujeto de experiencias que tiene la capacidad de reconocerse a sí mismo como sujeto individual de experiencias. Un yo posee conocimiento reflexivo. Por 'conocimiento reflexivo' entiendo, hablando laxamente, conocimiento de la propia identidad y estados mentales conscientes, es decir, conocimiento de quién es uno y de lo que está pensando y sintiendo. Como veremos, existen algunas complicaciones a la hora de especificar esta noción de un modo completamente satisfactorio, si es que al final esto puede lograrse. Pero —hablando laxamente de nuevo— la posesión del tipo de conocimiento reflexivo de uno mismo que le hace a uno una persona se presenta junto con la posesión de un concepto de 'primera persona' de *uno mismo*, que se refleja lingüísticamente en la capacidad de usar la palabra 'yo' —entendiéndola— para referirse a uno mismo.

Pero ¿a qué clase de cosa se refiere la palabra 'yo', suponiendo que realmente se refiera a algo? ¿Qué clase de cosa soy yo? Al formular esta pregunta volvemos, claro está, a algunas de las cuestiones que tratamos en el capítulo 2. Vimos allí que distintos filósofos han ofrecido respuestas muy diferentes a esta cuestión, sosteniendo algunos que 'yo' se refiere a un determinado cuerpo —'mi' cuerpo—, otros que se refiere

a algo completamente no físico, como un alma o espíritu inmaterial, y otros que se refiere a algo que es una combinación de cuerpo y alma. Otra respuesta que también es posible es que 'yo' se refiere a una colección de pensamientos, sensaciones y sentimientos —'mis' pensamientos, sensaciones y sentimientos. Pero podría sorprendernos que los filósofos estén tan radicalmente en desacuerdo acerca de a qué clase de cosa se refiere 'yo' y estén, sin embargo, tan seguros de que se refiere a *algo*. Tal vez debiéramos cuestionar ese supuesto, o, aunque lo aceptemos, tal vez debiéramos poner en cuestión el supuesto de que 'yo' se refiere a una cosa del *mismo* tipo siempre que se la usa para referirse a algo. Quizá una persona o un 'yo' no constituya un *tipo* de cosas cuyos ejemplos estén todos sujetos a las mismas condiciones de identidad. Tal vez ser una persona o un yo sea desempeñar un cierto papel o realizar una función —un papel o una función que podría ser desempeñado o realizado por cosas de muchos tipos distintos. Por ejemplo, podría sostenerse que ser una persona es un papel que ahora desempeña mi cuerpo y que ha desempeñado mi cuerpo durante la mayor parte de mi existencia, pero que durante las primeras semanas o meses de mi existencia no desempeñaba ese papel. Ello implicaría que, al menos en mi caso, 'yo' se refiere a mi cuerpo, pero que hubo momentos de tiempo en los que yo (es decir, mi cuerpo) existía pero no era (todavía) una persona. En el curso del presente capítulo examinaremos más detalladamente estas y otras posibilidades.

Otro conjunto de temas que tendremos que explorar concierne al conocimiento de nuestros propios estados mentales y de su contenido. ¿Cómo es que parece que tenemos un conocimiento incontestable de lo que estamos pensando y sintiendo? ¿O es que nuestra creencia de que tenemos tal conocimiento carece de hecho de fundamento? Si ello fuera así, aún habríamos de explicar la frecuencia y tenacidad de esa creencia. Sin embargo, comencemos por el fenómeno de la autoreferencia, es decir, con el uso comprensivo de la palabra 'yo'.

LA PRIMERA PERSONA

Aunque los niños adquieren el uso del pronombre de primera persona 'yo' en un estadio muy temprano de su desarrollo lingüístico, entender la semántica del discurso de primera persona es algo sorprendentemente difícil. Resulta fácil decir cuál es la función lingüística de la palabra 'yo', de un modo que satisfará a la mayoría de las personas. 'Yo' es la palabra (del español) que todo el mundo utiliza para referirse a sí mismo. Parece que todas las lenguas humanas tienen una palabra o una expresión que equivale a esto. Pero la dificultad que los filósofos tienen en la comprensión del significado de la palabra 'yo' surge de la dificultad de especificar, de una manera que no sea circular y que resulte iluminadora, lo que es 'referirse a sí mismo', en el modo especial que

acompaña al uso comprensivo de la palabra 'yo'. Hablo aquí deliberadamente del uso *comprehensivo* de la palabra 'yo' (del uso de la palabra 'yo' entendiéndola), porque, por ejemplo, se podría decir que incluso un ordenador que carece completamente de mente se 'refiere a sí mismo' cuando muestra en pantalla un mensaje como 'Estoy preparado'.¹ De igual modo, es concebible que se pudiera enseñar a un loro a proferir las palabras 'Yo tengo hambre' siempre que tenga hambre, en cuyo caso podría decirse que utiliza el pronombre de primera persona para referirse a sí mismo, pero ello no implicaría que tuviera un concepto de primera persona de sí mismo, como parece que se requiere para usar el pronombre 'yo' con comprensión.

La dificultad a la que estoy aludiendo puede tal vez ponerse de manifiesto al comparar el uso comprensivo de la palabra 'yo' para referirse a uno mismo con el uso comprensivo de otros términos singulares para referirse a uno mismo. Por ejemplo, yo puedo utilizar mi nombre de persona, 'Jonathan Lowe', para referirme a mí mismo. Igualmente, puedo utilizar determinadas descripciones para referirme a mí mismo, como 'el autor de este libro' o 'la persona que está sentada en esta silla'. Sin embargo, un rasgo curioso de la palabra 'yo' es que parece que está *garantizado* que se refiera a una persona totalmente específica en cualquier ocasión de uso y de tal modo que la persona que la usa no puede equivocarse sobre a qué persona se refiere —a saber, a ella misma—. Ese rasgo no acompaña a los otros modos de referirnos a nosotros mismos. Yo puedo olvidarme de que soy Jonathan Lowe o llegar a dudar de que yo sea el autor de este libro, pero no puedo dudar de que yo soy yo. Se podría desechar esto calificándolo de trivial, como al hecho de que no puedo dudar de que Jonathan Lowe es Jonathan Lowe. Pero esta sería una respuesta demasiado precipitada y superficial. Puede ponerse de manifiesto que no es un asunto trivial al considerar el siguiente tipo de ejemplo. Supongamos que, tras entrar en una casa ajena y al caminar por un pasillo, veo una figura humana que se me está acercando y me formo el siguiente juicio: 'Esa persona tiene una apariencia sospechosa'. Me doy entonces cuenta, de repente, que la persona en cuestión soy *yo mismo*, reflejado en un espejo que hay al final del pasillo. Aunque, en un cierto sentido, yo sabía a quién me estaba refiriendo al utilizar la expresión demonstrativa 'esa persona' —a saber, a la persona que parecía estar acercándose—, está claro que existe otro sentido en el cual yo *no* sabía a quién me estaba refiriendo, puesto que me estaba refiriendo a mí mismo sin saberlo. Por tanto, es claro que, cuando me 'refiero a mí mismo' al modo que comporta este ejemplo, no me refiero a mí mismo al modo en que lo hago cuando utilizo la palabra 'yo', ya que en este último caso no existe una posibilidad similar de no saber a quién me estoy refiriendo.

¹ En este punto y a lo largo de todo el capítulo el lector habrá de tener en cuenta, claro está, que en español —al contrario que en inglés— el pronombre personal resulta muy frecuentemente elidido (N. de la T.).

Este rasgo del uso de la palabra 'yo' se encuentra estrechamente asociado al hecho de que determinados juicios de primera persona muestran lo que se denomina a veces 'inmunidad al error debido a identificación equivocada'.² Es perfectamente posible hacer un juicio de primera persona erróneo, de la forma 'Yo soy *F*', de la misma manera que es posible hacer un juicio erróneo de tercera persona, de la forma '*S* es *F*'. Sin embargo, mientras que '*S* es *F*' puede ser erróneo de dos modos diferentes, parece que determinados juicios de la forma 'Yo soy *F*' sólo pueden ser erróneos de una sola manera. Comparemos, por ejemplo, los dos juicios 'Juan está irritado' y 'Yo estoy irritado'. Al hacer el primero de estos juicios puedo estar equivocado acerca de *qué* es lo que Juan siente o acerca de *quién* está irritado. Por un lado, tal vez Juan esté celoso en vez de irritado. Por el otro, quizás no sea Juan quien está irritado, sino alguna otra persona a la que he confundido con Juan. Pero consideremos ahora el segundo juicio, 'Yo estoy irritado'. Aquí también es posible que esté equivocado sobre *qué* es lo que yo siento; puede ser que crea que estoy irritado cuando en realidad lo que yo siento son celos. Pero lo que no parece tener ningún sentido es que pueda confundir la irritación de ninguna otra persona con la mía. No obstante, no todos los juicios de primera persona tienen esta propiedad de inmunidad al error debido a una identificación equivocada; según parece, los únicos que la tienen son los que comportan la atribución a uno mismo de un estudio psicológico consciente. Consideremos, por ejemplo, el juicio 'Estoy tocando la mesa con la mano', realizado sobre la base de lo que puedo ver y sentir. Aunque sea improbable, es concebible que de hecho no sea mi mano la que está tocando la mesa, sino la de la persona que se sienta a mi lado, que es muy parecida a la mía, y que lo que realmente siento es que mi mano toca otro objeto cercano. En ese caso, estoy equivocado acerca de *quién* está tocando la mesa, de un modo en que no podría estar equivocado acerca de *quién* está irritado cuando hago el juicio '(Yo) estoy irritado'.

Algunos filósofos parecen pensar que, cuando la palabra 'yo' se utiliza como al hacer el juicio '(Yo) estoy irritado', no puede realmente estar funcionando como una expresión referencial, precisamente debido a que en tales casos la referencia errónea está, según parece, excluida.³ Esos fi-

² La expresión 'inmunidad al error debido a identificación equivocada' se debe a Sydney Shoemaker; véase su 'Self-Reference and Self-Awareness', *Journal of Philosophy* 65 (1968), pp. 555-67, reimpresso en su *Identity, Cause, and Mind: Philosophical Essays* (Cambridge: Cambridge University Press, 1984) y en Quassim Cassam (ed.) *Self-Knowledge* (Oxford: Oxford University Press, 1994). Esta noción se inspira en Wittgenstein; véase *The Blue and Brown Books*, 2^a edición (Oxford: Blackwell, 1969), pp. 66 y siguientes. Véase también Gareth Evans, *The Varieties of Reference* (Oxford: Clarendon Press, 1982), pp. 179-91.

³ Para dudas acerca de que 'yo' sea una expresión referencial, véase G. E. M. Anscombe, 'The First Person' en Samuel Guttenplan (ed.), *Mind and Language* (Oxford: Clarendon Press, 1975), reimpresso en G. E. M. Anscombe, *Metaphysics and the Philosophy of Mind* (Oxford: Blackwell, 1981). Para una discusión crítica de esta cuestión, véase Andy Hamilton, 'Anscombian and Cartesian Scepticism', *Philosophical Quarterly* 41 (1991), pp. 39-54.

lósofos abrazan la idea de que un genuino acto de referir a algo sólo puede darse si existe la posibilidad de error o de fallo; al modo como podría expresarse, donde no hay posibilidad de fallo, tampoco hay posibilidad de éxito. Consideran por ello que la palabra 'yo' en tales contextos no tiene más función referencial de la que tendría el presunto sujeto elidido de un enunciado impersonal como 'Llueve'. ¿Cuál sería, pues, la función de la palabra 'yo' en tales contextos, según esos filósofos? No está muy claro cómo contestarían a esta pregunta, pero la tendencia sería afirmar que oraciones como '(Yo) estoy irritado' se usan, en realidad, no para expresar *juicios* que pudieran ser verdaderos o falsos, sino para *expresar sensaciones internas*. De acuerdo con esta concepción, decir '(Yo) estoy irritado' es expresar *las propias sensaciones* y no expresar un juicio acerca de lo que se siente. Tal 'expresión' se considera como una expresión verbal de la emoción, comparable con 'expresiones' de emoción no verbales como puedan serlo las miradas y los gestos de irritación.⁴

No obstante, según creo, la mayoría de los filósofos no quedarían persuadidos por esta doctrina, y no sería la menor de las razones el que no se subscribirían al principio de que todo acto genuino de referencia debe dejar un espacio para la posibilidad de error. Además, considerarían que es intrínsecamente implausible suponer que la misma palabra 'yo' pueda tener dos funciones radicalmente distintas, toda vez que parece difícilmente discutible que la palabra 'yo' tenga una función referencial en muchos contextos de su uso. ¿Cómo explicarían entonces estos filósofos que la palabra 'yo' no pueda, según parece, ser capaz de referencia? Algunos de ellos lo atribuirían simplemente a que se trata de una expresión por así decir *reflexiva del ejemplar*, semejante a expresiones como 'aquí' y 'ahora'. De igual manera que toda manifestación de la palabra 'aquí' se refiere normalmente al lugar en el que se realiza la manifestación y que toda manifestación de la palabra 'ahora' se refiere al momento de tiempo en que se profiere esa palabra, según esa explicación una manifestación de la palabra 'yo' se refiere a la persona que la está profiriendo. En consecuencia, la regla semántica que gobierna el uso de la palabra 'yo' impide la posibilidad de que alguien utilice erróneamente la palabra 'yo' para referirse a ninguna otra cosa que no sea él mismo.⁵ Sin embargo, aunque sea importante reconocer el carácter refle-

⁴ En las *Investigaciones filosóficas* dice Wittgenstein en un momento dado: 'Cuando digo "Siento dolor", no señalo a una persona que tenga dolor ... no nombro una persona. Del mismo modo en que no nombro a nadie al gemir de dolor. Aunque alguna otra persona pueda ver por los gemidos quién siente dolor.' Véase Ludwig Wittgenstein, *Philosophical Investigations*, traducción de G. E. M. Anscombe, 2^a edición (Oxford: Blackwell, 1958), p. 404.

⁵ Para más información sobre la reflexividad testimonial y sobre la comparación entre 'yo' y 'ahora', véase D. H. Mellor, 'I and Now', *Proceedings of the Aristotelian Society* 89 (1989), pp. 79-94, reimpresso en su *Matters of Metaphysics* (Cambridge: Cambridge University Press, 1991). Véanse también los artículos en la parte I de Palle Yourgrau (ed.), *Demonstratives* (Oxford: Oxford University Press, 1990) y John Campbell, *Past, Space, and Self* (Cambridge, Mass.: MIT Press, 1994), capítulos 3 y 4.

xivo del ejemplar de la palabra 'yo', no parece que aduciendo sólo esto puedan explicarse todos los rasgos distintivos de la referencia de primera persona que hemos venido examinando. Ello se debe a que no se puede captar lo que de especial tiene el uso comprensivo de la palabra 'yo' para referirse a sí mismo citando meramente la regla semántica de que cualquier manifestación de la palabra 'yo' se refiere de manera estándar a la persona que la profiere; puesto que esta regla no impone condición alguna acerca de que la persona —o la cosa— que profiera la palabra 'yo' debería tener una concepción de primera persona de sí mismo como un sujeto de experiencias autoconsciente, mientras que el uso comprensivo de la palabra 'yo' para referirse a uno mismo sí que requiere esto.

Tengo que confesar que no conozco ningún modo satisfactorio de proporcionar un análisis que no sea circular del concepto de auto-referencia propio del uso comprensivo de la palabra 'yo'. La capacidad para pensar *en uno mismo como uno mismo* y de atribuirse a uno mismo, concibiéndose uno de este modo, diversos pensamientos y sensaciones conscientes, parece ser una capacidad primitiva e irreducible, en el sentido de que no se puede tomar como modelo de esta capacidad ninguna otra capacidad más general de pensar en objetos particulares y atribuirles propiedades. En alguna fase del desarrollo intelectual del niño, éste adquiere esta capacidad, pero el modo en que la adquiere seguirá siendo un tanto misterioso en la medida en que no tengamos ningún modo de caracterizar la capacidad en cuestión excepto en sus propios términos.⁶

PERSONAS Y CRITERIOS DE IDENTIDAD

Dado que la palabra 'yo' tiene garantizada su referencia en toda ocasión en que se la use con comprensión, ¿a qué clase de cosa se refiere cuando se la usa? La respuesta rápida sería que se refiere a una persona o yo particular. Pero, en el modo en que he venido usando los términos 'persona' y 'yo', esto es, en efecto, verdad simplemente por definición y por tanto no resulta muy revelador, dado que he caracterizado una persona como un sujeto de experiencias que tiene un concepto de primera persona de sí mismo. No hemos establecido aún que las personas, o más generalmente los sujetos de experiencias, constituyan un tipo o clase diferenciado de cosas al modo en que, por ejemplo, las estrellas, los robles y las sillas constituyen tipos diferenciados de cosas. Los especímenes de una clase o tipo genuino deben compartir como mínimo

⁶ Para una discusión del concepto y del desarrollo de la conciencia de uno mismo desde múltiples perspectivas, tanto filosóficas como psicológicas, véase José Luis Bermúdez, Anthony Marcel y Naomi Eilan (eds.), *The Body and the Self* (Cambridge, Mass.: MIT Press, 1995). Véase también José Luis Bermúdez, *The Paradox of Self-Consciousness* (Cambridge, Mass.: MIT Press, 1998).

las mismas condiciones de identidad, lo cual nos posibilita contarlos y seguirles la pista a través del tiempo. Si se me pregunta cuántos robles hay en un determinado bosque en un cierto momento de tiempo, sé, al menos en principio, cómo descubrir la respuesta a esta cuestión, porque sé qué es lo que hace que un roble sea numéricamente distinto de otro en un cierto momento de tiempo, a saber, el hecho de que ocupa una región propia y separada en el espacio, adecuado para acomodar exactamente a un roble. De igual modo, si se me pregunta si el roble que se encuentra ahora en una determinada parte del bosque es el *mismo* roble que el que se encontraba allí hace cuarenta años, conozco, cuando menos en principio, el modo de descubrir la respuesta, porque entiendo las condiciones de persistencia de los robles, es decir, entiendo cuáles son los tipos de cambios que puede experimentar un roble y cuáles no para sobrevivir a través del tiempo. Por ejemplo, sé que los robles pueden sobrevivir un transplante y por tanto sé que sería precipitado suponer que este roble es idéntico a aquél con el que encontré un día sólo porque tiene el tipo de edad adecuado y porque se encuentra en el mismo lugar.

La pregunta que debo hacer ahora es la de si los sujetos de experiencia, y más específicamente las *personas*, constituyen un tipo genuino de cosas, cuyos especímenes en su totalidad comparten las mismas condiciones de identidad que nos posibilitan contar personas y seguirles la pista a través del tiempo. A primera vista la respuesta a esta pregunta es claramente 'Sí'. Nos parece suponer que contar personas como mínimo no es tan fácil como contar robles y que las personas tienen condiciones de persistencia que determinan qué tipos de cambios pueden experimentar y cuáles no para pervivir a través del tiempo. Tal vez suceda que no tengamos completamente claro qué decir respecto a determinados casos fronterizos, pero lo mismo sucede con los robles a los cuales se considera, a pesar de ello, como un tipo genuino y natural de cosas.

¿Cuáles son, pues, los principios que rigen el modo en que contamos las personas y les seguimos la pista a través del tiempo? Lo que aquí buscamos es lo que muchos filósofos llamarían un *criterio de identidad personal*.⁷ La forma general que toma un criterio de identidad para los cosas de un determinado tipo *T* es la siguiente:

- (C.) Si *x* e *y* son cosas del tipo *T*, entonces *x* es idéntico a *y* si y sólo si *x* e *y* se encuentran uno con otro en la relación *R_t*.

De manera que la pregunta que hay que hacer es ésta: ¿de qué modo han de estar relacionadas una con otra la persona *P₁* y la persona *P₂* para que *P₁* sea *idéntica* a *P₂*? El problema presenta dos aspectos —con-

⁷ Para mayor información acerca de la noción general de criterio de identidad, véase mi 'What Is a Criterion of Identity?', *Philosophical Quarterly* 39 (1989), pp. 1-21, reimpresso en Harold W. Noonan (ed.), *Identity* (Aldershot: Dartmouth, 1993) y mi 'Objets and Criteria of Identity', en Bob Hale y Crispin Wright (eds.), *A Companion to the Philosophy of Language* (Oxford: Blackwell, 1997).

cerniente el uno a la identidad de las personas *en el mismo momento de tiempo* (identidad 'sincrónica') y concerniente el otro a su identidad *a través del tiempo* (identidad 'diacrónica')—, el segundo de los cuales se considera como el más difícil. Está claro que existen muchas respuestas triviales y nada informativas a la pregunta que se acaba de hacer. Sería verdadero, aunque trivial, decir que, para ser idénticas la una a la otra, P_1 y P_2 deben relacionarse por su *identidad*. Lo que buscamos es una respuesta que no sea trivial y que sea informativa a la pregunta en cuestión. Más generalmente, la relación R , que se menciona en un criterio de identidad de la forma (C), debe no ser la propia relación de identidad ni una relación que pueda solamente enunciarse en términos que comporten o presupongan la identidad de las cosas del tipo T .

Algunos filósofos han mantenido que las personas tienen un criterio *corporal* de identidad. Por ejemplo, ciertos de ellos han sostenido que una persona P_1 y la persona P_2 son idénticas, bien sea en el mismo momento de tiempo o a través del tiempo, si y sólo si P_1 y P_2 tienen el mismo cuerpo. En contra de esto podemos señalar que las personas pueden sobrevivir a la pérdida de muchas de las partes de su cuerpo. Sin embargo, esta no es una consideración decisiva, puesto que es igualmente plausible decir que el *cuerpo* de una persona puede sobrevivir a la pérdida de muchas de sus partes. Una objeción más convincente, si es que puede ser justificada, sería que personas distintas pueden compartir el mismo cuerpo —como se dice algunas veces que ocurre en los casos del denominado síndrome de personalidad múltiple— o que una persona puede intercambiar su cuerpo con otra.⁸ Esta última posibilidad la sugiere el procedimiento, al parecer factible pero hasta ahora no realizado, de trasplante total del cerebro, en el cual el cerebro de una persona se trasplanta al cráneo de otra y a la inversa. A la vista de esta posibilidad, algunos filósofos se inclinan a juzgar que es el tener el mismo *cerebro*, en lugar del mismo *cuerpo*, lo que contribuye a la identidad personal. Sin embargo, tal posición parece inestable, por la razón que sigue (aunque dejemos de lado cualquier problema que puedan plantear los casos de síndrome de personalidad múltiple, en los que el mismo cerebro parece ser compartido por dos personas distintas). Está claro que la razón por la que nos inclinamos a juzgar que una operación de doble trasplante de cerebro constituiría el intercambio de *cuerpos* entre dos personas en lugar del intercambio de sus *cerebros* es que suponemos que todos los aspectos vitalmente importantes de la personalidad humana se basan en rasgos del cerebro; en particular, suponemos que los recuerdos y el temperamento de una persona se basan en rasgos de su cerebro. Pero esto sugiere que lo que *realmente* pensamos que determina la identidad de una persona son estos aspectos de la personalidad.

⁸ Para una discusión de las consecuencias del síndrome de personalidad múltiple, véase Kathleen V. Wilkes, *Real People: Personal Identity Without Thought Experiments* (Oxford: Clarendon Press, 1988), capítulo 4.

humana y, en consecuencia, que se estimaría como posible, al menos en principio, que una persona adquiriera un nuevo cerebro, supuesto que el nuevo cerebro sirviera para basar en él todos los aspectos vitales de la personalidad humana que se basaban en el viejo.⁹

A través de un par de pasos plausibles nos hemos visto llevados desde un criterio corporal de identidad personal, por la vía de un criterio cerebral, hasta un criterio *psicológico* de identidad personal. Según las versiones simples de este último tipo de criterio, una persona P_1 y una persona P_2 son idénticas, bien sea en un momento determinado de tiempo o a través del tiempo si y sólo si P_1 y P_2 comparten determinados aspectos presuntamente vitales de la personalidad humana. Pero ¿cuáles son estos presuntos aspectos vitales de la personalidad humana? ¿Existe, de hecho, algún aspecto de la propia personalidad que no pueda cambiar con el tiempo? Parece difícil suponerlo. Las personas pueden experimentar cambios radicales de temperamento después de sufrir un daño en el cerebro y pueden también padecer formas extremas de amnesia o de pérdida de memoria. No obstante, aunque podamos estar de acuerdo en que la personalidad de alguien pueda verse alterada completamente en un período extenso de tiempo, es menos evidente que alguien pudiera sobrevivir a un cambio repentino y radical de su personalidad que comportara tanto un cambio completo de temperamento como un intercambio global de recuerdos 'viejos' por 'nuevos'. Esto parecería constituir más un caso de *substitución* de una persona por otra que un caso en el que la misma persona sobrevive a un cambio. De modo que quizás un criterio psicológico aceptable de identidad personal debería permitir que una persona experimentara cambios importantes de personalidad con el tiempo, pero sólo si esos cambios tuvieran lugar poco a poco, de una manera gradual. Por ejemplo, según este enfoque, una persona que existe en un cierto momento de tiempo podría *no conservar ninguno* de los recuerdos que poseía esa persona en un momento de tiempo muy anterior, supuesto que, en el segundo de cada dos momentos intermedios que estuvieran separados por un corto intervalo, conservara muchos de los recuerdos que poseía en el primero de esos dos momentos.¹⁰

Sin embargo, incluso esta versión más plausible del criterio psicológico de identidad personal está sujeta a una objeción al parecer devastadora, ya que si suponemos que los aspectos de la personalidad humana que son relevantes para la identidad personal, como el temperamento y la memoria, se basan en rasgos del cerebro, ¿qué ex-

⁹ Para una ulterior discusión de las consecuencias sobre la identidad personal de los trasplantes de cerebro y otros procedimientos relacionados, véase Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, NY: Cornell University Press, 1963), pp. 22 y siguientes, y Bernard Williams, 'The Self and the Future', *Philosophical Review* 79 (1970), pp. 161-80, reimpreso en su *Problems of the Self* (Cambridge: Cambridge University Press, 1973).

¹⁰ Para mayores detalles sobre esta idea de *continuidad* de la memoria, véase Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 204 y siguientes.

cluiría la posibilidad de que hubiera dos cerebros distintos en la que los *mismos* aspectos de la personalidad humana se basaran simultáneamente? Supongamos, por ejemplo, que los mismos aspectos de la personalidad humana —el mismo temperamento y los mismos recuerdos— estuvieran completamente basados en cada uno de los dos hemisferios de un único cerebro, pero que fueran entonces separados y trasplantados a los cráneos de dos cuerpos humanos distintos, a cada uno de los cuales se le hubiera extraído previamente el cerebro.¹¹ Entonces, según parece, nos veríamos confrontados con dos personas diferentes, cada una de ellas con sólo la mitad de un cerebro, que compartirían el mismo temperamento y los mismos recuerdos. Pero esto parece ser incompatible con el criterio psicológico de identidad personal que se ha sugerido, el cual implica que las personas que comparten el mismo temperamento y los mismos recuerdos son idénticas. Con lo que aquí nos las habemos es al parecer con un caso de *fisión personal*, es decir, la división de una persona, P_1 , en dos personas distintas, P_2 y P_3 . Las leyes lógicas de la identidad nos impiden decir que P_2 y P_3 son *ambas* idénticas a P_1 , dado que no son idénticas una con otra, y sin embargo tanto P_2 como P_3 se relacionan psicológicamente con P_1 de una manera que el criterio psicológico de identidad personal propuesto estima que es suficiente para la identidad. Al parecer, pues, el criterio debe estar equivocado.

Una respuesta plausible a esta objeción consiste en revisar el criterio psicológico de identidad personal propuesto de tal manera que una persona que exista en un momento de tiempo posterior se considere idéntica a una que exista en un momento anterior sólo si no se ha dado un proceso de fisión, es decir, sólo si la persona existente en el momento posterior es la *única* persona que existe en ese momento de tiempo que se relaciona al modo psicológicamente pertinente con la persona que existía en el momento anterior. Sin embargo, si se adopta esta línea de respuesta, parece tener la consecuencia de que las cuestiones de identidad personal son mucho menos importantes, tanto moralmente como emocionalmente, de lo que intuitivamente pensamos que son,¹² toda vez que, de acuerdo con esta concepción, el que yo exista o no exista mañana puede depender de la respuesta a la cuestión aparentemente sin importancia de si *algún otro* muy parecido a mi existirá mañana. Si estoy destinado a experimentar una fisión dejaré de existir, pero algún

¹¹ Las consecuencias más generales para el concepto de identidad personal de la posibilidad de bisección cerebral las trata Thomas Nagel en 'Brain Bisection and the Unity of Consciousness', *Synthese* 22 (1971), pp. 396-413, reimpresso en su *Mortal Questions* (Cambridge: Cambridge University Press, 1979). El caso hipotético de la bisección del cerebro seguido por el trasplante separado de los dos hemisferios lo analiza Derek Parfit en su 'Personal Identity' *Philosophical Review* 80 (1971), pp. 3-27, reimpresso en Jonathan Glover (ed.), *The Philosophy of Mind* (Oxford: Oxford University Press, 1976). Véase también Parfit, *Reasons and Persons*, capítulo 12.

¹² Esta es la opinión de Derek Parfit, quien sugiere también la noción de 'supervivencia' que vamos a tratar en breve; véase de nuevo su *Reasons and Persons*, capítulos 12 y 13.

otro existirá mañana que se relacionará conmigo de manera psicológicamente exacta, del mismo modo en que yo me relaciono con mi yo de ayer. En realidad, mañana existirán *dos* personas tales. Sin embargo, si lo que a mí me importa, cuando considero las perspectivas de mi supervivencia, es simplemente que mañana exista alguien que se relacione conmigo psicológicamente en el mismo modo en que yo me relaciono con mi yo de ayer, ¿por qué debería preocuparme el que exista mañana *más de una* persona tal? Pero si esto no debiera preocuparme, se sigue, de acuerdo con la concepción de la identidad personal que ahora se está proponiendo, que no debería necesariamente preocuparme el que no exista nadie mañana que sea *idéntico* a mí. En realidad, tal vez yo no debiera pensar en mi 'supervivencia' en términos de *identidad* —es decir, en términos de que exista en el futuro alguien que sea idéntico a mí— sino más bien en términos de que haya *al menos una* (pero posiblemente más de una) persona que exista en el futuro que se relacione conmigo psicológicamente del mismo modo en que yo me relaciono con mi yo en el pasado. Según este modo de entender el término 'supervivencia', si experimentase un proceso de fisión 'sobreviviría' dos veces, aunque nadie que existiera después de la fisión sería idéntico a mí.

El resultado de nuestra discusión hasta el momento es que no hemos encontrado un criterio de identidad personal, corporal o psicológico, que sea totalmente satisfactorio en el sentido de que sus veredictos no entren en conflicto con nuestras creencias intuitivas concernientes a la naturaleza de la identidad personal y su significación moral y emocional. Algunos filósofos se inclinan por extraer la conclusión de que nuestras creencias intuitivas concernientes a la identidad personal son confusas o contradictorias, sugiriendo incluso que el concepto mismo de 'persona' es de algún modo confuso. Una sugerencia alternativa es que algunos de los casos imaginarios que parecen crear dificultades para algunos de los criterios de identidad propuestos están ellos mismos mal concebidos y no representan posibilidades genuinas.¹³ Pero una posible tercera línea de respuesta, que no deberíamos desechar demasiado a la ligera, consiste en aventurar la sospecha (sin por ello pretender en modo alguno impugnar el concepto de persona) de que de hecho *no hay ningún* criterio de identidad personal que no sea trivial y que sea informativo. Quizá la identidad personal sea algo básico que no puede explicarse en términos más fundamentales. No intentaré aquí decidir entre estas tres respuestas, aunque mis propias simpatías recaen en la última.¹⁴

¹³ Para una crítica del recurso que se hace a los experimentos mentales en las discusiones filosóficas sobre la identidad personal, véase Wilkes, *Real People*, capítulo 1.

¹⁴ Desarrollo de manera más completa mis propias ideas sobre la identidad personal en mi *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms* (Oxford: Blackwell, 1989) capítulo 7 y en mi *Subjects of Experience* (Cambridge: Cambridge University Press, 1996) capítulo 2. Para más información sobre la identidad personal en general, véase Harold W. Noonan, *Personal Identity* (Londres: Routledge, 1989) y Brian Garrett, *Personal Identity and Self-Consciousness* (Londres: Routledge, 1998).

LA MEMORIA PERSONAL

El concepto de *memoria* ha tenido gran importancia en nuestra discusión de la identidad personal, pero no hemos analizado ese concepto con el cuidado que se merece. De hecho, no hay un *único* concepto de memoria sino muchos distintos. La memoria de hechos es diferente de la memoria de habilidades prácticas y ambas son igualmente distintas de la que se denomina a veces memoria *personal* o *autobiográfica*. Es esta última clase de memoria la más patentemente relevante para las cuestiones de identidad personal. La memoria personal es el recuerdo de alguna experiencia pasada o acción como propia, es decir, es el recuerdo de *experimentar o hacer algo*, y no meramente un recuerdo de *que se experimentó o se hizo algo*. Yo podría recordar *que hice una determinada cosa cuando era niño porque alguien que me observó habiéndolo me lo ha dicho y recuerdo este hecho sobre mí mismo*. Esto es evidentemente muy distinto de simplemente recordar *haber hecho esa cosa*. Como acabo de indicar, los recuerdos de hechos y los personales se relatan característicamente de distintas maneras, esto es, mediante oraciones que tienen distintas estructuras gramaticales. Así, por una parte podría decir 'Recuerdo que el pasado verano fui a la playa', y por la otra podría decir 'Recuerdo haber ido a la playa el pasado verano'; la primera oración se interpreta de modo más natural como un informe del recuerdo de un hecho, mientras que la interpretación más natural del segundo es como un informe de un recuerdo personal.¹⁵

Otra distinción que deberíamos hacer es la que distingue entre memoria *disposicional* y memoria *en acto*, es decir, entre tener la capacidad para recordar algo y recordarlo realmente. Ahora mismo hay muchas cosas que *puedo* recordar haber hecho en el pasado, aunque no estoy en el presente recordándolas porque estoy pensando en otras cosas. El acto mental de recordar alguna experiencia o acción pasada comporta, según parece, elementos de la experiencia *presente*. Esa es la razón por la que las personas describen a veces tal acto como uno en que se 'vuelve a vivir' una experiencia pasada, aunque raramente sucede que la experiencia de 'volver a vivir' una experiencia pasada sea tan 'vivida' como la original. Por tanto, en algunos aspectos las experiencias de la memoria personal son como las experiencias involucradas en actos de *imaginación* (véase el capítulo 7). En ambos casos describimos las experiencias implicadas como algo que exhibe características sensoriales o como algo que *cae en* la competencia de diferentes modalidades sensoriales; discriminaremos así entre recuerdos e imaginaciones *visuales* y *auditivas*. En

¹⁵ Para mayor información sobre la memoria personal o (como se la llama a veces) memoria *experiencial*, véase Richard Wollheim, *The Thread of Life* (Cambridge: Cambridge University Press, 1984), capítulo 4.

un acto de recuerdo visual se recuerda la *apariencia visual* de algo, mientras que en un acto de recuerdo auditivo se recuerda cómo *sonaba* algo.

Aquí puede plantearse la siguiente pregunta: dado el parecido entre las experiencia de recuerdos personales y las experiencia involucradas en los actos de imaginación, ¿cómo distinguimos entre las dos cuando nosotros mismos estamos teniendo esas experiencias? ¿Cómo diferenciamos si estamos recordando algo o solamente imaginándolo? La pregunta sería sensata si supusiéramos que recordar o imaginar no son nada más que simplemente tener algún tipo de experiencia pero, por supuesto, son mucho más que eso. Recordar e imaginar son *actos mentales*, cosas que *hacemos*, muy a menudo, de manera totalmente intencionada. En consecuencia, si se me pide que trate de recordar haber ido a la playa el pasado verano, esto es algo que resultará que puedo hacer fácilmente, por el simple hecho de querer hacerlo. Similarmente me encuentro con que puedo *imaginar* haber ido a la playa. Pero si se me pregunta entonces *cómo sé* que 'lo que estoy haciendo es en el primer caso recordar, en lugar de imaginar, haber ido a la playa el pasado verano', no estoy nada seguro de qué sentido puedo darle a esa pregunta. El hecho de que las experiencias concomitantes puedan ser muy similares es al parecer irrelevante, puesto que yo no juzgo si estoy recordando o imaginando sobre la base del tipo de experiencia que tengo. Más bien sé si estoy recordando o imaginando porque estas son cosas que puedo hacer intencionadamente, y no se puede hacer algo intencionadamente a menos que se sepa que se está haciendo. La pregunta que realmente deberíamos hacer es, posiblemente, no cómo sé si estoy recordando o meramente imaginando haber ido a la playa el pasado verano, sino cómo sé que realmente *fui a la playa el pasado verano*, dado que 'recuerdo' haber ido.

Nos podríamos ver tentados a responder a esta pregunta diciendo que, si *recordamos* haber ido a la playa el pasado verano, tenemos que haber ido, porque no se puede 'recordar' haber hecho algo que no se haya hecho, igual que no se puede 'saber' algo que no sea el caso o 'ver' algo que no existe. Una manera de decir esto es que 'saber', 'ver' y 'recordar' son todos ellos verbos de *éxito* o *logro*. Sin embargo, esta respuesta se apoya en una observación puramente lingüística concerniente al significado convencional de la palabra 'recuerdo', y por tanto no parece ir al fondo del asunto. Dado que podemos como mínimo *parecer* recordar haber hecho algo que no hicimos, ¿cómo sabemos, sobre la base de ese parecer recordar, que realmente lo hicimos? Tal vez la respuesta a esta pregunta sea que *no podemos* saber con seguridad que hicimos algo meramente sobre la base de ese parecer recordar haberlo hecho, pero que, a pesar de ello, es razonable creer sobre esa base que lo hicimos, a menos que poseamos indicios en contra. Es dudoso que podamos coherientemente poner en cuestión la veracidad de *todo* lo que parecemos recordar, pero ciertamente deberíamos reconocer la falibilidad de cualquier intento particular de recordar lo que ocurrió en el pasado.

Dado que erróneamente nos puede parecer haber hecho algo que de hecho no hicimos, se plantea entonces la pregunta de si nos puede parecer recordar haber hecho algo que de hecho hizo *algún otro*. Realmente esto es así. Pero, ¿podría ocurrir que nos pareciera recordar haber hecho algo y que *estuviéramos* recordando haberlo hecho, si no fuera por el hecho de que fue algún otro el que lo hizo? Si la fisión personal del tipo que contemplábamos antes fuera posible, la respuesta a esta pregunta parece ser 'Sí', ya que en tal caso a *las dos* personas, P_2 y P_3 , en las que la persona original, P_1 , se divide, les parece recordar haber hecho lo que P_1 hizo y todo lo que nos impide decir que tanto P_2 como P_3 *recuerdan* haber hecho lo que P_1 hizo es que ninguna de ellas lo hizo de hecho, ya que fue P_1 quien lo hizo y ninguna de ellas es idéntica a P_1 . Algunos filósofos introducen el término 'cuasi-recuerdo' para describir lo que ocurre en tales casos.¹⁶ El concepto de cuasi-recuerdo se supone que es más amplio que el de recuerdo personal tal como a éste se lo concibe comúnmente. Sólo podemos tener un recuerdo personal *verídico* de lo que de hecho hicimos nosotros, pero uno puede tener un *cuasi-recuerdo* verídico de haber hecho algo que de hecho hizo alguna otra persona, supuesto que uno esté relacionado con esa persona de la misma manera en que uno está relacionado usualmente con el propio yo anterior, como supuestamente sucede en el caso de la fisión personal. Sin embargo, no todos los filósofos están satisfechos con la noción de cuasi-recuerdo, pues algunos de ellos sospechan que es incoherente. Si la fisión personal fuera posible, entonces sería posible teóricamente que —sin que yo lo sepa— yo mismo fuera un producto de tal fisión, en cuyo caso muchos de mis 'recuerdos' personales serían realmente sólo cuasi-recuerdos. Pero es tentador afirmar que la memoria personal muestra una forma de 'inmunidad al error debido a identificación equivocada', en el sentido de que si a mí me parece recordar haber hecho algo, no puede suceder que esté equivocado sólo por el hecho de que no fui yo quien hizo eso.¹⁷ Si ello es así, entonces parecería seguirse que, después de todo, la fisión personal es una imposibilidad. No intentaré, sin embargo, resolver aquí esta difícil cuestión.

Una razón por la cual algunos filósofos defienden la noción de cuasi-memoria es la de evitar la acusación de circularidad que se hace a veces contra cualquier criterio psicológico de identidad personal que apele a la memoria personal. Cualquier criterio de identidad personal que sostenga que una persona, P_2 , existente en un momento de tiempo dado, sea idéntica a una persona, P_1 , existente en un momento anterior,

¹⁶ La noción de 'cuasi-recuerdo' se debe a Sydney Shoemaker; véase su 'Persons and their Pasts' *American Philosophical Quarterly* 7, (1970), pp. 269-85, reimpresso en su *Identity, Cause and Mind*. Derek Parfit adopta esa noción; véase su 'Personal Identity' y su *Reasons and Persons*, pp. 220 y siguientes.

¹⁷ Para las dudas acerca de la coherencia de la noción de cuasi-recuerdo, véase Evans *The Varieties of Reference*, pp. 242-8, Wollheim, *The Thread of Life*, pp. 111-17 y Andy Hamilton, 'A New at Personal Identity', *Philosophical Quarterly* 45 (1995), pp. 332-49.

sólo si P_2 recuerda haber hecho algunas de las cosas que P_1 hizo, parece comportar circularidad, porque una condición necesaria de que P_2 recuerde haber hecho algo es que P_2 debe realmente haber hecho eso. Por tanto, no se puede establecer que P_2 realmente *recuerda* haber hecho algo que hizo P_1 , a menos que se haya establecido que P_2 es en realidad *idéntica* a P_1 , la cual cosa es lo que se suponía que el criterio nos posibilitaba determinar. (Recuérdese mi observación anterior de que la relación R , que se menciona en un criterio de identidad de la forma de (C), no debe ser la propia relación de identidad, ni ninguna relación que pueda enunciarse solamente en términos que comporten o presupongan la identidad de las cosas del tipo T ; es esta segunda restricción la que parece que se viola en el caso que estamos discutiendo.) La acusación de circularidad puede evitarse substituyendo en tal criterio la referencia al recuerdo por la referencia al cuasi-recuerdo. Pero si resultase que el concepto de cuasi-recuerdo es incoherente, ello parecería tener la consecuencia de que no disponemos de ningún criterio psicológico de identidad personal que no sea circular, dado que el concepto de recuerdo personal debería entonces figurar centralmente en cualquier criterio de esa naturaleza.

MEMORIA Y CAUSALIDAD

Una cuestión que no hemos tratado hasta ahora es la siguiente. ¿En virtud de qué un acto o estado mental mío presente es un *recuerdo* personal de alguna experiencia o acción mía anterior? Es plausible pensar que se requiere una conexión *causal* entre la experiencia o la acción anterior y el acto o el estado mental presentes.¹⁸ Pero, ¿qué clase de conexión? No serviría cualquier tipo de conexión causal, ya que supongamos que alguna otra persona hubiera sido testigo de una acción anterior mía y me dijera posteriormente que la había realizado, aunque yo mismo hubiera olvidado haberla hecho. Es concebible que, transcurrido un tiempo, yo pudiera olvidar que se me dijo que hice eso y podría empezar a *parecer* recordar haberlo hecho, aunque realmente no lo estuviera recordando. En tal caso, mi estado mental presente de parecer recordar haber hecho eso está conectado causalmente con mi acto original de hacerlo, pero evidentemente no lo está del modo correcto para que ese estado mental pueda considerarse como un *recuerdo* personal genuino de haber hecho lo que fuere. ¿Cuál es, pues, el tipo 'correcto' de conexión causal?

Nos encontramos aquí de nuevo con el problema de las 'cadenas causales desviadas', que ya encontramos al tratar las teorías causales de la percepción (en el capítulo 6) y las teorías causales de la acción (en el

¹⁸ En el importante artículo de C. B. Martin y Max Deutscher, 'Remembering', en *Philosophical Review* 75 (1966), pp. 161-96, se defiende una explicación causal del recuerdo.

capítulo 9). Sabemos que ese problema es en general un problema difícil, de modo que lo que cabe esperar es que resulte ser también difícil en el caso particular del recuerdo. No intentaré resolverlo aquí. Pero es obvio que el tipo de conexión causal que comporta el recuerdo es una conexión 'interna' a la persona que tiene el recuerdo. Es esto lo que sustenta a diversas teorías de la memoria que recurren a la noción de *huella* de la memoria, concebida como un estado físico del cerebro que ha dejado un episodio original de experiencia sensorial y que es capaz de inducir la subsiguiente experiencia de recuerdo de ese episodio original. Pero es preciso decir que algunos filósofos son contrarios a esta noción y en general a las teorías causales del recuerdo, por razones similares a las que motivan su hostilidad a las teorías causales de la percepción.¹⁹ En particular consideran que tales teorías fomentan el escepticismo acerca de la memoria y por tanto acerca de nuestro conocimiento del pasado. Del mismo modo en que tales filósofos propondrían una teoría 'disyuntiva' de la percepción (véase el capítulo 6) que presentase a la percepción verídica como 'directa', podrían defender la idea de que el recuerdo verídico constituye un conocimiento 'directo' del pasado. Sin embargo, sus oponentes tenderán a considerar esa idea como oscura y difícilmente compatible con una concepción naturalista de los seres humanos y de sus capacidades psicológicas. De nuevo, esta no es una disputa que se vaya a tratar de resolver aquí.

EL ANIMALISMO

Anteriormente vimos que no es fácil encontrar un criterio de identidad personal que no presente problemas. Algunos filósofos se inclinarían por extraer la conclusión de que el término 'persona' no denota un tipo diferenciado de cosas, cuyos especímenes comparten las mismas condiciones de identidad. Podemos aquí sacar a colación la tesis, mencionada al comienzo de este capítulo, de que ser una persona o un yo consiste en desempeñar un papel o realizar una función —un papel o una función que podría ser desempeñado o realizado por muchas cosas distintas. La cada vez más popular doctrina que se conoce como *animalismo* se adhiere a una versión de esta tesis.²⁰ Según el animalismo, *somos* realmente nuestro cuerpo, es decir, no somos ni más ni menos que organismos humanos vivos que empiezan a existir como células únicas

¹⁹ Para una panorámica sobre los ataques a la noción de huella de la memoria, véase John Sutton, *Philosophy and Memory Traces: Descartes to Connectionism* (Cambridge: Cambridge University Press, 1998), capítulo 16. Véase también Norman Malcolm, 'A Definition of Factual Memory' en su *Knowledge and Certainty: Essays and Lectures* (Ithaca, NY: Cornell University Press, 1963).

²⁰ Posiblemente, la defensa del animalismo más plenamente desarrollada hasta la fecha sea la de Eric T. Olson; véase su *The Human Animal: Personal Identity Without Psychology* (Nueva York: Oxford University Press, 1997).

(un zigoto) y que continúan existiendo hasta su muerte biológica. Durante parte de ese tiempo, diversos tipos de actividad de nuestro cerebro y nuestro sistema nervioso sustentan el tipo de conciencia —incluyendo el consciente percatarse de uno mismo— que posibilita que se nos describa como 'personas'. Pero existimos antes de ser personas y puede ser también que existamos durante algún tiempo después de dejar de ser personas. Por ello no somos *esencialmente* personas, puesto que una propiedad esencial es una propiedad sin la cual el que la tiene no podría existir. Nuestras condiciones de identidad son, por lo tanto, simplemente las de un organismo humano. Pero si hubiera cosas de otro tipo —tal vez robots adecuadamente programados— que pudieran ser personas, podrían tener condiciones de identidad totalmente diferentes de las nuestras. Por consiguiente, según esta concepción es sencillamente un error buscar un criterio de identidad que se aplique a todas las cosas que podrían ser personas y sólo a ellas, y es tentador atribuir a este presunto error los problemas que encontramos anteriormente al pretender establecer tal criterio.

Una ventaja del animalismo es, según parece, la siguiente. Si suponemos que *no somos* idénticos a nuestro cuerpo, ni a ninguna parte del mismo, como el cerebro, pero que aun así somos cosas físicas que se extienden en el espacio y que tienen propiedades físicas como el peso y la solidez, entonces parece que hemos de *coincidir* con algo distinto de nosotros mismos. Por ejemplo, si mi peso y mi forma son los mismos que los de mi cuerpo, entonces yo debo coincidir con mi cuerpo, que estamos suponiendo que es distinto a mí. Pero, ¿cómo pueden dos cosas distintas coincidir exactamente, es decir, ocupar exactamente la misma región del espacio en el mismo momento temporal? Además, si tanto yo como mi cuerpo tenemos un determinado peso y sin embargo somos dos cosas distintas, ¿cómo es que nuestro peso combinado no es el doble del de mi cuerpo? Ninguna de estas evidentes dificultades se plantea con el animalismo, dado que sostiene que yo soy realmente idéntico a mi cuerpo.

No obstante, estas presuntas dificultades pudieran no ser tan graves como sostiene el animalista, pues, como ya indiqué en el capítulo 2, el adversario del animalismo podría apoyarse en la analogía entre una estatua y el trozo de bronce del cual ésta está compuesta. También la estatua y el trozo de bronce parecen ser dos cosas distintas que coinciden exactamente, siendo su distinción una consecuencia de las diferencias entre sus respectivas condiciones de persistencia. La estatua puede sobrevivir a cambios a los que no puede sobrevivir el trozo de bronce, y a la inversa, y la razón por la que el peso 'combinado' de la estatua y el trozo de bronce es sólo el del trozo de bronce es que la estatua debe enteramente el peso que pueda tener a la materia de la que está compuesta. Aunque la estatua y el trozo de bronce sean objetos *distintos* —es decir, no sean idénticos uno a otro—, no sucede como si fueran objetos *separados*, como podrían serlo dos trozos distintos de bronce. Por ello,

una concepción posible de la relación entre nosotros y nuestro cuerpo que compite con el animalismo es la de que estamos *compuestos* o *constituidos* por nuestro cuerpo. Como dejé claro en el capítulo 2, este no es mi punto de vista, pero aquí todo lo que quiero sostener es que el animalismo merece menos crédito del que algunos de sus partidarios imaginan que tiene, sólo porque evita las presuntas dificultades que hemos venido comentando, ya que estas presuntas dificultades han sido, según creo, exageradas por los partidarios del animalismo.

Otra presunta dificultad que el animalismo parece evitar es la siguiente. Está claro que somos los sujetos de determinados estados mentales, es decir, que tenemos determinados pensamientos, determinadas sensaciones y determinados sentimientos. Pero, según sostiene el animalista, es igualmente verdad que el organismo humano que es nuestro cuerpo tiene determinados pensamientos, sensaciones y sentimientos, en virtud de la actividad de su cerebro y su sistema nervioso. Sin embargo, si no somos idénticos a nuestro cuerpo, entonces parece que hay dos cosas distintas que tienen exactamente los mismos pensamientos, sensaciones y sentimientos en el mismo momento temporal, a saber, cada uno de nosotros y su cuerpo. Pero esto parece absurdo, y este absurdo parece que se evita al mantener, como hace el animalista, que cada uno de nosotros es (idéntico a) su cuerpo.

Sin embargo, un adversario del animalismo podría cuestionar la afirmación del animalista de que el organismo humano que es nuestro cuerpo tenga pensamientos y sensaciones. Dista mucho de ser una verdad evidente que tanto mi cuerpo como un todo como una cualquiera de sus partes —como mi cerebro— tengan literalmente pensamientos y sensaciones. Ciertamente parece que es verdad que no podríamos tener pensamientos y sensaciones si no tuviéramos un cerebro que funcionase de determinada manera. Pero es igualmente verdad que no podríamos correr de no tener unas piernas que funcionan de una determinada manera, y sin embargo, este último hecho no implica que *nuestras piernas* corran, por lo que ¿cuál es la razón por la que debería pensarse que el hecho anterior implica que *nuestro cerebro* tenga pensamientos y sensaciones? Tenemos que distinguir entre un *sujeto* de pensamientos y sensaciones —algo que piensa y siente— y algo —como un cerebro— cuya actividad hace posible y sustenta los pensamientos y las sensaciones de un sujeto.

Ahora bien, el animalista podría conceder que no es *nuestro cerebro* el que tiene pensamientos y sensaciones, porque lo que él quiere decir es que somos *nosotros* los que tenemos pensamientos y sensaciones y que cada uno de nosotros es idéntico a su cuerpo —un organismo vivo—, como un todo en vez de un órgano que se halla dentro de ese cuerpo, es decir, su cerebro. Sin embargo, cuando se mantiene que *el cuerpo como un todo* puede literalmente tener pensamientos y sensaciones, parece difícil negar que *un cerebro* pueda literalmente tener pensamientos y sensaciones, porque parece que en principio es perfectamen-

te posible que un cerebro separado del resto pudiera ser conservado con vida y seguir funcionando de modo que sirviera para sustentar pensamientos y sensaciones. En tal caso, el animalista no puede decir que sea la persona de cuyo cerebro se trata y no el cerebro mismo el que tenga pensamientos y sensaciones, porque entonces se vería forzado a admitir la afirmación que él rechaza, a saber que una persona pueda coincidir con algo distinto a ella misma (en este caso, un cerebro). De manera que deberá decir que es el cerebro el que tiene pensamientos y sensaciones en un caso así. Pero si un cerebro separado que sigue funcionando tiene pensamientos y sensaciones, se sigue claramente que un cerebro que se encuentre en un cuerpo y que funcione tiene también pensamientos y sensaciones, lo que significa que podemos volver la propia objeción del animalista contra sí mismo, ya que ahora parece verse forzado a admitir que *tanto* mi cuerpo en su totalidad como su correspondiente cerebro tienen pensamientos y sensaciones, es decir, de nuevo, que existen dos cosas distintas que tienen exactamente los mismos pensamientos y sensaciones en el mismo momento temporal.

En vista de esta consecuencia, es bastante urgente que el animalista revise su posición y mantenga que, después de todo, es el cerebro de cada uno de nosotros y *no* el cuerpo como un todo el que tiene pensamientos y sensaciones. Pero entonces se ve forzado a aceptar la nada atractiva doctrina de que somos idénticos a nuestro cerebro, puesto que sería absurdo negar que *nosotros* tenemos pensamientos y sensaciones. Lo que creo que se desprende de estas consideraciones es que el animalismo es una teoría inestable, pues se ve amenazado de quedar confundido con la teoría de que yo soy idéntico a mi cerebro. Podemos evitar esta declaración de identidad distinguiendo entre nosotros mismos, sujetos de pensamientos y sensaciones, y nuestro cerebro, el órgano corporal cuya actividad hace posible y sustenta nuestros pensamientos y sensaciones. Pero si aceptamos esta distinción, no hay realmente nada que motive la afirmación original de los animalistas de que cada uno de nosotros es idéntico a su cuerpo como un todo, un organismo vivo. Realmente, si, como parece plausible intuitivamente, pudiéramos sobrevivir con nuestro cerebro separado del resto de nuestro cuerpo, con certeza no podríamos ser idénticos a nuestro cuerpo como un todo. El animalismo simplemente es contrario a nuestras más profundas intuiciones relativas a nuestras propias condiciones de persistencia: creemos que podríamos sobrevivir a cambios a los que nuestro cuerpo no podría sobrevivir. Por más que pueda ser difícil especificar esas condiciones de persistencia de un modo preciso y no sujeto a controversia, parece suficientemente claro que difieren de las condiciones de persistencia de los organismos vivos. Podemos incluso concebir que sobrevivimos a un proceso de cambio en el cual nuestro cuerpo y nuestro cerebro orgánicos se ven substituidos gradualmente por un cuerpo y un cerebro inorgánicos, compuestos de metal y de silicio. Además, cuando reflexionamos sobre ello, el animalismo simplemente parece equivocado del todo

en su afirmación de que podríamos seguir existiendo como un cuerpo inconsciente después de que nuestro cerebro haya dejado de sustentar cualquier pensamiento o sensación consciente. Según creo, después de meditarlo bien, nuestra opinión sería que un cuerpo que se encuentre en esa condición no es sino una especie de cadáver viviente y que la persona de cuyo cuerpo se trataba ha dejado de existir.

EL CONOCIMIENTO DE LA PROPIA MENTE

Sugerí anteriormente que una persona, o 'yo', se distingue entre los sujetos de experiencia por tener no sólo pensamientos y sensaciones, si no también un conocimiento de al menos algunos de esos pensamientos y esas sensaciones, en particular de sus pensamientos y sus sensaciones conscientes. Las personas pueden identificar sus pensamientos y sus sensaciones conscientes como suyos y a menudo tienen, o parecen tener, un conocimiento fiable y autorizado de qué es lo que conscientemente están pensando y sintiendo.²¹ Suponiendo que tal conocimiento realmente existe, ¿cómo es posible? Es muy tentador ofrecer un modelo *observacional* de la adquisición de tal conocimiento, es decir, suponer que descubrimos hechos acerca de nuestros propios pensamientos y sensaciones mediante un proceso de 'introspección' que es análogo a los procesos perceptuales mediante los que descubrimos hechos acerca de acaecimientos y estados de cosas en el mundo en general. Pero existen dificultades en ese modelo de las fuentes del conocimiento de nosotros mismos. Para empezar, no es nada evidente que tengamos de hecho ningún tipo de 'sentido interno' análogo a los 'sentidos externos' de la visión, el oído, el olfato, etcétera. Además, un modelo observacional de las fuentes del conocimiento de nosotros mismos parece estar mal preparado para explicar el tipo especial de fiabilidad y autoridad que se supone que caracteriza a nuestro conocimiento acerca de nuestros propios estados mentales conscientes. Nuestra tendencia es suponer que cada persona es en general un juez mejor de lo que él mismo está conscientemente pensando y sintiendo de lo que pueda serlo ninguna otra, aunque concedamos que nadie es completamente infalible en tales cuestiones. Pero no está claro que el modelo observacional pueda explicar este hecho, si es que realmente es un hecho, porque la observación es por regla general una fuente de información no muy fiable, sujeta como está a diversas variedades de prejuicios e ilusiones.

A favor del modelo observacional podría aducirse que cada persona puede 'observar' *directamente* sus propios pensamientos y sus sensaciones conscientes de un modo que no está a disposición de un 'obser-

²¹ Para más comentarios sobre nuestra capacidad de identificar nuestros pensamientos, sensaciones y sentimientos conscientes como nuestros, véase mi *Subjects of Experience*, capítulo 7.

vador externo', el cual siempre se ve obligado a *inferir* lo que otra persona está pensando y sintiendo a partir de observaciones de la conducta corporal de esa persona, y que esa es la razón por la que cada persona tiene una autoridad especial acerca de los propios pensamientos y sensaciones. Pero, aun dejando a un lado el supuesto cuestionable de que la observación 'directa' sea de alguna manera especialmente fiable, esta afirmación parece considerar los pensamientos y las sensaciones como objetos necesariamente *privados* de la observación 'interna' de un modo que se ve expuesto a fuertes críticas.²² Si mis propios pensamientos y sensaciones son cosas que sólo yo puedo observar directamente, ¿cómo puedo ni siquiera entender la proposición de que otras personas, además de mí mismo, pueden tener sus propios pensamientos y sensaciones? Si trato de pensar en 'sus' pensamientos y sensaciones por analogía con los míos, entonces —según el modelo observacional— habré de concebirlos como cosas que otras personas pueden observar directamente pero que necesariamente yo no puedo observar así. Pero, ¿qué me permite a mí denominar a esas cosas hipotéticas *pensamientos* y *sensaciones* cuando, por hipótesis, se trata de cosas que es imposible que yo pueda comparar, para ver si tienen algún parecido con mis propios pensamientos y sensaciones? Parece imposible que yo pueda adquirir conceptos perfectamente generales de pensamiento y sensación, aplicables a otras personas además de a mí mismo, solamente a partir de una extraña clase de autoobservación mental, y sin embargo eso es lo que parece requerir el modelo observacional de las fuentes de conocimiento de uno mismo.

Pero, ¿qué alternativa hay frente al modelo observacional? Un primer paso hacia una posible alternativa consiste en centrarnos en la observación de que parecemos tener una forma de conocimiento especialmente fiable y autorizada sólo respecto de nuestros propios pensamientos y sensaciones conscientes, toda vez que —recordando ahora nuestros comentarios sobre la conciencia en el capítulo 3— podría sugerirse que el hecho de que un estado mental sea *consciente* consiste justamente en que es el objeto de otro estado mental de orden superior. Del mismo modo en que podemos creer que *p*, podemos, según parece, creer que creemos que *p* y, según la concepción que estamos ahora discutiendo, en esto es realmente en lo que consiste el hecho de que nuestra creencia de que *p* sea una creencia 'consciente'. Similarmente, el hecho de que una determinada sensación mía, como lo es una sensación de dolor, sea consciente consiste, según esta concepción, únicamente en que no sólo la tengo, sino que *creo* que la tengo. Se estaría sosteniendo que algunas criaturas son capaces de tener sólo creencias 'de primer orden' y que en consecuencia care-

²² La crítica de las concepciones que consideran los pensamientos y las sensaciones como 'objetos privados' debe mucho a la obra de Wittgenstein; para una exposición y comentario de las partes relevantes de la obra de Wittgenstein, véase Malcolm Budd, *Wittgenstein's Philosophy of Psychology* (Londres: Routledge, 1989), capítulo 3.

cen de pensamientos conscientes de acuerdo con ese modelo, pero que nosotros los humanos somos seres pensantes conscientes porque tenemos también creencias 'de segundo orden', es decir, creencias acerca de nuestras creencias de primer orden.²³

Si el hecho de que una creencia de que *p* sea consciente sólo consiste en que su sujeto crea que cree que *p*, entonces la razón por la que los sujetos tienen un conocimiento especialmente fiable y autorizado de sus propias creencias conscientes no parecerá nada misteriosa, toda vez que es fácil que pensemos que es un principio evidentemente correcto el de que si *S* cree que cree que *p*, entonces, ciertamente, *S* cree que *p*, y si este principio es correcto, entonces siempre que *S* crea que cree que *p*, su creencia de que cree que *p* será *verdadera*, con lo que constituirá un *conocimiento* (bajo el supuesto de que las creencias verdaderas que son fiablemente producidas constituyan un conocimiento). Sin embargo, aunque pensemos que este principio sea evidente, lo que realmente se requiere es que averigüemos *por qué* es correcto, si es que realmente lo es; ¿por qué habría de suceder que si *S* crea que cree que *p*, entonces *S* cree que *p*? No estamos simplemente dando por supuesto aquí el propio fenómeno del conocimiento de nosotros mismos que estamos tratando de explicar? Con todo, quizás estemos al menos dirigiendo ahora nuestra mirada al lugar adecuado para encontrar una solución a nuestros problemas, que es algo que muy bien puede argumentarse que no estábamos haciendo cuando nos hallábamos tentados por el modelo observacional. Por ejemplo, podría afirmarse que se puede explicar sobre bases evolutivas por qué las creencias de una criatura acerca de sus propias creencias, si es que las tiene, deberían en general ser sumamente fiables; ello se debería a que cabría esperar que cualquier criatura que creyera frecuentemente tener creencias que de hecho no tuviera mostrará un comportamiento tan autodestructor que su supervivencia sería improbable. De acuerdo entonces con esta sugerencia, el principio de que si *S* cree que cree que *p* entonces *S* cree que *p*, aunque no sea una verdad lógica o *a priori*, se aproxima mucho a una ley psicológica; puede que tenga contraejemplos, pero tenemos razones para esperar que sean pocos y que éstos se den aisladamente.

LA PARADOJA DE MOORE Y LA NATURALEZA DE LAS CREENCIAS CONSCIENTES

Existen, sin embargo, dificultades también en este enfoque alternativo del problema del conocimiento de nosotros mismos. Para empezar, no es nada evidente que una explicación de la conciencia en términos

²³ Una defensa de una teoría de la conciencia como pensamiento de orden superior es la de Peter Carruthers en su *Language, Thought and Consciousness: An Essay in Philosophical Psychology* (Cambridge: Cambridge University Press, 1996), capítulos 5, 6 y 7.

de estados mentales de orden superior sea totalmente satisfactoria; como ya comenzábamos a ver en el capítulo 3. ¿Cómo podría ser consciente una creencia sólo en virtud de que su sujeto tuviera una creencia acerca de esa creencia? El hecho de que una creencia sea una creencia *consciente* tiene algo que ver con la naturaleza intrínseca de ese estado mental, y no tiene nada que ver con el hecho de que sea objeto de otra creencia del mismo sujeto. Si tiene sentido suponer que pudiera haber un sujeto cuyas creencias fueran *todas* inconscientes —lo cual parece presuponerlo la teoría que estamos ahora examinando—, entonces ¿por qué no podría suceder que algunas de esas creencias fueran creencias acerca de otras creencias de ese mismo sujeto? Sin embargo, la teoría en cuestión lo excluye, puesto que, de acuerdo con ella, esas creencias acerca de las que el sujeto tiene creencias serían por esa misma razón consideradas como creencias *conscientes*.

En cualquier caso, al reflexionar se podría llegar a sospechar de la noción misma de creencia 'de orden superior', al menos en el sentido que ahora es pertinente. Para empezar, no parece que haya circunstancia alguna en la que podamos afirmar sensatamente 'creo que creo que *p*'. Supongámos que yo le pregunto a usted si usted cree que *p* o no. Usted podría responder: 'Sí, creo que *p*', o 'No, no lo creo', o incluso tal vez 'No estoy seguro'. Pero, ¿qué es lo que podría querer decir si respondiera 'Creo que creo que *p*'? Nótese, sin embargo, que no existe ninguna dificultad análoga para entender qué es lo que podría querer decir si dijera 'Creo que *sólo* cree que *p*', o 'Creo que *creeré* que *p*', comentando su propio estado mental pasado o el por venir. Similmente, no hay ninguna dificultad en entender qué es lo que usted querría decir si dijera 'Creo que *usted* cree que *p*', comentando el estado mental de otra persona presente.

Estas observaciones parecen relacionarse con lo que se conoce como la *paradoja de Moore*, es decir, el hecho de que, según parece, no tiene ningún sentido que alguien afirme 'Creo que *p* pero no es verdad que *p*', aunque tenga un perfecto sentido que alguien afirme '*Sólo* cree que *p* pero no es verdad que *p*' o '*Usted* cree que *p* pero no es verdad que *p*'.²⁴ Pudiera pensarse que de este hecho deberíamos extraer la lección de que aseverar 'Creo que *p*' no equivale a informar de que uno tiene una cierta creencia, sino más bien aseverar que *p*, si bien de algún modo provisional. Podríamos entonces entender la razón por la que resulta virtualmente contradictorio aseverar 'Creo que *p* pero no es verdad que *p*', toda vez que esto equivaldría tanto a aseverar que *p* (si bien provisionalmente) como a aseverar que no es verdad que *p*. Sin embargo,

²⁴ La paradoja de Moore (denominada así porque su descubridor fue G. E. Moore) y sus consecuencias para el concepto de creencia las comenta Jane Heal en su 'Moore's Paradox: A Wittgensteinian Approach', *Mind* 103 (1994), pp. 5-24. Véase también Arthur W. Collins, 'Moore's Paradox and Epistemic Risk', *Philosophical Quarterly* 46 (1996), pp. 308-19.

tanto si aceptamos este análisis como si no, la paradoja de Moore nos podría persuadir de que hemos de cuestionar el supuesto de que las propias creencias presentes sean cosas acerca de las cuales se pueda adoptar una posición cognitiva, es decir, que sean cosas acerca de las cuales se pueden tener creencias concurrentes. Esta duda puede verse reforzada por la siguiente consideración. Cuando se nos pregunta si creemos o no que *p*, no intentamos determinar la respuesta a esa pregunta pensando acerca de *uno mismo*, sino más bien pensando acerca de la proposición de que *p*. Si nos sentimos persuadidos de que existen buenos indicios en favor de la verdad de *p*, nos veremos inclinados a juzgar que es verdad que *p* y en consecuencia responderemos afirmativamente a la pregunta que se nos ha hecho. En la obtención de esa respuesta no parece estar implicado proceso alguno de introspección ni la formación de una creencia 'de orden superior'. Nos interesamos totalmente en la verdad o falsedad de la proposición de que *p* y de los indicios que puedan ser pertinentes para esa verdad o falsedad.

Pero si, como sugieren las anteriores consideraciones, saber qué es lo que uno está pensando o sintiendo no estriba en tener creencias verdaderas, formadas de forma fiable, acerca de lo que se está pensando o sintiendo, entonces ¿qué es? Tal vez sea simplemente cuestión de la *mánera* en que se está pensando y sintiendo, a saber, *conscientemente*. Esta propuesta hace aún menos misteriosa que la anterior la razón por la que habrán de ser nuestros pensamientos y sensaciones *conscientes* aquellos acerca de los cuales tengamos un conocimiento especialmente autorizado, ya que estamos sugiriendo ahora que el hecho de que un sujeto tenga 'conocimiento' de este tipo consiste simplemente en que sus pensamientos y sensaciones sean conscientes. Por supuesto, esto equivale a presuponer la noción de conciencia en lugar de explicarla, que es lo que intentaba hacer la teoría de los pensamientos de orden superior. Pero ya tenemos razones para pensar que la noción de conciencia puede ser tan fundamental que no pueda disponerse de ninguna explicación o ningún análisis de la misma que no sea circular.

EXTERNISMO Y AUTOCONOCIMIENTO

Este es un lugar conveniente para mencionar otro problema relativo al conocimiento de nuestros propios pensamientos conscientes. Se trata de un problema relacionado con el debate acerca de las explicaciones *externista* e *internista* del contenido mental que comentamos en el capítulo 4. El externista sostiene que los contenidos proposicionales de los pensamientos de alguien están, en general, determinados parcialmente por rasgos del entorno de ese alguien que existen independientemente de él. Por ejemplo, según el externismo, mi pensamiento de que el vaso que tengo ante mí contiene *agua* tiene el contenido que tiene debido en parte a que mi entorno físico contiene una determinada clase

de substancia, H_2O . Si mi entorno físico no hubiera contenido esa substancia, entonces, aunque hubiera contenido otro tipo de substancia XYZ, que tuviera la misma apariencia y el mismo sabor que el agua, yo no podría, de acuerdo con el externista, haber tenido un pensamiento con exactamente el mismo contenido que el que realmente tengo relativo a este vaso. Pero ahora el problema es éste. Parezco tener un conocimiento autorizado acerca de *qué* es lo que estoy pensando —el contenido proposicional de mi pensamiento— cuando tengo el pensamiento consciente de que este vaso contiene agua, y sin embargo puedo no tener ningún conocimiento de química y, en particular, ningún conocimiento de la diferencia entre H_2O y XYZ, dado que este último tipo de conocimiento tiene un carácter empírico y científico. En contraste con esto, mi conocimiento de *qué* es lo que estoy pensando parece no ser ni empírico ni científico, sino inmediato y *a priori*. Sin embargo, el conocimiento implica la verdad: si sé que estoy pensando que ese vaso contiene agua, se sigue que eso es realmente lo que estoy pensando, y si eso es realmente lo que estoy pensando, esto implica a su vez, según el externismo, que en mi entorno físico hay H_2O en lugar de XYZ. De ese modo parece que una verdad empírica, científica, resulta implicada por mi conocimiento no empírico de *qué* es lo que estoy pensando junto con la verdad de una doctrina filosófica, la doctrina del externismo. Pero esto es imposible, en cuyo caso, al parecer, deberíamos abandonar el externismo o bien la afirmación de que sabemos de un modo no empírico *qué* es lo que estamos pensando, es decir, abandonar la afirmación de que tenemos un conocimiento especialmente autorizado de los contenidos de nuestros propios pensamientos.

Sin embargo, quizás el problema se plantea aquí sólo para el modelo 'observacional' de las fuentes de nuestro autoconocimiento. Si se supone que saber lo que estamos pensando es una consecuencia de la inspección interna de nuestros propios estados mentales, entonces, el hecho —si es que es un hecho— de que los contenidos de esos estados estén parcialmente determinados por rasgos de nuestro entorno físico parece presentar una dificultad, ya que difícilmente podrá suponerse que los presuntos procesos de inspección interna revelen cosa alguna acerca de los rasgos del entorno. Por otra parte, si el hecho de que tengamos un pensamiento o una creencia consciente de que esto es un vaso de agua consistiera simplemente en creer que creemos que esto es un vaso de agua, como sostiene la teoría de los pensamientos de orden superior, entonces parece que no habría ningún problema, toda vez que el contenido de una tal creencia de segundo orden estará con seguridad determinado del mismo modo que el contenido de la creencia de primer orden que constituye su objeto. De modo que, si mi creencia de que esto es un vaso de agua tiene en parte determinado su contenido por el hecho de que mi entorno físico contiene H_2O , entonces también mi creencia de segundo orden de que yo creo que esto es un vaso de agua tendrá en parte determinado su contenido por el hecho de que mi en-

torno físico contiene H₂O; según parece, los contenidos de las dos creencias no pueden de ningún modo separarse. Ello parece suprimir la dificultad. Aunque mi creencia de segundo orden no sea una creencia que yo haya formado sobre bases empíricas o científicas, el que tenga el contenido que tiene implica, según el externismo, que mi entorno físico contiene H₂O, exactamente por la misma razón por la que lo implica el contenido de mi creencia de primer orden, y sin embargo, ello no parece comportar la nada bien recibida consecuencia de que pueda utilizar de algún modo mi conocimiento de qué es lo que creo para llegar a un conocimiento no empírico de la verdad científica en cuestión.²⁵

Sin embargo, si una solución así funciona en el caso de la teoría de los pensamientos de orden superior, parece que una solución similar también habría de funcionar para la teoría de que saber lo que estamos pensando es simplemente cuestión de la *manera* en que lo pensamos —a saber, conscientemente—, toda vez que también, según esta teoría, los contenidos de nuestros pensamientos conscientes pueden estar determinados por rasgos del entorno sin que esto parezca implicar que podemos de algún modo adquirir un conocimiento no empírico de cuáles son esos rasgos. Con todo, debo insistir en que las cuestiones que se han planteado en esta sección siguen siendo el tema de fuertes y complejos debates y que no ha surgido todavía un consenso sobre las mismas.

EL AUTOENGANZO

La otra cara de la moneda del autoconocimiento es el *autoengaño*, el cual constituye un fenómeno un tanto enigmático. Algunas personas parecen engañarse a sí mismas acerca de lo que realmente creen o desean. Por ejemplo, una madre podría rehusar enfrentarse al hecho de que su hijo, que era un soldado, ha muerto en acción, a pesar de haber recibido y leído las cartas oficiales que informan de su muerte. Ella quiere creer que todavía está vivo y en algún sentido *cree* que todavía está vivo, aunque en otro sentido está claro que debe saber que está muerto. Podríamos decir que tal persona 'se engaña a sí misma', pero es difícil interpre-

²⁵ Dos importantes artículos que intentan reconciliar el autoconocimiento con el externismo, en parte rechazando el modelo observacional del autoconocimiento, son los de Donald Davidson, 'Knowing Ones Own Mind', *Proceedings and Addresses of the American Philosophical Association* 60 (1987), pp. 441-58, y Tyler Burge, 'Individualism and Self-Knowledge', *Journal of Philosophy* 85 (1988), pp. 649-63, reimpressos ambos en Casam (ed.), *Self-Knowledge*. El tema se discute también en varios de los artículos de Crispin Wright, Barry C. Smith y Cynthia Macdonald (eds.), *Knowing our own Minds* (Oxford: Clarendon Press, 1998). Para comentarios extremadamente claros y útiles, véase John Heil, *The Nature of True Minds* (Cambridge: Cambridge University Press, 1992), capítulo 5. Una nueva e interesante solución al problema del autoconocimiento la propone André Gallois en su *The World Without, the Mind Within: An Essay on First-Person Authority* (Cambridge: Cambridge University Press, 1996).

tar de modo literal esta descripción de su situación. Una persona puede engañar a otra persona, pero sólo puede hacerlo porque sabe cosas que la segunda desconoce. En el caso de la madre de nuestro ejemplo, al parecer 'realmente' sabe que su hijo está muerto, de modo que no parece que tenga literalmente sentido describirla como alguien que guarda para sí mismo ese conocimiento, al modo en que uno puede guardar para uno mismo algo sin comunicárselo a otra persona. Tampoco resultará útil sugerir que parte de su mente sabe que su hijo está muerto e impide que este conocimiento lo tenga la otra parte de su mente, pues de hecho esto sería tratar el autoengaño como si fuera meramente un tipo de engaño de otros, lo que claramente no es. Aparte de cualquier otra consideración, parece haber algo de intrínsecamente irracional en el autoengaño, lo que no sucede con el engañar a otros. Por otra parte, no deberíamos confundir el autoengaño con el 'hacerse ilusiones', que también es irracional. Hacerse ilusiones consiste en creer algo sólo porque se quiere que sea verdad y no porque se tenga ningún indicio de que lo sea, y a la madre de nuestro ejemplo no parece aplicársele esa descripción, ya que posee buenos indicios en favor de la *falsedad* de lo que quiere creer y con todo, en algún sentido todavía lo cree frente a todos los indicios, aunque al mismo tiempo sepa 'realmente' que debe ser falso.

Es difícil decidir cómo se debería comenzar siquiera a describir adecuadamente un caso así.²⁶ Una enseñanza que podría extraerse de esto es que nuestro modo cotidiano de hablar acerca de 'creencias', 'conocimiento' y 'deseos' posee una complejidad que hasta el momento ha escapado a los intentos de los filósofos de proporcionarle un fundamento teórico claro. No necesariamente hemos de concluir, con los materialistas eliminativos, que los estados de actitud proposicional son una ficción nacida de una teoría de la mente 'de sentido común' que es científicamente inadecuada (véase de nuevo el capítulo 3). Pero tal vez deberíamos contemplar con mirada más escéptica el cómodo supuesto de que las creencias y los deseos claramente son 'estados' de personas, o de sus mentes, en analogía con estados físicos de sus cuerpos, como la forma, la masa y la velocidad.

CONCLUSIONES

En este capítulo final hemos cubierto una buena extensión de un territorio difícil, aunque en parte lo hayamos inspeccionado de un modo necesariamente superficial. Nuestras reflexiones sobre la naturaleza de nosotros mismos y del conocimiento de nosotros mismos han resultado

²⁶ Para más información sobre el autoengaño, véase David Pears, *Motivated Irrationality* (Oxford: Clarendon Press, 1984), capítulo 3, y Alfred R. Mele, 'Real Self-Deception', *Behavioral and Brain Sciences* 20 (1997), pp. 91-136. Este artículo contiene una extensa bibliografía.

ser perturbadoras en lugar de algo con lo que nos podamos encontrar a gusto y que resulte tranquilizador. Parecemos saber menos acerca de nosotros mismos de lo que podríamos haber imaginado al comienzo de nuestras indagaciones. Estamos ahora inseguros sobre qué clase de cosas somos y acerca de nuestras propias condiciones de persistencia. La memoria, en la que nos apoyamos para el conocimiento acerca de nuestro pasado, resulta ser un asunto complicado y controvertido, e incluso nuestra capacidad de saber lo que estamos pensando y sintiendo conscientemente parece ser algo que es difícil de entender. Con todo, estas incertidumbres no nos deberían conducir a desesperar de poder hacer progresos en la filosofía de la mente. Por el contrario, prueban cuán importante es la filosofía de la mente como correctivo de la complacencia que caracteriza a menudo tanto al pensamiento cotidiano acerca de la mente como a los supuestos teóricos de los psicólogos y los científicos cognitivos.

Bibliografía

- Anscombe, G.E.M., 'Causality and Determination', en Anscombe, *Metaphysics and the Philosophy of Mind. Intention*, segunda edición (Oxford: Blackwell, 1963) [Hay traducción al español, *Intención*, Barcelona: Paidós, 1991.]
Metaphysics and the Philosophy of Mind: Collected Philosophical Papers, vol. II (Oxford: Blackwell, 1981).
'The First Person', en Guttenplan (ed.), *Mind and Language*, reimpresso en Anscombe, *Metaphysics and the Philosophy of Mind*.
Armstrong, D. M., 'Perception, Sense Data and Causality', en G.F. Macdonald (ed.), *Perception and Identity* (Londres: Macmillan, 1979).
Universals: An Opinionated Introduction (Boulder, CO: Westview Press, 1989).
Aune, B., *Reason and Action* (Dordrecht: D. Reidel, 1977).
Austin, J.L., *Sense and Sensibilia*, ed. G.J. Warnock (Oxford: Clarendon Press, 1962).
Ayer, A.J., *Language, Truth and Logic*, segunda edición (Londres: Victor Gollancz, 1946) [Hay traducción al español, *Lenguaje, verdad y lógica*, Valencia: Universitat de València, 1991.]
The Foundations of Empirical Knowledge (Londres: Macmillan, 1940).
Baker, G. y Morris, K.J., *Descartes' Dualism* (Londres: Routledge, 1996).
Baker, L.R., *Saving Belief* (Princeton, NJ: Princeton University Press, 1987).
Barwise, J. y Perry, J., 'Scenes and Other Situations', *Journal of Philosophy* 78 (1981), pp. 369-97.
Situations and Attitudes (Cambridge, MA: MIT Press, 1983) [Hay traducción al español, *Situaciones y actitudes*, Madrid: Visor, 1992.]
Beauchamp, T.L. y Rosenberg, A., *Hume and the Problem of Causation* (Nueva York: Oxford University Press, 1981).
Bechtel, W. y Abrahamsen, A., *Connectionism and the Mind: An Introduction to Parallel Processing in Networks* (Oxford: Blackwell, 1991).
Benacerraf, P., 'Mathematical Truth', *Journal of Philosophy* 70 (1973), pp. 661-80, reimpresso en Benacerraf y Putnam (eds.), *Philosophy of Mathematics*.
Benacerraf, P. y Putnam, H. (eds.), *Philosophy of Mathematics: Selected Readings*, 2^a edición (Cambridge: Cambridge University Press, 1983).
Bermúdez, J.L., *The Paradox of Self-Consciousness* (Cambridge, MA: MIT Press, 1998).