

¿Consiste en algo ser una y la misma persona?

ENRIQUE ROMERALES(*)

Resumen: Aunque el problema de la identidad personal es antiguo, algunos de los enfoques actuales para resolverlo son novedosos; entre ellos destacan las llamadas teorías del nexo causal. Tales teorías suponen una modificación de la teoría psicologista de la identidad personal, que se remonta a Locke, y que durante bastante tiempo había pasado por ser la teoría estándar. En lo que sigue 1) examinaremos un par de versiones de la teoría causal de la identidad, aunque para comprenderlas mejor comenzaremos viendo en qué modifican la teoría psicologista tradicional, y haremos también alguna referencia a las teorías fisicalistas de la identidad personal; 2) exponemos la teoría de la identidad personal como algo básico e irreductible, como una alternativa a la anterior; 3) analizaremos algunos ejemplos diseñados para discriminar entre las teorías rivales; y 4) concluiremos afirmando que, dado que las teorías causales fallan, habrá o que tomar en serio la teoría no reduccionista o que buscar la respuesta en otro lugar.

Abstract: Although the problem of personal identity is old, some present-day approaches to solve it are new; among them stand out the so called causal link theories. Such theories are a modification of the psychological theory of personal identity, that goes back to Locke, and that for a rather long time had apparently been the standard theory. In what follows: 1) we will survey a couple of versions of the causal theory of identity, though for better understanding we will begin by regarding in what modifies it the traditional psychological theory, and will also make some reference to the physicalist theories of personal identity; 2) we will expound the theory of personal identity as something basic and non-reducible, as an alternative to the former; 3) we will analyze some examples designed to discern between rival theories; and 4) we will conclude claiming that, since causal theories fail, either non-reductionist theory should be seriously taken into account, or the answer should be looked for elsewhere.

La teoría psicologista de la identidad personal

¿Qué es lo que hace que Sócrates sea Sócrates y ningún otro? ¿Por qué yo sigo siendo yo a través del tiempo? ¿A qué se debe que una persona sea una y la misma persona por mucho que cambie? Para los defensores de la teoría psicologista de la identidad personal, ésta se basa en *determinadas* propiedades cualitativas, especialmente en la memoria y la autoconciencia. Así, según Locke, una persona es "un ser inteligente pensante, que posee razón y reflexión, y puede considerarse a sí mismo como sí mismo, la misma cosa pensante, en diferentes tiempos y lugares"¹. Según esta versión, para seguir siendo una misma persona deben permanecer algunas propiedades que se consideran esenciales, las demás pueden alterarse o desaparecer. Las propieda-

¹ *An Essay Concerning Human Understanding* II, XXVII, § 9, Oxford, Clarendon, 1990; p. 335.

(*) Dirección para correspondencia: Enrique Romerales Espinosa. Dep. de Filosofía, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid (Campus de Cantoblanco), 28049 Madrid (España).

© Copyright 1993 Secretariado de Publicaciones e Intercambio Científico, Universidad de Murcia. Murcia (España). ISSN: 1130-0507.

des esenciales son todas psíquicas (memoria y autoconciencia), de forma que si nos descorporeizáramos, o adquiriésemos un "cuerpo espiritual" (san Pablo) o, simplemente, nos transplantaran a otro cuerpo, seguiríamos siendo la misma persona, siempre y cuando conserváramos nuestros recuerdos y tuviéramos conciencia de seguir siendo el mismo.

Esta teoría da cabida a la profunda intuición de que nuestra memoria es algo fundamental, de que pocas cosas nos dolerían más que saber que nos iban a borrar todos nuestros recuerdos —incluido el de quiénes somos—. La memoria personal nos identifica a la vez que nos diferencia de los demás, igual que la (auto)conciencia de ser yo conlleva la de no ser otro.

Con todo, posee dificultades insuperables. La primera la puso claramente de manifiesto el obispo Samuel Butler en su *First Dissertation to the Analogy of Religion* de 1736: "la conciencia de la identidad personal presupone, y por ende no puede constituir, la identidad personal, más de lo que el conocimiento, en cualquier otro caso, puede constituir la verdad, la cual presupone"². Como afirma Butler, la circularidad de pretender fundamentar la identidad personal sobre la conciencia de dicha identidad personal es palmaria.

Otra objeción crucial es que la identidad es una relación que posee las propiedades de reflexividad, simetría y transitividad. Pero la continuidad —o mejor **conectividad**— de conciencia y memoria no es transitiva. En efecto, según Locke yo soy la misma persona a lo largo del tiempo si cada una de mis fases está psicológicamente conectada con la anterior, de forma que conservo los recuerdos de la fase anterior (o buena parte de ellos) y la conciencia de seguir siendo el mismo. Pero puede ocurrir que hoy esté psicológicamente conectado con mi yo de ayer, y ayer con el de anteayer, y anteayer con el del día anterior, y así sucesivamente, sin que ello garantice que hoy esté psicológicamente conectado con mi yo de hace diez o veinte años. Esta relación de conectividad entre fases sucesivas de mi historia temporal no es transitiva, y por tanto no puede constituir la identidad personal que ha de serlo.

Una tercera dificultad no menos seria la constituyen la amnesia y la paramnesia³. Según Locke, si la persona *P* es el mismo en t_2 que en t_1 , entonces puede recordar (todo o al menos algo de) lo que hizo en t_1 . Y viceversa, si *P* en t_2 puede recordar haber hecho o vivido *x* en t_1 , es porque *fue él* quien hizo o vivió *x* en t_1 . Ahora bien, puede muy bien ocurrir que yo haya hecho algo que sea incapaz de recordar —incluso aunque sea algo tan significativo como un asesinato—. Cabría replicar que el 'puede' es lógico: si has hecho *x* *por definición* puedes recordar haber hecho *x*. Pero entonces, como advierte Flew, el enunciado se convierte en una tautología vacía. Además, puede ocurrir que yo "recuerde" haber hecho o vivido cosas que no he hecho ni vivido (paramnesia). ¿Cómo saber entonces si mi recuerdo de haber hecho *x* *demuestra* que fui yo quien hizo *x*? De nuevo cabe apelar a las definiciones, y decir p.e. que si Juan recuerda haber hecho algo que no hizo, lo que tiene no es un recuerdo, sino un "recuerdo". Pero este expediente tampoco aquí sirve de nada, pues la cuestión es cómo diferenciar los recuerdos —constitutivos de la identidad personal— de los "recuerdos" —no constitutivos—. En suma, el hecho de que yo puedo en un momento dado no acordarme de nada de lo que he hecho o vivido, incluso no acordarme de quién soy, sin que por ello deje de ser quien soy; junto con el hecho —menos frecuente— de que puedo recordar hechos o vivencias que en realidad nunca experimenté, incluso que nunca le han ocurrido a nadie, sin que por ello sea yo el autor de tales hechos o el sujeto de dichas vivencias, muestra que algo falla en la teoría de Locke.

² Primer apéndice: "Of Personal Identity" [en John Perry (ed.): *Personal Identity*, Berkeley, Univ. California Press, 1975; p. 100].

³ Cf. A. Flew: *The Logic of Mortality*, Oxford, Blackwell, 1987; pp. 136-42.

La teoría psicologista de la identidad personal produce, además, numerosos contraejemplos. Por ejemplo el siguiente: supongamos que un neurólogo maniaco me atrapa y me amenaza con urgar en mi cerebro para causarme directamente durante un buen rato el máximo de dolor, y que, por si fuera poco, antes me va a borrar la memoria y me va a implantar la memoria de otra persona —digamos Stalin— así como su personalidad, aunque guardará escrupulosamente todos mis recuerdos para luego reimplantármelos a la vez que borrará del cerebro los recuerdos y la personalidad de Stalin. ¿Me quedaré contento pensando "le va a dar una buena lección a ese Stalin" como se deduce de la teoría lockeana?⁴ Más bien parece que soy yo quien va a sufrir terribles dolores, incluso aunque no me pare a pensar quién está sufriendo o aunque crea que soy Stalin. Según Williams —y parece estar en lo cierto— efectivamente tenemos razones para temer tal experimento (aparte de los problemas técnicos) que van contra la teoría psicológica de la identidad personal. Como alternativa él ofrece una teoría fisicalista de la identidad personal según la cual es la identidad de cerebro (o de la mayor parte suya) lo que proporciona la identidad personal.

La teoría materialista de la identidad personal

Para los reduccionismos fisicalistas la identidad de una persona consiste en la mismidad —continuidad— de su cuerpo y/o su cerebro. La persona adquiere su identidad gracias a la unidad corporal, y en esto se asemeja a los demás objetos, cuya identidad viene asimismo dada por la continuidad y contigüidad espacio temporales. Como en las personas pueden sustituirse varias partes del cuerpo sin que esto afecte a su identidad, se considera que lo definitivo para la identidad es el cerebro o, al menos, la mayor parte (o la parte más esencial) de él.

Los fisicalistas insisten en que los casos *reales* de identidad personal implican siempre continuidad y contigüidad cerebrales y en que su teoría se amolda al uso ordinario de los conceptos y al método de reconocimiento: nuestro modo de identificar a otras personas se basa siempre en criterios materiales, sean puramente físicos o biológicos. Esto último es simplemente falso, además de irrelevante. Falso, porque a menudo para identificar a una persona utilizamos criterios psicológicos, como averiguar sus creencias, intenciones, deseos etc., es decir su personalidad, y sobre todo preguntamos por sus recuerdos para constatar si su memoria corresponde con la de la persona que supuestamente es. Irrelevante, porque una cosa es la identificación, esto es, *cómo podemos saber si una persona es tal persona*, y otra la identidad, o sea, *si una persona es o no tal persona* (qué razón o fundamento hay para que esa persona sea la persona que es independientemente de que tales razones puedan o no ser verificadas por nosotros)⁵.

Pero hay otras dificultades que han reducido notablemente el número de fisicalistas. Una célebre, de corte cartesiano, es puesta de manifiesto por Sydney Shoemaker: mientras que la identificación de las demás personas se realiza usando descripciones físicas o descripciones de

⁴ El ejemplo, con variaciones, aparece en B. Williams: «The Self and the Future», *Philosophical Review*, 79, (1970). Ahora en su *Problems of the Self*, Cambridge Uni. Press, 1973; p. 48 y ss.

⁵ Por ejemplo, supongamos que se me presenta en una pantalla la escena de una guardería infantil y se me dice: "ahí está jugando Cantor de bebé". Obviamente yo no puedo *reconocer* a Cantor, pues no poseo ningún criterio de identificación —no sé cómo era de pequeño—. Pero eso es *irrelevante* para su identidad: Cantor está ahí o no está, independientemente de que alguien pueda reconocerlo: es aquel que andando el tiempo demostrará p.e. la correspondencia entre el conjunto de los números reales y los puntos de una línea.

modos de comportamiento que siempre hacen referencia a sus cuerpos, en los enunciados acerca de nuestra propia identidad es del todo diferente; 'yo' no se usa nunca como una abreviatura de la descripción física de mí mismo (la referencia a uno mismo no es la referencia al cuerpo de uno) y, además, uno sabe la *verdad* de los enunciados psicológicos en primera persona sin saber nada acerca de su cuerpo. Pero el sujeto de todos estos enunciados es la palabra 'yo', que se refiere necesariamente a una persona. Parece, pues, que puedo formular enunciados acerca de una persona que no se refieren a ningún cuerpo y cuya verdad puedo conocer sin referirme a ningún cuerpo. Pero esto mismo pueden hacerlo los demás; "si de cada persona puede haber un conocimiento (el suyo) que no es un conocimiento de un objeto material, parece seguirse que las personas no son, o al menos no esencialmente, objetos materiales"⁶.

La crítica más general es que tales teorías son innecesariamente restrictivas al prohibir como incoherente o lógicamente imposible lo que normalmente se considera simplemente no real. Así, tanto un Shoemaker (teórico de la teoría causal de la identidad personal) como un Swinburne (teórico de la teoría de la identidad personal como algo último e irreductible) convienen en que cosas tales como los intercambios de cuerpo —cerebro incluido—, la existencia de personas sin cuerpo, la supervivencia tras la muerte del cuerpo etc. son concebibles y posibles⁷. Y es que para un fisicalista la transmigración de personas o la reencarnación no es simplemente algo que de hecho no sucede, sino algo incoherente, debido a que según su definición de 'persona', dado que una persona es un cuerpo vivo determinado con tales y cuales características, si ese cuerpo se destruye y se sustituye por otro *ex hypothesi* hay otra persona. Pero los conceptos de transmigración y reencarnación *son claramente coherentes*, luego lo que falla es tal definición.

Alguno podrá, empero, dudar de que poseamos p.e. criterios de identidad para las personas sin cuerpo⁸. Pero tanto si las personas están constituidas por la continuidad y conectividad psicológicas, como si son una entidad última e inanalizable, en uno u otro caso eso constituye su identidad sin tener que recurrir a un cuerpo. Más aún, contamos incluso con *criterios de identificación* para personas sin cuerpo. P.e. yo puedo reconocer a un espíritu por su voz, por el contenido de sus mensajes, o por sus recuerdos de anteriores encuentros. Supongamos que alguien objeta: ¿qué sentido tiene que un espíritu afirme p.e. que ayer estuvo presente en la boda de tu primo, si el que ni tu primo, ni tú, ni nadie le viese no refuta tal aserción? Bien, siempre puedo preguntarle: "¿cómo era el vestido de la novia?", "¿quién llevaba la cola del vestido?" etc. Y si contesta correctamente tendré razones para pensar que realmente estuvo presente en la ceremonia.

El fisicalista puede replicar que de lo que se trata es de *nuestro concepto de persona*, esto es, del que se emplea en el lenguaje ordinario, y no de un constructo filosófico. Y que en el

⁶ S. Shoemaker: *Self-Knowledge and Self-Identity*, Ithaca, Cornell Uni. Press, 1963; p. 16-17. No obstante, Shoemaker advierte que no hay que sacar conclusiones cartesianas al respecto: de ahí no se sigue ni que la persona sea algo enteramente no material, ni que la identidad personal sea esencialmente diferente de la identidad de los demás objetos físicos (p. 19-20). Siempre cabe afirmar que la persona ni es esencialmente material *solamente* ni esencialmente inmaterial *solamente*, sino que incluye esencialmente ambos aspectos. Esta, en su versión lingüística, es la tesis de Strawson: "La condición [para considerarlos personas]... es que los individuos en cuestión, incluido uno mismo, sean de un cierto tipo único: a saber, de un tipo tal que a cada individuo de ese tipo le han de ser adscritos, o adscribibles, tanto estados de conciencia como características corpóreas"; *Individuos*, Madrid, Taurus, 1989; p. 106-7.

⁷ S. Shoemaker & R. Swinburne: *Personal Identity*, Oxford, Blackwell, 1984.

⁸ Una excelente argumentación a favor de la coherencia del concepto de persona sin cuerpo (espíritu) aparece en R. Swinburne: *The Coherence of Theism*, Oxford, Clarendon, 1977; pp. 99-102 y 105-125. También en *The Evolution of the Soul*, Oxford, Clarendon, 1986; p. 152 y ss.

lenguaje ordinario 'persona' se refiere siempre a un ser humano, y por consiguiente a un animal necesariamente corporal⁹. Esto es cierto, pero de nuevo irrelevante. Porque lo que quiere el filósofo no es simplemente averiguar cómo se usan los conceptos en el lenguaje ordinario —que puede equivocarse tan a menudo como el "sentido común"— sino construir conceptos *últimamente adecuados a su objeto*. Y lo que está en cuestión aquí es la cosa misma, la identidad personal, algo sobre cuya naturaleza ningún uso lingüístico, por bien establecido que esté, puede legislar.

Las teorías del nexa causal

Visto que las teorías materialistas de la identidad personal son inaceptables por demasiado restrictivas, y que la teoría psicologista adolecía de graves defectos, inspeccionemos cómo las teorías causales de la identidad superan estos defectos. En un artículo ya clásico Sydney Shoemaker abordó —y hay consenso en que con éxito— el problema de la circularidad de la teoría psicologista, a la vez que el problema de separar los falsos recuerdos de los genuinos¹⁰. Dado que el recuerdo de haber experimentado o realizado algo lógicamente sólo puede ser de algo previamente experimentado o ejecutado *por uno mismo*, es cierto que la memoria presupone la identidad personal. Shoemaker define entonces "cuasi-recuerdo" como aquel recuerdo tan semejante a un recuerdo como sea posible, con la salvedad de que la experiencia puede haber sido experimentada, o la acción ejecutada, por otra persona. En nuestro mundo seguramente todos los cuasi-recuerdos son recuerdos. Pero supongamos un mundo posible en el que la memoria fuera portada en las células de la sangre. Entonces, mediante una transfusión de otro, yo podría cuasi-recordar haber hecho o vivido cosas que no he hecho o vivido yo, sino él. Esto muestra que los cuasi-recuerdos no presuponen la identidad personal (solucionada, pues, la objeción de Butler). La cuestión clave ahora es cómo diferenciar aquellos cuasi-recuerdos que son genuinos recuerdos de aquellos que no lo son. Para que un cuasi-recuerdo de alguien sea un genuino recuerdo hace falta, en primer lugar, "que su presente recuerdo sea de algún modo *debido a*, que acontezca *a causa de*, un estado cognitivo y sensorial que quien recuerda tuvo en el momento en que experimentó el acontecimiento"¹¹. Shoemaker denomina a tal cuasi-recuerdo causalmente conectado con su origen 'cuasi-recuerdo_c' o 'recuerdo_d'¹² ('d' por *en sentido débil*). La conexión causal *apropiada* es, pues, garantía de que el recuerdo ha sido producido por haber vivido la experiencia en cuestión. Dado este requisito de conexión causal, sólo una cosa puede frustrar la identidad entre quien casi-recuerda_c haber hecho o vivido algo y quien de veras lo experimentó: que haya habido alguna fisión o fusión de personas (o transfusión de sus recuerdos). Así, si una persona *P* se divide en dos iguales *P*₁ y *P*₂ en un tiempo *t*, tras *t* tanto *P*₁ como *P*₂ cuasi-recuerdan_c las vivencias y acciones de *P*, aunque obviamente no pueden ser ambos *P*. El segundo requisito es, entonces, que no haya tales ramificaciones: "Podemos aseverar como una verdad lógica que si *S* recuerda_d el acontecimiento *E* (o recuerda_d la acción *A* desde el inte-

⁹ David Wiggins propone que el concepto de persona se construya añadiendo a la definición de Locke que una persona es siempre una especie de animal con determinadas características (*Sameness and Substance*, Oxford, Blackwell, 1980; p. 188).

¹⁰ «Persons and their Pasts», *American Philosophical Quarterly*, 7, (1970); pp. 269-285.

¹¹ «Persons and their Pasts»; loc. cit. p. 272.

¹² «Persons and their Pasts»; p. 278 y 281.

rior), y si no ha habido ramificación de las cadenas causales [apropiadas] durante los tramos relevantes de la historia de *S*, entonces *S* es uno de los testigos de *E* (o la persona que hizo *A*)¹³. La respuesta de Shoemaker constituye un buen paso adelante; pero como veremos, el *quid* de la cuestión está en cuál haya de ser el tipo de conexión causal.

Según la teoría causal, pues, la identidad de una persona a través del tiempo se basa, además de en la semejanza psíquica que proporciona la conciencia y la memoria, en la *relación causal* con sus estados (mentales) antecedentes¹⁴. La teoría psicologista clásica consideraba a la persona como un conjunto de fases o "escenas" distribuidas a lo largo del tiempo y psicológicamente conectadas cada una con la siguiente (**conectividad**). Pero vimos cómo esto no lograba producir la necesaria transitividad: yo puedo acordarme a los veinte años de lo que hice a los diez, y a los treinta de cómo era a los veinte, pero ser incapaz de recordar a los treinta mi vida a los diez. Para conseguir la transitividad hace falta *además* la **continuidad psicológica**: "el mantenimiento de cadenas superpuestas de conectividad fuerte"¹⁵. Esto quiere decir que haya varias líneas de continuidad superpuestas entre varios grupos de fases, como los hilos de una soga, de forma que cualquier fase esté bien conectada con cualquier otra, aunque no haya un mismo hilo que vaya de la primera a la última fase. De este modo, si en t_2 Marta tiene lazos de continuidad con t_1 y en t_3 con t_2 , los tiene en t_3 con t_1 .

La continuidad resuelve también el problema de la amnesia. Que yo no pueda recordar ahora, supongamos en t_{13} , lo que hice p.e. en t_4 no es problema, siempre y cuando haya al menos una fase anterior psicológicamente continua con la presente en la que yo sí podía recordar lo que me ocurrió en t_4 . Supongamos que hasta t_9 yo podía recordar a voluntad lo ocurrido en t_4 ; pero t_9 —como cualquier otra fase— es psicológicamente continua con t_{13} —la fase actual; luego yo soy el mismo que era en t_4 .

Queda el problema de la paramnesia, y aquí está la clave de estas teorías: *la continuidad psicológica ha de deberse al género correcto de causa*. ¿Qué quiere decir esto? En principio algo simple. Supongamos que yo recuerdo haber sido Napoleón; tengo ante mi mente con gran vivacidad Waterloo, la campaña de Rusia, el aburrimiento en Santa Elena y muchas otras cosas y, especialmente, todas estas vivencias las veo "desde dentro". Según Locke eso basta para demostrar que yo soy el mismo que Napoleón. Ahora bien, supongamos que soy un estudioso y obseso del tema, y me he "empapado" hasta tal punto de la biografía de Napoleón que he acabado por identificarme con él reviviendo sus hazañas como propias. ¿Cómo saber entonces si Napoleón y yo somos o no la misma persona? La respuesta es: si todos estos recuerdos-vivencias han sido causados en mi mente por haber estado efectivamente en Waterloo, Rusia o Santa Elena hace unos 200 años, entonces yo soy Napoleón. Si se trata meramente de una identificación psicológica producida por la obsesión, no lo soy. De esta forma, la *conexión causal* entre mis recuerdos y las *experiencias realmente vividas* por mí es lo que garantiza la mismidad de la persona. Pero esto parece circular: preguntar si mi recuerdo de haber asesinado a Kennedy ha sido producido por mi experiencia de asesinarlo es sencillamente preguntar si el asesino de Kennedy

¹³ Shoemaker; op. cit. p. 281.

¹⁴ "Sólo allí donde suponemos que los rasgos de las cosas en momentos diferentes están causalmente relacionados de este modo estamos en nuestro derecho de considerar la similitud de algo en un momento y algo en otro momento como evidencia de la identidad"; Shoemaker, op. cit. p. 282-3.

¹⁵ D. Parfit: *Reasons and Persons*, Oxford, Clarendon, 1984; p. 206. 'Fuerte' se refiere a que como la conectividad admite grado hemos de estipular un grado suficiente para salvaguardar la identidad. Parfit expuso algunas de sus ideas anteriormente en su «Personal Identity», *The Philosophical Review*, 80 (1971).

y yo somos la misma persona, y no hay modo de contestar afirmativamente a lo primero sin presuponer lo segundo. Si esto es así, ¿qué fundamenta qué: la causalidad a la identidad o al revés? El problema es que la identidad y la causalidad están interconectadas, de modo que cuando la una es borrosa, la otra es dudosa.

Para obviar esta objeción, algunos admiten que la conexión causal no tiene por qué ser la normal, sino que puede ser cualquier causa fiable, o incluso cualquiera en absoluto¹⁶. Según esto, si mi recuerdo se produce porque me han manipulado convenientemente el cerebro, yo soy el asesino. ¡Pero esto es algo antiintuitivo! ¿Cómo voy a ser yo el asesino de Kennedy porque alguien manipule mi cerebro y me implante los recuerdos del asesino y su conciencia (de culpabilidad)? O consideremos, por ejemplo, el caso de la transfusión sanguínea portadora (parcial) de la memoria ideado por Shoemaker. Para Parfit tal caso sería un ejemplo de identidad personal borrosa (tanto más borrosa cuanto más se aproxime la memoria transferida al 50% del total): al llevarme parte de la memoria de otro mediante una transfusión me llevo parte de su identidad personal, luego la cuestión de si fui *yo* quien hizo o experimentó lo que *ahora* recuerdo haber hecho o experimentado (pero antes de la transfusión *no podía* recordar) no tiene una respuesta definida, dado que parte de la otra persona se ha transvasado a mí. Para Shoemaker, empero, en un caso así la conexión causal sería obviamente inadecuada para caracterizarme a mí como autor de los hechos o sujeto de las vivencias del donante de sangre, pues tal conexión causal no tiene nada que ver con las conexiones causales ordinarias. Intuitivamente parece que Shoemaker lleva razón y que si, p.e., el donante de sangre hubiera cometido un asesinato antes de la transfusión, habría que castigarlo "a él" y no "a mí", aunque tras la transfusión sólo yo pudiera recordar el asesinato e incluso sentirme culpable por ello. Pero lo verdaderamente difícil para esta teoría es determinar qué conexiones causales son suficientemente parecidas a las ordinarias como para ser calificadas de portadoras de la identidad personal —o incluso justificar que sólo en las relaciones causales ordinarias se conserva la identidad personal—. La posición de Parfit (más radical para ganar en coherencia) obvia este problema admitiendo *cualquier* conexión causal, ordinaria o no.

Sin embargo, como han señalado Kolak y Martin¹⁷, lo decisivo es que la conexión causal admite grado, y si estamos dispuestos a ampliar al máximo la laxitud de la conexión causal para considerar el mayor número de casos como casos de identidad personal, como hace Parfit, el nexo causal acaba por disolverse, con lo que volvemos a la teoría psicologista estándar. Pero si queremos detenernos en un punto y fijar un grado mínimo de conexión causal como condición necesaria de la identidad personal, no hay criterio alguno para fijar ese punto más bien que algún otro.

En este punto Parfit reconoce que la identidad personal no equivale a la conectividad más continuidad psicológicas, como muestra el caso de la fisión. Volvamos al caso en que una persona *P* se divide cual ameba en un momento dado en dos personas cualitativamente idénticas *P'* y *P''*. En el instante siguiente tanto *P'* como *P''* poseen conectividad y continuidad psicológica con *P*: ambos recuerdan todo lo que *P* y ambos tienen conciencia de ser *P*. Evidentemente,

¹⁶ Esta última, denominada la concepción más amplia del nexo causal, es seguramente la más coherente y la defendida por Parfit, (*Reasons and Persons*; p. 207 y ss.). Frente a ella, David Lewis («Survival and Identity», en A. Rorty (ed.): *The Identities of Persons*, Berkeley, Uni. of California Press, 1976; p. 17) defiende que la conexión causal ha de ser la normal; y esa parece ser también la opinión de Shoemaker.

¹⁷ D. Kolak y R. Martin: «Personal Identity and Causality: Becoming Unglued», *American Philosophical Quarterly*, 24 (1987); p. 343-344.

por el principio de necesidad de la identidad¹⁸, no pueden ser ambos *P*. Al menos uno de ellos — digamos *P'* — no es *P* (aunque seguramente ninguno lo es) a pesar de poseer conectividad y continuidad con *P*: luego la identidad personal no equivale a tales propiedades. A la relación que se da entre *P'* y *P* la denomina Parfit *supervivencia*. *P'* posee todos los rasgos físicos y psíquicos de *P*, toda su memoria y la conciencia de ser *P* ¿qué más puede faltar? Nada, según Parfit. Ello muestra que *la supervivencia es lo que importa*. A los demás efectos, la identidad personal es del todo irrelevante.

Que la identidad personal y la supervivencia no son lo mismo puede concedérsele a Parfit¹⁹. Como él bien señala, la identidad personal exige una relación biunívoca: un individuo sólo puede ser idéntico a un individuo; mientras que varios individuos pueden ser los *supervivientes* de un individuo (en la fisión son dos, pero podrían igualmente ser 37). Además, la identidad es una relación que no admite grado: o dos cosas son idénticas o no lo son; mientras que la supervivencia lo admite: si *P''* recuerda dos o tres cosas menos que *P* (o si tiene una cana que *P* no tenía) *P''* es un superviviente menos claro de *P* que *P'*.

Que la supervivencia sea lo único que importa es lo que hay que discutir. Supongamos que un genio maligno me advierte que me va a desintegrar y que acto seguido va a integrar en el mismo punto del espacio-tiempo a dos personas cualitativamente idénticas a mí en lo físico y en lo psíquico. Como no existe una relación biunívoca entre esos dos y yo, ellos no serán la misma persona que yo. Como son idénticos, ninguno es mejor candidato a la identidad conmigo que el otro, luego presumiblemente ninguno de ellos soy yo. Al instante siguiente, antes de que ninguno de ellos pueda apercibirse de la existencia del otro ni de lo que está ocurriendo, el genio desintegrará a uno. De esta forma el que quede —el superviviente— no se habrá apercibido de nada, seguirá siendo cualitativamente idéntico a mí, teniendo la conciencia de ser yo, y los mismos proyectos, deseos, intenciones etc. ¿Me dejará indiferente la amenaza del genio? Tiendo a pensar que no. Según Parfit, la supervivencia es lo único importante, y la supervivencia en grado sumo está garantizada en este caso. Pero ¿lo único importante para quién? Para los demás, para todo el mundo, mi supervivencia es lo único importante (si es que tiene alguna importancia en absoluto); pero para mí es crucial la identidad. Según todas las apariencias *es otro quien en adelante va a vivir por mí* y, lo que es peor, *creyendo ser yo*. Parfit, Lewis, Perry y otros insisten en que esta "sensación" carece de fundamento, aunque está muy arraigada. Perry postula que probablemente ha sido conservada como factor de supervivencia en la evolución biológica, porque desde un punto de vista lógico si me sobrevive alguien que puede llevar a cabo todas mis

¹⁸ $\forall x \forall y [(x=y) \rightarrow \Box (x=y)]$; o sea, si dos cosas son idénticas, son necesariamente idénticas —o dicho en términos de mundos posibles, si dos cosas son una en un mundo posible, son una en todos los mundos posibles—. Este principio se sigue de la ley de substitutividad de la identidad o ley de Leibniz: $\forall x \forall y [(x=y) \rightarrow (Fx \rightarrow Fy)]$, junto con la necesidad de la autoidentidad: $\forall x \Box (x=x)$. En «Identity and Necessity» (en M.K. Munitz (ed.): *Identity and Individuation*, New York Uni. Press, 1971; p. 135-64) Saul Kripke ha defendido la necesidad (*de re*) de la identidad contra varias objeciones de corte lógico-lingüístico y epistemológico. Aunque la necesidad de la identidad es generalmente admitida, su aceptación está lejos de ser unánime. Para una defensa de la identidad contingente cf. Allan Gibbard: «Contingent identity», *Journal of Philosophical Logic*, 4 (1975); p. 187-221.

¹⁹ Lewis defiende, en cambio, que son lo mismo. Pero él construye una nueva noción de identidad: la *identidad temporizada* (tensed identity), según la cual dos personas pueden ser idénticas en t_1 y diferentes en t_2 . Entonces, en los casos de fisión de personas lo que ocurre es que tanto antes como después de la fisión hay dos personas: dos personas que son t-idénticas primero y no t-idénticas después. En nuestro caso, *P'* sería un superviviente de *P* y sería idéntico con *P* en t_1 y no idéntico en t_2 ; lo mismo ocurriría con *P''* respecto de *P*. Cf. su «Survival and Identity», en A. Rorty: *The Identities of Persons*, Berkeley, California Uni. Press, 1976; p. 17-40.

tareas y proyectos *exactamente igual* que yo, no tendría por qué preocuparme²⁰. Pero esto último parece claramente inaceptable.

Para la versión reduccionista fuerte de la identidad personal, como la que defienden Parfit, Lewis y Perry, podemos asimismo explicar la unidad de la vida mental de modo impersonal; la referencia lingüística a un sujeto es siempre posible, pero no necesaria: "Podríamos redescubrir la vida de cualquier persona en términos impersonales. Al explicar la unidad de esta vida, no tenemos por qué afirmar que es la vida de una persona particular. Podríamos describir lo que, en diferentes ocasiones, era pensado, sentido, observado y hecho, y cómo estaban interrelacionados estos diversos acontecimientos. Las personas serían mencionadas aquí sólo en las descripciones del contenido de muchos pensamientos, deseos, recuerdos, y demás. No tiene por qué afirmarse que las personas sean los pensadores de ninguno de estos pensamientos"²¹.

Otros más moderados, como Shoemaker o Strawson (éste siguiendo a Kant) piensan que no podemos referirnos a experiencias particulares sin referencia a su sujeto, esto es, que no podemos dar una descripción completa del mundo en términos puramente impersonales²². No obstante, para todos ellos el sujeto no existe independientemente de su cuerpo-cerebro y de su memoria-conciencia. Del mismo modo, piensan, que no hay estados físicos o mentales sin sujeto, no hay sujeto sin estados físicos o mentales²³. Hirsch ha destacado que Parfit insiste en que no sólo podemos reemplazar nuestro modo de hablar ordinario en términos de identidad personal y del yo por un modo impersonal en términos de etapas o fases de personas interconectadas por una relación (más o menos fuerte de conectividad y continuidad), sino que este segundo modo de hablar es preferible. Ahora bien, según Hirsch, aunque la identidad personal no se base en un hecho último metafísico o lógicamente necesario, hay una *necesidad psicológica* en los seres humanos de emplear el concepto de yo y su lenguaje correspondiente tal y como lo usamos²⁴. Esto haría que, si bien en casos especiales anormales, como los escrutados por Parfit, su lenguaje sería factible o quizá preferible, "es psicológicamente imposible para un ser humano tener una actitud parfitiana como su orientación primaria. Si uno posee un sentido normal del yo, entonces sus juicios espontáneos de preocupación²⁵ deben, al menos típicamente, ser acuñados en términos del concepto ordinario de auto-identidad..."²⁶.

²⁰ «The Importance of Being Identical», en A. Rorty (ed.): *The Identities of Persons*; p. 83.

²¹ *Reasons and Persons*; p. 251. Esto recuerda a la tesis de Hume de que "la mente es una especie de teatro en el que distintas percepciones se presentan en forma sucesiva... No existe en ella con propiedad ni *simplicidad* en un tiempo, ni *identidad* a lo largo de momentos diferentes, sea cual sea la inclinación natural que nos lleve a imaginar esa simplicidad e identidad" (Hume: *Tratado de la naturaleza humana*, I, Madrid, Editora Nacional, 1977; p. 401). Y de ahí hay sólo un pequeño paso a aceptar una teoría negadora de la identidad personal que afirme *simpliciter* que la identidad personal o el yo es totalmente irreal, como se hace en algunas versiones del budismo. Pero, mientras que afirmar que el sujeto es siempre *el sujeto de* (la experiencia, la acción, el lenguaje...) es coherente, afirmar que el sujeto *no es*, esto es, que puede haber experiencia, acción o lenguaje *sin sujeto*, me parece claramente incoherente.

²² Cf. P.F. Strawson: *Los límites del sentido*, Madrid, Revista de Occidente; p. 145 ss.; S. Shoemaker: *Self-Knowledge and Self-Identity*; p. 99 ss.; y Eli Hirsch: *The Concept of Identity*, New York, Oxford Uni. Press, 1982; p. 307 ss. En efecto, parece que la primera persona del singular, la autorreferencia (sea con el 'yo' explícitamente delante o sin él) es una especie de universal lingüístico, lo que da pie a considerarlo una suerte de "trascendental" para el pensamiento.

²³ Algunos como Shoemaker admiten que puede haber sujeto con sólo estados mentales; vimos que otros como Strawson requieren que haya de ambos tipos para poder hablar de 'personas'.

²⁴ Eli Hirsch: *The Concept of Identity*; p. 302 ss.

²⁵ También Perry sostiene que nos preocuparemos de lo que le pase a una persona en el futuro en la medida en que nos identifiquemos con esa persona. Cuando la identificación es total —allí donde se da una mismidad de perspectiva interior— la preocupación es máxima («The Importance of Being Identical», en A. Rorty, op. cit. p. 73 y ss.).

²⁶ Eli Hirsch: *The Concept of Identity*; p. 311. Con todo, Hirsch conviene con Parfit en que su análisis de la identidad

Hay una versión de la teoría causal que, manteniendo el interés por la identidad, reconoce que en la fisión no puede conservarse la identidad personal. Nos referimos a la teoría del continuador más próximo (en adelante TCMP) de Robert Nozick. Según él, la conectividad y continuidad causales entre dos entidades a lo largo del tiempo no bastan para la identidad, porque puede haber más de una entidad que tenga los mismos lazos causales con la entidad precedente. Es necesario que la relación de semejanza y causalidad sea más fuerte entre dos entidades que entre la primera y cualquier otra para que podamos decir que son la misma entidad, que la segunda es la continuadora causal de la primera²⁷. Así, p.e. si tras la segunda Guerra Mundial se reúnen en Estambul 3 o 4 miembros del Círculo de Viena y siguen trabajando en lo mismo con la misma orientación, ellos son ahora el Círculo de Viena. Pero si resulta que al mismo tiempo en Chicago se encuentran trabajando de igual forma 7 u 8 de sus miembros, son ellos, y no los de Estambul, el verdadero Círculo de Viena. Esto quiere decir que la identidad de algo *no es independiente* de la identidad de todo lo demás. Aplicándolo a personas supongamos el siguiente caso. Fettuccini, que es un gran prestidigitador, entra en una caja fuerte y ante la vista del público se cierra la caja. Al minuto abren la caja y Fettuccini no está. Acto seguido abren otra caja que está cinco metros a la izquierda y ahí aparece una persona física y psíquicamente indistinguible de Fettuccini. Aunque nadie pueda explicar cómo ha ocurrido, todos acaban por considerar que el de la caja de la izquierda es Fettuccini. Según la TCMP efectivamente lo es. Pero acto seguido empiezan a oírse ruidos en una caja que está cinco metros a la derecha de la primera y al abrirla, ante la sorpresa de todos, hay una persona también indiscernible tanto física como psíquicamente de Fettuccini. Entonces según la TCMP el otro ya no es Fettuccini: al haber dos candidatos igual de bien relacionados con Fettuccini en cuanto a la semejanza y la continuidad causal, ninguno puede ser Fettuccini. Esto es lo que quiere decir que la identidad personal no es independiente: la identidad de los demás individuos es relevante para la identidad de un individuo en particular. Pero esto es absurdo: que el de la caja de la izquierda sea o no Fettuccini no puede depender de que aparezca o no aparezca alguien en otra parte (¿si el de la derecha tiene un grano en la oreja del que carecía Fettuccini, puede el de la izquierda respirar tranquilo sabiéndose después de todo Fettuccini?)²⁸. El principio de que la identidad de algo depende de la de otras cosas es válido cuando se aplica a entidades tales como conjuntos, grupos y demás, pero produce resultados antiintuitivos cuando se aplica a personas. Lo que en realidad ocurre es que la relación *ser el continuador más próximo de* no equivale a la relación *ser la misma persona que*. Conceder a Nozick la no independencia de la identidad personal es *eo ipso* conceder la gradualidad y por ende la relatividad de la identidad personal. Pero justamente eso es lo que está en disputa.

En general, todos los defensores de teorías causales coinciden en que, si bien en los casos *reales* la identidad personal está bien definida y no plantea problemas, en los casos posibles

personal podría aceptarse como un modo *derivado* de hablar del yo, una traducción posible de nuestro hablar primario; lo cual sería indicativo de que la identidad personal no es algo lógica o metafísicamente último e inanalizable.

²⁷ *Philosophical Explanations*, Cambridge (Mass), Harvard Uni. Press, 1981; p. 33 y ss.

²⁸ El principio de que la identidad personal ha de ser *determinada* (que para cualquier caso dado alguien o es la persona *x* o no lo es, sin que quepa una tercera posibilidad) e *independiente* (que para la identidad de alguien la identidad de las demás personas es irrelevante) es defendido en R. Chisholm *Person and Object: A Metaphysical Study*, La Salle, Open Court, 1976. Para una reivindicación de la independencia de la identidad frente a los ataques de Nozick, véase H. Noonan: *Personal Identity*, London, Routledge, 1989; caps. 7 y 11.

problemáticos que plantean los experimentos mentales²⁹ la identidad personal es algo gradual, vago, a menudo objeto de convención o estipulación. Es célebre el problema planteado por el barco de Teseo: si cada tablón de su nave es sustituido poco a poco, el día en que no queda un solo tablón original ¿sigue siendo la misma nave? Pensamos que sí, porque aunque no hay identidad numérica hay identidad cualitativa y continuidad y contigüidad espacio-temporal. Pero ¿y si con los tabloncillos que han ido quitando han montado otro barco igual? Entonces dudamos y, seguramente, no hay una respuesta definitiva. Todo es cuestión de a qué factores *decidamos* dar más importancia, si a la mismidad (identidad numérica) de materia o a la contigüidad espacio-temporal (la forma es la misma en ambos). Pues bien, según Nozick la persona en los ejemplos problemáticos estaría en el mismo caso: no habría una respuesta *determinada* a la cuestión de su identidad, habría o que dejarla indeterminada o que estipularla (estableciendo límites precisos pero arbitrarios para lo que en sí es vago). ¿Hemos de aceptar que la identidad personal es en última instancia una cuestión de estipulación, vale decir, de *convención*? Quienes se niegan a aceptarlo defienden una concepción de la identidad personal como algo irreductible.

La concepción de la identidad personal como algo irreductible

Para los defensores de este enfoque la identidad personal es algo primitivo u originario, simple e indefinible, no analizable ulteriormente en términos de otras propiedades o relaciones ni, menos, reducible a ellas. Así, p.e., lo que constituye la identidad de Sócrates, lo que le hace diferente a cualquier otra persona por muy parecida, o inclusive cualitativamente idéntica, que fuese es *la socrateidad*: la propiedad de ser autoidéntico a Sócrates. Si esto es así, las cuestiones acerca de la identidad personal siempre tienen una respuesta taxativamente afirmativa o negativa, o sea, la identidad personal es determinada³⁰. De esta forma, cada persona posee una esencia individual o *estidad* (*haecceidad*) que la singulariza como única e irreplicable: una persona podría cambiar todas sus propiedades cualitativas (y a fortiori las relacionales), su memoria, su conciencia y su cuerpo y seguir, empero, siendo la misma persona. La concepción sustancialista de la persona como un alma espiritual simple —desde el Fedón de Platón hasta el ego cartesiano de la Meditación segunda— guarda una estrecha relación con esta teoría³¹, pero no está necesariamente implicada en ella. De igual modo, la versión aquí expuesta parece en principio irrefutable, pues de ella no se deduce ningún hecho. ¿Qué sentido tiene hablar de identidad personal si *todo lo observable* en una persona —incluso por ella misma— puede modificarse sin que se modifique su identidad? No obstante, veremos que hay ciertos hechos reales o posibles que se explican mucho mejor sobre la base de esta hipótesis que con las hipótesis alternativas, con lo cual no parece ser del todo cierto que la teoría sea completamente impenetrable por los hechos.

Por último, si bien es justo decir que estas teorías son hoy minoritarias, no lo es menos decir

²⁹ Y a veces los reales. Ver *infra* los casos de comisurotomía cerebral.

³⁰ Cf. R. M. Chisholm: «Reply to Strawson's Comments», en H.E. Kiefer & Milton K. Munitz (eds): *Language, Belief and Metaphysics*, Albany (New York), State Uni. of New York Press, 1970; p. 188 ss.

³¹ Este es fundamentalmente el caso de R. Swinburne en su contribución a S. Shoemaker & R. Swinburne: *Personal Identity*, y en los caps. 8-10 de su *The Evolution of the Soul*. Swinburne recoge la idea escotista de la haecceidad, aunque piensa que no es defendible que los objetos posean estidades, sino únicamente los organismos y en particular las personas.

que cuentan con algunos representantes de primera fila³². Las esencias individuales se han adoptado no solamente para resolver los problemas de la identidad personal, sino también en la semántica modal para dar cuenta de la identidad transmunda³³, para dotar de referencia a los términos singulares³⁴, y para otras diversas cuestiones de lógica filosófica³⁵. Con esto sólo pretendo decir que hay razones filosóficas de otra índole para creer en la existencia de esencias individuales además de las que puedan aparecer en un análisis de la identidad personal.

La concepción no reduccionista de la identidad personal tiene tres grandes ventajas sobre todas las demás. Una es que es capaz de "resolver" casi cualquier caso problemático sin "pillarse los dedos". Otra es que conecta directamente con lo que creemos de modo intuitivo acerca de la naturaleza de las personas. Y la tercera es que seguramente coincide con aquello que nos gustaría creer acerca de nosotros mismos: que somos algo único, individual e irreplicable y, por ello mismo, más fácilmente conservable o reconstruible. Sin embargo, es sabido que el que una creencia coincida con nuestros intereses la convierte automáticamente en sospechosa. Pero no sólo esta tercera "ventaja" es un defecto. También lo es la segunda, en la medida en que las creencias intuitivas o de "sentido común" suelen tener unas credenciales hartamente dudosas. Pero es sobre todo en la primera en la que me quiero concentrar: si esta teoría parece tener respuesta para todo es simplemente porque postula una entidad indetectable que constituiría tal respuesta, y eso posee un insostenible carácter ad hoc que cuenta negativamente en su contra. Éste es su primer gran defecto.

El segundo es que este carácter ad hoc va ligado a una peculiar carencia de contenido empírico: por definición no hay modo de definir la identidad personal, es algo último e inanalizable, compatible con todo hecho. Ello hace a la teoría no susceptible de verificación y, aunque no por ello carente de significado, doblemente sospechosa de vacuidad: parece que estamos postulando una entidad tan innecesaria como en su momento lo fueron el flogisto o el éter (aunque más adelante veremos cómo esta crítica no es del todo justa).

Junto a la *ad-hocidad* y a la carencia de contenido empírico, su tercer defecto es que "resuelve" todos los problemas a costa de no ofrecer nunca una respuesta *determinada*. P.e. en el caso de la caja fuerte, como en cualquier caso de fisión de personas, el defensor de la irreducibilidad de la identidad personal afirmará tranquilamente que Fettuccini es tras la duplicación aquel *que sea* idéntico con el Fettuccini de antes de la duplicación, o sea, el que posea su misma *estidad*, con lo cual Fettuccini puede ser el de la izquierda, el de la derecha, o ninguno. Es un modo elegante de resolver el problema, pero ¿no es un modo de devolver la pregunta? ¿No se trata precisamente de averiguar quién es ahora Fettuccini? El antirreduccionista observará que su teoría no constituye un *método de identificación*, sino sólo un criterio de identidad personal que no le permite en cada caso descubrir quién es quién. O mejor aún, el antirreduccionista niega que pueda haber, y por tanto que deban exigirse, criterios de identidad para las personas: para los objetos podemos ofrecer criterios de identidad en términos de tales o cuales propiedades o relaciones, pero no hay modo de reducir los enunciados de identidad sobre las personas a enuncia-

³² Aparte de los citados, G. Madell: *The Identity of the Self*, Edinburgh Uni. Press, 1981; y el libro de Chisholm *The First Person*, Minneapolis, Univ. of Minnesota Press, 1981. Y más recientemente *The Immaterial Self*, de J. Foster (London, Routledge, 1991); y *Indivisible Selves and Moral Practice*, de V. Haskar (Edinburgh Uni. Press, 1991).

³³ Cf. D. Kaplan: «Transworld Heir Lines», en M. Loux (ed.): *The Possible and the Actual*, Ithaca, Cornell Uni. Press, 1979.

³⁴ Cf. A. Plantinga: *The Nature of Necessity*, Oxford, Clarendon, 1974; esp. cap. V.

³⁵ Para una descripción general de estas tareas y defensa de las esencias individuales, véase M. Lososky: "Individual Essences", *American Philosophical Quarterly*, 24, (1987); 253-60.

dos sobre propiedades o relaciones distintas de la identidad³⁶. Así, Shoemaker mostró hace tiempo en que los enunciados de la memoria en tiempo pasado sobre uno mismo —p.e. "ayer rompí esa ventana"— no se trata en general de la identificación de un yo presente con uno pasado, de forma que estuviéramos aplicando algún criterio de identidad psicológico (basado en la semejanza, la continuidad o lo que sea), sino que no estamos aplicando ningún criterio de identidad porque tales enunciados no son la conclusión de inferencia alguna, sino que simplemente informan de lo que el hablante recuerda³⁷.

Una última objeción menor a la teoría es la de cómo determinar qué objetos poseen *estidades*. Parece claro que los objetos materiales no las poseen, pues yo puedo, p.e. partir mi coche por la mitad y con otras dos mitades construir dos nuevos, y la cuestión ¿cuál de ellos es ahora mi coche? carece de una respuesta determinada. Puesto que yo puedo reemplazar una por una las partes de un objeto, sin que éste deje de ser el mismo debido a la contigüidad espacio-temporal, no parece que podamos hablar de sus *estidades*. En las personas, por contra, la perspectiva interna, la autoconciencia de la identidad o el yo, constituiría algo peculiar e irremplazable que constataría —aunque no necesariamente constituiría— su *estidad*. Pero ¿y en los animales superiores? ¿Cabe también ahí hablar de *estidades*? ¿Y en los animales inferiores o incluso en el resto de los organismos? Todas estas son cuestiones difíciles que no podemos abordar aquí, pero que constituyen un serio problema de delimitación para las teorías no cualitativas de la identidad.

A pesar de todas estas objeciones, sí quiero dejar sentado que la teoría posee significado y es coherente. Esto se demuestra al ver que muchos hechos posibles (concebibles e imaginables) son solamente explicables con esta teoría —o al menos las demás teorías tienen muchas más dificultades para dar cuenta de ellos—, lo cual viene a mostrar que la teoría no carece enteramente de contenido empírico. Por ejemplo, la noción de transmigración parece claramente coherente: que una misma persona viva varias vidas en (sucesivos) tiempos y lugares diferentes, con cuerpos y cerebros diferentes, e incluso con memorias y conciencias de identidad personal muy diferentes. ¿Tiene entonces sentido decir que Julio César y yo somos la misma persona? Creo que la respuesta es sí, aunque no haya ningún hecho que nos induzca a pensarlo. Desde luego hay contrafácticos que reforzarían la hipótesis: si yo recordara hechos o vivencias acerca de Julio César que en esta vida es imposible que haya conocido o vivido y que pueda demostrarse de modo independiente que son verdaderos, eso sería una excelente razón para creer que yo soy esa otra persona (o p.e. si mañana me apercibiera de que soy Gautama Buda, el que de repente pudiera hablar perfectamente sánscrito sería una excelente razón en favor de ello). Pero si la transmigración fuera un hecho ¿hay alguna hipótesis alternativa mejor³⁸ para explicarla distinta de que la persona es una entidad independiente del cuerpo, del cerebro y de los estados mentales?³⁹.

³⁶ Cf. S. Shoemaker: *Self-Knowledge and Self-Identity*; p. 37-38.

³⁷ *Self-Knowledge and Self Identity*; p. 136-142.

³⁸ Que hay hipótesis alternativas es obvio. Por ejemplo, el conductismo, que no deja de ser una versión del materialismo, puede responder muy bien de la transmigración. Si una persona es sencillamente un conjunto de disposiciones conductuales, ese mismo conjunto de disposiciones puede ser poseído por varios sujetos a lo largo de la historia (Cf. Shoemaker, op. cit. p. 26). Pero eso mismo produce resultados antiintuitivos: varios individuos que existan simultáneamente podrían ser la misma persona si tuvieran las mismas disposiciones conductuales; y a la vez, uno no sería la misma persona a los 10 años que a los 60, pues sus disposiciones conductuales son, seguramente, muy diferentes.

³⁹ Curiosamente el budismo afirma la transmigración y niega entidad al sujeto. En este caso es *una personalidad* o un carácter (normalmente egoísta y obcecado) el que transmigra de un cuerpo a otro hasta que se purifica. Pero en

Ahora bien, incluso en el caso de que un determinado estado de cosas *lógicamente posible* sólo fuera explicable admitiendo la verdad de esta hipótesis, de ello no se sigue la verdad de la hipótesis, sino sólo su posibilidad lógica. Swinburne razona así:

"dado que para cualquier persona presente actualmente consciente, no hay imposibilidad lógica... de que esa persona continúe existiendo sin su cuerpo, se sigue que esa persona debe ahora realmente poseer una parte distinta de la parte corporal que pueda continuar, y que podemos llamar su alma... Porque no hay ni siquiera una posibilidad lógica de que si yo ahora consisto en nada salvo la materia y la materia es destruida, no obstante yo continuara existiendo. De la mera posibilidad lógica de mi existencia continuada se sigue el hecho real de que hay ahora en mí algo más que mi cuerpo; y ese más es la parte esencial de mí mismo"⁴⁰.

Pero ahí hay un error. Desde luego, si yo consisto sólo en mi cuerpo, no existe la *posibilidad real* de que yo siga existiendo una vez suprimido éste. Pero dado que, aunque yo de hecho sea solamente un cuerpo, siempre es lógicamente posible que yo consista en algo distinto del cuerpo, es asimismo lógicamente posible que siga existiendo sin el cuerpo, aunque no sea real o metafísicamente posible tal cosa.

Otro ejemplo que muestra la coherencia de esta hipótesis es la posesión (p.e. diabólica). La posesión implica que otra persona invade un cuerpo humano y pugna por controlarlo. No discuto la realidad de tales hechos, sino su coherencia: si el que un mismo cuerpo pueda estar regido por más de una persona sucesivamente o a la vez (aunque estas personas pugnen) es algo concebible e imaginable —como sin duda lo es— entonces la teoría de la identidad personal como algo diferente del cuerpo-cerebro y de la memoria-conciencia es coherente. Desde luego un caso de esquizofrenia completa o de personalidad múltiple se explica normalmente como si se tratara de una persona con problemas psicológicos, pero las teorías psicologistas de la identidad personal deben admitir que ahí hay dos personas, pues hay dos corrientes *disjuntas* de conciencia. Los antirreduccionistas, en cambio, pensarán que puede tratarse de dos personas ocupando un cuerpo, aunque lo más probable es que una persona posea dos corrientes independientes de conciencia (esto último es un contrasentido para los reduccionistas psicológicos). De igual modo, la teoría no reduccionista de la identidad personal explicaría muy bien el caso posible de una persona que ocupara más de un cuerpo (cf. infra).

Como quiera que las teorías materialistas de la identidad personal parecen poco aceptables, que las psicologistas han sido superadas por las del nexo causal, y que las negadoras de la identidad personal en su versión débil —la única coherente— se asemejan mucho a estas últimas, las alternativas parecen reducirse a dos: alguna versión de la teoría del nexo causal o alguna versión de las teorías no cualitativas de la identidad personal. Analicemos algunos ejemplos a modo de experimentos cruciales para optar por una u otra de las teorías.

Decisión entre teorías mediante experimentos mentales

Uno de los métodos usuales para optar entre estas teorías consiste en imaginar un caso problemático (por extravagante que parezca con tal de que sea lógicamente posible) y comparar las respuestas que nos dan las diversas teorías con lo que nuestras intuiciones previas nos dicen que es la respuesta correcta. Contra este proceder se han suscitado algunas objeciones:

1) ¿Qué sentido tiene plantearse posibilidades totalmente implausibles en vez de resolver los

principio no se ve por qué dos seres humanos de distintas épocas con caracteres y personalidades bien distintas no podrían ser la misma persona.

⁴⁰ *The Evolution of the Soul*; p. 154.

casos *reales* de identidad personal? Nuestros conceptos de persona e identidad personal son adecuados a los casos normales y funcionan perfectamente bien en el lenguaje ordinario. Plantear cómo resolverían *esos mismos conceptos* los casos extraordinarios no tiene ningún sentido, pues para tales casos *no hemos tenido aún que desarrollar ni conceptos ni términos*⁴¹. Bien, pero lo que quiere el filósofo —o al menos algunos— es no solamente averiguar cómo se usan las palabras y se forjan los conceptos en el lenguaje ordinario y cuán adecuados son a sus necesidades actuales, sino si el concepto se adecua *en última instancia* a su objeto para, de no ser así, modificar el concepto o construir uno enteramente nuevo. Se trata, pues, de dos tareas distintas, la segunda con más importe filosófico (o metafísico).

- 2) No siempre hay intuiciones uniformes acerca de las respuestas intuitivamente correctas en un caso dado. Esto es simplemente cierto, y muestra una de las limitaciones de los experimentos mentales.
 - 3) Algunos de tales experimentos son tan complejos que, lejos de aclarar nada, confunden aún más. La receta obvia en este caso es: empléense solamente ejemplos más claros que las nociones a dilucidar.
 - 4) Algunos plantean situaciones cuya coherencia lógica (concebibilidad) es más que dudosa. De nuevo en tales casos habrá que negar validez, no a las consecuencias, sino al ejemplo mismo.
- En suma, con ciertas restricciones y como método complementario de la descripción de los rasgos de la identidad personal, muchos experimentos mentales no carecen de utilidad iluminadora para elegir entre las teorías. Veamos algunos de ellos.

Un primer caso es el que denominaré la paradoja de la resurrección. Supongamos que Dios quiere resucitarme el día del Juicio final: ¿qué hará? Según la teoría materialista de la identidad personal crear un cuerpo y un cerebro cualitativamente idéntico⁴² al mío en alguna fase de mi vida (posiblemente justo el del instante antes de la muerte, aunque quizá mejor el de cuando tenía veinte años). Esto parece poco plausible: primero porque yo podría seguir viviendo aun sin cuerpo, y segundo —y más fácilmente aceptable— porque parece claramente concebible e imaginable que yo podría volver a vivir con otro cuerpo, incluido otro cerebro, máxime si ese otro cerebro está en el mismo estado neuro-psicológico en el que yo estuve en un momento dado. Pero si eso es posible, un ser omnipotente puede realizarlo.

Según la teoría psicologista Dios creará un hombre con mis mismos recuerdos, conciencia y demás disposiciones mentales, y ése seré yo. Pero lo que yo quiero no es que Dios resucite a un individuo con mis recuerdos, conciencia, intenciones etc. y que crea ser yo, *sino que me resucite a mí*, incluso aunque haya perdido muchos recuerdos, intenciones, y demás propiedades psíquicas. La teoría del nexo causal añadiría que haya una continuidad garantizada causalmente entre yo ahora y el que resucite en el futuro. Desde luego, esa continuidad no puede ser real, pues seguramente de aquí al día del Juicio habrá muchos siglos de desconexión (a menos que pensemos que en el ínterin existe la persona sin cuerpo como espíritu puro, cosa que sólo defienden algunos teóricos de la identidad no cualitativa). Pero la continuidad psicológica podría

⁴¹ Esta crítica, que coincide con la posición del segundo Wittgenstein de que el lenguaje "está bien como está", aparece en A. Flew: *The Logic of Mortality*; p. 123-24, y es la base de *Real People: Personal Identity without Thought Experiments* (Oxford Uni. Press, 1988) de K.W. Wilkes.

⁴² Si se trata de una teoría materialista estricta, como la de T. Nagel (cf. su *Mortal Questions*, Cambridge Uni. Press, 1979) podría ser necesaria además la identidad numérica de mi materia, lo cual puede ser muy problemático si esos mismos componentes han formado parte a lo largo de la historia de algún otro ser humano que Dios también desee resucitar: *¿en tal caso ni siquiera Dios podría resucitar a ambos!* Para Williams, en cambio, la identidad cualitativa de la materia es suficiente.

ser como la que se da al despertarnos (aunque en el caso del sueño siempre queda el cuerpo como conector *real* y se trata de un período de sólo unas horas). ¿Admitiría el teórico causal de la identidad que Dios constituye una garantía como causa de la continuidad psicológica entre mi yo ahora y mi yo cuando resucite?

La teoría no reduccionista de la identidad personal lo tiene todo más fácil: Dios simplemente tiene que volver a actualizar la misma esencia individual, volver a dar vida al mismo individuo.

Segundo caso: ¿cómo tratan las diversas teorías la duplicación —p.e. el referido caso de Fettuccini en su caja fuerte? Todas las teorías aceptarían que los dos no pueden ser Fettuccini⁴³. Obviamente, si se trata de un caso de *duplicación*, entonces para todas las teorías —aunque por distintos fundamentos— Fettuccini es *el original* y el otro es sólo su réplica. Pero supongamos que se trata de una genuina fisión: los dos son igual de contiguos espacio-temporalmente con el original. Para los materialistas estrictos (Nagel) Fettuccini sería aquél cuyo cuerpo-cerebro sea numéricamente idéntico con el original. Si los dos son numéricamente distintos —aunque cualitativamente idénticos— ninguno es Fettuccini. Si cada uno tiene la mitad de la materia del original, el materialista no lo tiene fácil. Tampoco lo tiene fácil el materialista en sentido amplio (Williams) que sólo exige la identidad cualitativa de la materia, porque los dos resultantes son cualitativamente idénticos a Fettuccini en lo corporal, pero, por el principio de la necesidad de la identidad, no pueden ser ambos Fettuccini. Para la teoría causal, si se trata de un caso de genuina fisión, el problema de la identidad es espurio: decir que ninguno es Fettuccini, o que es uno, o que es el otro, son simplemente tres modos de describir un mismo hecho que conocemos de modo cabal: no queda ningún hecho ulterior en el que pudiera consistir la identidad personal de Fettuccini⁴⁴. En esto tal teoría —al menos en la versión de Parfit— se aproxima mucho a las teorías negadoras de la identidad personal, para las que el problema simplemente no surgiría. Para la variante del continuador más próximo (Nozick), al estar los dos causalmente igual de emparentados con Fettuccini, ninguno es Fettuccini. Para los antirreduccionistas Fettuccini será el que posea su *estidad*: uno, u otro o, probablemente, ninguno. Esta teoría es la única que permite aceptar algo que parece relativamente plausible: que Fettuccini *podría* ser tras la fusión *uno y no el otro* aunque ambos sean indiscernibles.

Tercer ejemplo. ¿Qué pasaría en un caso de irradiación de personas o teletransporte? (donde una persona es descompuesta en sus átomos, la información irradiada a la velocidad de la luz, y luego recompuesta en otro lugar). Para un materialista estricto, puesto que no hay continuidad espacio-temporal de la materia (la identidad numérica no se conserva en el teletransporte: repárese en que se irradia la información en forma de energía, pero no la materia misma) la identidad personal no se conserva. Para un materialista no estricto, dado que los átomos de la persona integrada son cualitativamente idénticos con los de la teletransportada (y están en el mismo orden y disposición), sí se conservaría. Para un teórico del nexo causal —que en este caso coincide con Nozick— si al integrarse en otro planeta se destruye el individuo de aquí, el del otro planeta es ahora él, pero si algo falla y el de aquí no se destruye, éste es la persona en cuestión, y el otro sólo su réplica, pues la conexión causal del de aquí es más clara. Para los no reduccionistas que el teletransportado sea o no la misma persona dependerá una vez más de que conserve o no su esencia individual. En principio parece lógico que lo sea, dado que esa esencia

⁴³ Una excepción la constituye A. Prior, para quien la identidad personal no es una relación transitiva en los casos de fisión, de modo que *P* puede ser el mismo que *P'* y el mismo que *P''*, pero *P'* y *P''* ser diferentes (cf. su «Time, Existence and Identity», *Proc. Arist. Society*, 1965-6). Vimos también como D. Lewis defiende algo de esa guisa.

⁴⁴ Parfit: *Reasons and Persons*; p. 258.

individual es enteramente independiente de la identidad numérica de su materia.

Aunque en todos estos ejemplos el antirreduccionista sale airoso (incluso en muchos parece salir victorioso) hay un par de casos verdaderamente problemáticos para el antirreduccionismo. Supongamos⁴⁵ que Marta, una chica preciosa de veinte años y con un sorprendente parecido a Claudia Shiffer, padece un tumor progresivo generalizado, y que un gran cirujano amigo suyo decide irle extirpando progresivamente las células infectadas de su cuerpo y su cerebro, a la vez que las va sustituyendo por las células correspondientes de Claudia Shiffer, a la que previamente ha raptado. Al final del proceso, cuando todas las células han sido sustituidas, obviamente Marta ya no existe, sino que en su lugar está Claudia Shiffer, con su cuerpo y su conectividad y continuidad psicológicas. La cuestión es: ¿qué pasa en los casos intermedios? ¿Cuándo exactamente se convierte Marta en Claudia? Según Parfit esto representa un problema insuperable tanto para las teorías materialistas como para las no reduccionistas. Para las primeras, porque no hay un modo no arbitrario de establecer cuánta cantidad del cuerpo/cerebro original es necesaria para preservar la identidad personal. Para las segundas, porque según ellas debería haber un punto crítico (¿acaso el 50%?) tras el cual Marta dejaría de existir y en su lugar aparecería Claudia Shiffer. Pero, —advierte Parfit— aunque el experimento sea técnicamente irrealizable, es muy poco probable que ocurriera eso. Lo más plausible es que Marta fuera difuminando poco a poco su identidad y adquiriendo poco a poco la de Claudia Shiffer a medida que se le implantaba más y más del cerebro de Claudia, *sin que podamos determinar un momento exacto* en el que una se convierta en otra. En todos los casos intermedios la identidad personal sería dudosa, indeterminada.

¿Qué podría replicar el antirreduccionista? Sin duda este es un caso difícil. El antirreduccionista podría pensar que el punto crítico se hallará próximo al 50% del intercambio cerebral, aunque no podamos determinar a priori dónde exactamente, y que a partir de ese punto Marta habría dejado de existir y en su lugar estaría Claudia; y que el cambio de personalidad, carácter, memoria, conciencia etc. sería igualmente repentino en ese instante. Convengo con Parfit en que parece más esperable que el cambio psicológico fuera gradual —como el material— en vez de repentino.

Pero hay una alternativa que Parfit no considera. Podría ocurrir que cuando Marta tenga una cantidad *suficiente* del cerebro de Claudia (p.e. un cuarto, o un tercio, o lo que sea) aparezca Claudia sin que desaparezca Marta, esto es, que ese cuerpo esté poseído por dos personas a la vez, aunque mejor controlado por Marta que por Claudia mientras haya más de su cerebro que del de Claudia. Y así hasta que se llegue al punto crítico en el que Marta desaparezca (p.e. cuando quede menos de un cuarto de su cerebro). En los casos intermedios Marta y Claudia ocupan el mismo cuerpo, aunque cada vez lo va controlando más Claudia y menos Marta. Esto concuerda con una tesis previa de los antirreduccionistas: *que un mismo cuerpo puede estar ocupado por más de una persona a la vez*⁴⁶. Evidentemente, cuando hay un tercio del cerebro de Claudia en Marta, dos tercios de ese cerebro siguen en la Claudia original (que para entonces

⁴⁵ Este experimento, al igual que el que sigue, y sus consecuencias proceden de Parfit (op. cit. p. 236 y ss) aunque lo he modificado ligeramente.

⁴⁶ O tal vez sólo sucesivamente, pero entonces podría ocurrir que la Marta con un tercio del cerebro de Claudia alternara sus períodos de ser Marta con sus períodos de ser Claudia (tuviera alternativamente la personalidad y la continuidad de conciencia y memoria) en razón de dos a uno. Insisto en que el flaco de esta doctrina no es su coherencia, sino su verosimilitud. Si se pudiera establecer más allá de toda duda razonable p.e. un solo caso de posesión de una persona por otra, quedaría demostrada la verdad de tal hipótesis.

habrá perdido un tercio de su cuerpo), con lo cual Claudia está en dos cuerpos a la vez. Esto implica que una misma persona puede ocupar más de un cuerpo, algo menos fácil de digerir por los antirreduccionistas y que nos lleva al segundo caso crítico planteado por Parfit.

Este caso es más problemático, entre otras cosas por estar más cerca de la realidad. En las últimas décadas a varios pacientes epilépticos se les ha seccionado el cuerpo calloso que conecta los dos hemisferios cerebrales para evitar la extensión cerebral global de los ataques de epilepsia. Aunque sigue habiendo una conexión mediante el cerebro inferior, las vías normales y principales de comunicación interhemisférica han desaparecido. Un efecto colateral indeseado en los pacientes es que hay una desconexión psicológica entre las funciones de ambos hemisferios: la mano derecha no sabe lo que hace la izquierda. Así, si se le presentan visualmente al hemisferio derecho objetos familiares es capaz de reconocerlos pero no de nombrarlos, mientras que el izquierdo reconoce a los objetos por su nombre pero no por su imagen. En otro caso, un paciente abraza a su esposa con el brazo derecho, mientras la rechaza con el izquierdo. En suma, parece como si hubiera *dos corrientes de conciencia independientes y diferentes en un mismo sujeto*. Pero esto es crucial, porque se supone que la unidad del sujeto es lo que produce la unidad de la conciencia⁴⁷, pero al tener aquí dualidad de conciencia seguramente debemos admitir una dualidad del sujeto, surgiendo así la ineludible pregunta: ¿cuál soy yo?⁴⁸

Pues bien, supongamos que yo soy un caso raro en el que ningún hemisferio es dominante, y que cada uno de mis hemisferios se implanta en un cuerpo que es una clonación del mío, teniendo los dos sujetos resultantes plena continuidad psicológica conmigo antes del transplante. La cuestión es ¿sobrevivo yo y quién soy ahora?⁴⁹

En principio caben cuatro respuestas: 1) no sobrevivo; 2) soy uno; 3) soy el otro; 4) sobrevivo como los dos. Para cualquiera que piense que la identidad es determinada no hay más respuestas posibles. 4 es inaceptable, porque los dos tienen ahora una conciencia y unas experiencias diferentes (¡incluso ocurre así en los casos reales de comisurotomía cerebral!), y si fueran la misma persona, alega Parfit, tendrían una sola corriente de conciencia (la famosa unidad de la conciencia). 2 y 3 son igualmente inaceptables, pues los dos hemisferios, por hipótesis, cumplen las mismas funciones. Y 1 es muy poco plausible, pues si una persona puede sobrevivir con medio cerebro (como de hecho ha ocurrido) *a fortiori* sobrevivirá con las dos mitades. "Sugiero que la concepción reduccionista es cierta. Quizá aquí no hay posibilidades diferentes, cada una de las cuales podría ser lo que suceda, aunque nunca pudiéramos saber lo que realmente sucede. Quizá, al saber que cada persona resultante poseería una mitad de mi cerebro, y sería psicológicamente continua conmigo, sabemos todo. ¿Qué estamos suponiendo cuando sugerimos, por ejemplo, que una de las personas resultantes podría ser yo? ¿Qué haría que esto fuera la respuesta verdadera?"⁵⁰. Para Parfit, pues, la cuestión carece de sentido. Se trata de modos alternativos de describir un mismo hecho del que conocemos todo. Preguntar si yo sobrevivo es presuponer que la identidad personal es *un hecho ulterior*, algo que, aunque inverificable, me queda por conocer⁵¹. No obstante, admite que 1 es probablemente la descripción preferible del

⁴⁷ Cf. Kant: *Crítica de la razón pura*, B 132 y ss.

⁴⁸ Algunos piensan que en tales casos tenemos una persona con dos mentes, o incluso dos personas en un cuerpo, pero eso deja intacta la cuestión: ¿cuál de las dos mentes (personas, sujetos) soy yo?

⁴⁹ Parfit, op. cit. cap. 12.

⁵⁰ Parfit, op. cit. p. 258.

⁵¹ Es curioso que Parfit plantee la cuestión aquí en términos de supervivencia y no de identidad, pues este caso se parece bastante al de la fisión, y tanto allí como aquí los dos nuevos sujetos son los supervivientes del anterior al tener continuidad psicológica plena con él.

caso.

¿Qué alegrará en este caso el antirreduccionista? En primer lugar hay que hacer referencia a los hechos. Porque parece que en la mayoría de los casos la unidad de conciencia se conserva y nadie, salvo el paciente, se da cuenta de las pequeñas disfunciones⁵². Es verdad, empero, que si la "dualidad de conciencia" no es tan clara como afirma Parfit, ello puede deberse a que permanecen intactas conexiones de los hemisferios con regiones inferiores interconectadas, y a que ambos hemisferios han tenido una larga vida de experiencias comunes. Qué ocurriría en casos donde la desconexión fisiológica fuera total y no hubiera experiencia común previa es algo que ignoramos.

Otro hecho cuestionable es la supuesta conciencia de *los dos* hemisferios. De hecho parece que casi siempre es dominante un hemisferio —aquél que es portador del lenguaje—, y que ése y sólo ése es el portador de la autoconciencia, y con ello de la identidad personal o del yo. Esta, que es la respuesta de Eccles y Mackay, es bastante discutible, pues el hemisferio no dominante tiene funciones psíquicas bastante elevadas⁵³. Parece que habrá que investigar más a fondo los hechos para ver si el hemisferio menor posee capacidades superiores tradicionalmente atribuidas a las personas —como hacer proyectos, tomar decisiones voluntarias, juzgar según criterios de valores— que lo cualifiquen como una persona distinta.

Un último hecho importante que Parfit deja de lado es que varias actividades gobernadas por el hemisferio menor escapan al control voluntario del sujeto⁵⁴. Por ejemplo, la mano y el brazo izquierdos, gobernados por el hemisferio derecho, se mueven por su cuenta. Esto puede interpretarse, no como que están siendo gobernadas por otro sujeto (Parfit), sino como que están fuera de control, esto es, que funcionan de manera mecánica como si de un robot se tratara. Entonces el experimento mental de Parfit podría explicarse diciendo que uno de los dos individuos resultantes no sería una persona sino un autómatas, cosa que delatarían los oportunos tests psicológicos. Concediendo que los hechos fueran como Parfit supone, Swinburne contesta que "por mucho que supiéramos en tal situación acerca de lo que le sucede a las partes del cuerpo de una persona, no sabríamos con certeza qué le sucede a la persona"⁵⁵. Esto no quiere decir que no quede nada por saber, sino que no queda nada por descubrir experimentalmente. Pero queda por saber lo fundamental: quién (si alguno) sigo siendo yo. Qué sucedería de llevarse a cabo el experimento es algo que simplemente no lo sabemos. Recuérdese que la continuidad de memoria y conciencia no es garantía de la identidad personal para un antirreduccionista, sino sólo evidencia en su favor.

Hay tres puntos importantes tocantes a algunos de los experimentos mentales anteriores. Uno

⁵² "La evidencia más importante sobre la unidad de la conciencia es con mucho la suministrada por los estudios de Sperry y sus colaboradores realizados sobre pacientes que habían sido sometidos a una comisurotomía... Fue una inmensa fortuna que el seccionar... el cuerpo caloso causase tan poca incapacidad al paciente; tanto fue así que *no se reveló nada significativo* hasta que Sperry realizó sus altamente discriminativas investigaciones". J.C. Eccles: *La psique humana*, Madrid, Tecnos, 1986; p. 26 (subrayado mío).

⁵³ D.M. & V. Mackay: «Explicit Dialogue Between Left and Right Half-Systems of Split Brains», *Nature*, 295 (1982); p. 690 ss. J.C. Eccles: *La psique humana*; p. 34-5. Eccles reconoce lo problemático de esta respuesta, pues el hemisferio menor, aunque carece de capacidad de expresión lingüística, es capaz de reconocimiento verbal (en el cap. 1 de este libro aparece todo un catálogo de interpretaciones distintas o contrarias a las de Eccles). En *The Human Mystery* Eccles afirma: "El descubrimiento más destacable de estos experimentos fue que todas las actividades neuronales en el hemisferio derecho son desconocidas para el sujeto hablante, que sólo está en ligazón con las actividades neurales del hemisferio izquierdo" (London, Routledge, 1984; p. 219).

⁵⁴ Cf. K. Popper y J.C. Eccles: *El yo y su cerebro*, Barcelona, Labor, 1980; p. 354 y 369. Y *La psique humana*; p. 35.

⁵⁵ *The Evolution of the Soul*, p. 148-9.

es que en su descripción va a menudo camuflada su solución. P.e. en los casos de fisión de personas, no es lo mismo si la persona se divide de modo natural —por infrecuente que sea— que si se divide por una operación quirúrgica o por entrar en un reduplicador.

Otro, que muchas veces necesitamos más información para poder decidir acerca de qué ha sucedido realmente. P.e. tras la fisión de la persona puede ocurrir que cada uno de los fisionados tenga una corriente de conciencia independiente. Pero podría también ocurrir que las conciencias estuvieran perfectamente interrelacionadas. Si tras fisionarme yo, uno de mis continuadores se queda en Madrid y el otro se va a París, y el de Madrid percibe sensualmente todo lo que percibe el de París, lo almacena en su memoria como el de París, y experimenta intenciones, deseos, impulsos etc. provocados por tales experiencias; y al de París le ocurre lo mismo con el de Madrid, entonces calificaríamos el caso de esta fisión como que yo había pasado a habitar dos cuerpos simultáneamente. En un caso así la multiplicación de cuerpos no implicaría la de las personas. Esta posibilidad, que me parece claramente coherente, no suele ser tenida en cuenta en los estudios sobre la identidad personal. Creo que sólo la hipótesis no reduccionista explicaría casos así. Desgraciadamente, no tenemos ninguna noticia de nada parecido a tales casos (salvo quizá algunos informes sobre experiencias simultáneas en gemelos distantes, o de santos que supuestamente han estado en dos sitios a la vez)⁵⁶; y en los únicos casos de división real que conocemos se da justamente una falta de interacción entre las dos corrientes de conciencia, lo que nos hace pensar en dos personas distintas.

Un tercero, que es importante si en esta descripción más completa se incluye o no el aspecto causal. Si en la descripción del experimento incluimos, además de lo que es manifiesto para cualquier observador, la causa del hecho, entonces puede parecer aceptable la tesis de Parfit de que *ya no queda nada más que preguntar*. Estos tres factores se dan en su caso crucial de división cerebral contra las teorías no reduccionistas, y por ello lo hacen un escollo especialmente insalvable para ellas.

A modo de conclusión

En la medida en que puede haber una conclusión de todo el debate anterior creo que es esta: no contamos aún con una teoría plenamente satisfactoria de la identidad personal; todas poseen demasiados contraejemplos. En cierto sentido, las teorías causales de la identidad personal parecen más plausibles, pero para superar sus dificultades se hacen más y más permisivas tendiendo a borrar su propia identidad como teorías y a fundirse con las teorías negadoras del sujeto, teorías que en último extremo parecen inaceptables. Llegados a este punto, las teorías no reduccionistas pueden parecer preferibles, pues capturan muchas de nuestras intuiciones más profundas, si bien no están exentas de sus propias dificultades.

Creo que como requisitos mínimos para cualquier teoría de la identidad personal podrían establecerse los siguientes:

a) la necesidad de la identidad: que si dos cosas son idénticas, entonces son necesariamente idénticas.

⁵⁶ La cuestión de si una persona puede estar en dos lugares a la vez no es la misma que la de si un objeto puede ocupar a la vez dos espacios. Aunque no es fácil demostrar que su negación sea contradictoria, el que un mismo objeto no puede estar en más de un sitio a la vez es un principio que se deduce de las propiedades del espacio. Por contra, en los casos de dislocación no se trataría de un mismo cuerpo en dos lugares, sino de dos cuerpos en dos lugares pero gobernados por una sola mente (o pertenecientes a una sola persona).

ticas o, dicho de otro modo, que si dos cosas son una en el mundo real, en todo mundo posible en el exista una existe la otra y es una misma con la primera. Esto quiere decir que una cosa no puede ser (haber sido o llegar a ser) dos y seguir siendo *esa* cosa. Con ello en los casos de fisión o división en todas sus vertientes eliminamos la cuarta posibilidad, la que consistía en que el sujeto tras la división es *él mismo* los dos nuevos sujetos.

- b) La independencia lógica de la identidad, al menos para individuos: que la identidad de un individuo no puede depender de la de los demás individuos. Si hacemos depender la identidad de un individuo de la de otros, inmediatamente hemos convertido la identidad en relativa⁵⁷. Admitir la dependencia lógica de la identidad personal equivale a admitir que la identidad personal *no es un caso de genuina identidad*.
- c) La no relatividad de la identidad personal, su determinabilidad; o sea, que la pregunta de si yo existo o no siempre tiene una respuesta determinada: sí o no. Que hay algo absurdo en pretender que yo puedo sobrevivir "en mayor o menor parte".

Con estos requisitos, creo que podemos afirmar que la identidad de las personas es más fuerte que la de los organismos, y mucho más fuerte que la de los objetos, porque su individuación es mayor. Queda sin resolver si la individuación de las personas es algo absoluto o último. El sujeto humano o es una realidad última, simple e inanalizable, o las relaciones de semejanza y continuidad causal le dotan de unidad suficiente para constituir un individuo diferenciado —aun en ausencia de todo elemento corpóreo—. Claro que, decir esto, es probablemente dejar las cosas más o menos donde estaban: lejos aún de una teoría universalmente aceptable de la identidad personal. ¿Consiste en algo ser una y la misma persona? Como en otras cuestiones filosóficas, también aquí es por el momento prudente *una cierta dosis* de agnosticismo.

(Septiembre de 1992)

⁵⁷ En este sentido los objetos no poseen una verdadera identidad al carecer de una individuación absoluta: podemos sustituir progresivamente *todas* las partes de un objeto hasta que la cuestión de *si es o no el mismo objeto* carezca de sentido o su respuesta sea meramente convencional. El principio de independencia de la identidad *personal* está implícito en «The Self and the Future» the Bernard Williams (en *Problems of the Self*; pp. 46-64). En un artículo anterior («Personal Identity and Individuation», *ibidem* pp. 1-19) defendió un principio aún más fuerte: que una persona es la misma que otra en dos momentos diferentes sólo si *no puede* haber una tercera que esté en la misma relación con la primera. Aunque atractivo, tengo dudas acerca de la validez de este principio.