



UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

FACULTAD DE FILOSOFÍA

Máster Universitario en Filosofía Teórica y Práctica

Especialidad de Lógica, Historia y Filosofía de la ciencia

Trabajo Fin de Máster

**La identidad personal no es lo que importa
La concepción reduccionista de Derek Parfit**

Autor: Ramon Domènech Arévalo

Tutor: Julio C. Armero San José

Madrid, septiembre de 2018

Resumen:

El problema de la identidad personal a través del tiempo es uno de los más interesantes problemas metafísicos. Numerosos pensadores han tratado de determinar en qué consiste nuestra existencia continua, tratando de encontrar las condiciones de nuestra persistencia a lo largo del tiempo. En este trabajo se analiza la propuesta reduccionista de Derek Parfit y su idea de que la relación que importa no es la identidad personal sino la continuidad y/o conexividad psicológica. Se examinan las distintas afirmaciones en que consiste la teoría de Parfit y se analizan algunas de las objeciones que ha recibido. Se propone también la revisión de algunas de dichas afirmaciones para tratar de alcanzar una teoría que reúna las virtudes de la concepción de Parfit y que no se encuentre expuesta a las mismas objeciones que tienen éxito contra esta. Se analiza también la concepción animalista de la identidad personal y se trata de mostrar que la concepción reduccionista es superior a dicha teoría.

Abstract:

The problem of personal identity through time is one of the most interesting metaphysical problems. Many thinkers have tried to determine what our continuous existence consists in, trying to find the conditions of our persistence over time. This paper analyzes the reductionist proposal of Derek Parfit and his idea that the relationship that matters is not personal identity but psychological continuity and/or connectivity. The different ideas in Parfit's theory are examined and some of the objections he has received are analyzed. I also propose the revision of some of these ideas to try to reach a theory that combines the virtues of Parfit's conception and that is not exposed to the same objections that are successful against it. The animalist conception of personal identity is also analyzed and I try to show that the reductionist conception is superior to this theory.

Índice

Introducción	4
1. La concepción reduccionista de Derek Parfit	6
2. La objeción de la circularidad y la posibilidad de una descripción impersonal del mundo ...	10
3. Refutando la tesis no reduccionista	18
4. Cerebros divididos	20
5. Bernard Williams: La importancia del cuerpo para la identidad personal	29
6. La vaguedad lingüística y la indeterminación de la identidad personal: los espectros psicológico, físico y combinado	35
7. La identidad personal no es lo que importa.....	51
8. Similitud psicológica: eliminando la causalidad de la relación R.....	59
9. La concepción animalista de la identidad personal	64
10. Daniel Dennett: El yo como producto biológico.....	81
11. Conclusiones	86
Referencias.....	88

Introducción

Uno de los problemas metafísicos más interesantes es el problema de la identidad personal a través del tiempo. Resolver dicho problema consiste en averiguar cuáles son nuestras condiciones de persistencia, es decir, cuáles son las condiciones necesarias y suficientes para que dos personas que existen en dos momentos distintos sean una y la misma persona¹. Lo que se quiere saber es, por lo tanto, en que consiste la continuidad de nuestra existencia a lo largo del tiempo, es decir, que es lo que hace que podamos decir que somos la misma persona a lo largo de nuestra vida y no una sucesión formada por personas distintas.

Es necesario distinguir entre identidad cualitativa e identidad numérica. La identidad cualitativa no es más que similitud exacta. Dos personas son cualitativamente idénticas si son exactamente iguales. Si pudiéramos escanear la posición de las partículas de que está compuesta una determinada persona y pudiéramos crear una réplica suya, estaríamos, al menos en un primer momento, ante dos personas cualitativamente idénticas. Sin embargo, no estaríamos ante dos personas numéricamente idénticas, es decir, ambas personas no serían una y la misma persona. Esto último es obvio, puesto que una misma persona no puede estar en dos lugares distintos simultáneamente. Del mismo modo que puede haber identidad cualitativa en ausencia de identidad numérica, es también posible que se dé la última sin que esté presente la primera. A medida que pasa el tiempo, las personas vamos cambiando. Ninguno de nosotros somos cualitativamente idénticos a la persona futura a la que somos numéricamente idénticos. La identidad cualitativa no es, por lo tanto, condición necesaria ni suficiente para que haya identidad numérica, y lo mismo puede decirse de esta última con respecto a la primera. La pregunta por el criterio de identidad personal a través del tiempo es una pregunta por las condiciones que han de darse para que exista identidad numérica. La existencia o no de identidad cualitativa no es, en principio, relevante para el problema.

Normalmente, cuando realizamos juicios acerca de la identidad numérica de cualquier objeto a través del tiempo, solemos considerar que el criterio es la continuidad física de la cosa en cuestión o, por lo menos, de las partes que la componen. Sin embargo, este criterio, aun suponiendo que sea correcto en la mayor parte de los casos, no lo es en todos. Existen entidades, como las naciones o los clubs, a las que no es posible aplicar un criterio que se base en la continuidad física. En el caso de los organismos biológicos, en cambio, aunque dicha continuidad sí que tiene cierta importancia, no parece ser condición suficiente. Para que

¹ Como señala Eric Olson (2017), esta forma de plantear la pregunta presupone que somos personas esencialmente, por lo que sería mejor preguntar, como afirma el autor, por las condiciones necesarias y suficientes para que una persona existente en cierto momento sea numéricamente idéntica a algo que exista en un momento distinto.

consideremos que un organismo sigue existiendo, parece ser necesario que siga siendo un organismo vivo. El caso más complicado parece ser el que, para nosotros, es también el más importante. Se trata del caso de nuestra propia identidad numérica a través del tiempo. En dicho caso, nuestras intuiciones parecen llevarnos en distinta dirección dependiendo de la situación que nos planteemos. Se han propuesto multitud de soluciones al problema. Hay quienes creen que el criterio de identidad correcto es, en nuestro caso, el mismo que empleamos para la mayor parte de objetos, es decir, la existencia de continuidad física. Otros piensan que el criterio correcto de identidad personal es algún tipo de continuidad o conexión psicológica. Existe también la opinión de que somos entidades inmateriales como, por ejemplo, almas y de que nuestra persistencia consiste, en consecuencia, en la permanencia de dicha entidad. Otros autores defienden que no somos ningún tipo de entidad inmaterial pero que la identidad personal es un hecho adicional y que no es necesario, por lo tanto, recurrir a ningún tipo de continuidad para que sea el criterio de esta.

Una de las teorías de la identidad personal más interesantes es la concepción reduccionista que defiende Derek Parfit en la tercera parte de su libro *Razones y personas*². En este trabajo analizaré la teoría de Parfit con el objetivo de defender muchas de sus afirmaciones pero también con el de revisar algunas de ellas para tratar de llegar a una teoría que conserve los puntos fuertes de dicha concepción pero que no se encuentre expuesta a las objeciones a que esta es vulnerable. En el primer apartado expondré brevemente las principales afirmaciones en que consiste la teoría de Parfit, afirmaciones que examinaré más detenidamente a lo largo del trabajo. En el segundo apartado expondré la objeción de la circularidad, una de las más importantes objeciones al criterio psicológico, que es el que defiende Parfit. En ese mismo apartado analizaré también la afirmación del autor de que es posible una descripción impersonal completa del mundo y trataré de defenderla de algunas objeciones. En el tercer apartado me centraré en examinar los principales argumentos que dirige Parfit contra la tesis de que somos entidades que existen separadamente, como sería el caso si fuéramos almas inmateriales. El cuarto apartado lo dedicaré a tratar ciertos casos de escisión cerebral que resultaron en lo que parecen ser dos corrientes de conciencia distintas en un mismo cerebro, casos en los que Parfit se basa para defender la idea de que la imposibilidad de muchos de los casos imaginarios que emplea es sólo técnica y para socavar la tesis no reduccionista. En el quinto apartado analizaré la defensa que hace Bernard Williams de la importancia del cuerpo para la identidad personal, importancia que socavaría el criterio psicológico. En el sexto apartado me ocuparé de examinar la afirmación de Parfit de que la identidad personal puede ser, en algunos casos, indeterminada y defenderé que dicha afirmación debe ser llevada aún más lejos. En el séptimo apartado analizaré la idea de Parfit de que la identidad personal no es lo que importa y de que lo que

² Cuando hable de la teoría de Parfit sin concretar la fuente, estaré siguiendo este libro.

importa es la relación a la que el autor se refiere como relación R. En el octavo apartado defenderé la necesidad de realizar dos modificaciones en la relación R. En el apartado noveno analizaré la concepción animalista defendida por autores como Eric Olson, según la cual somos numéricamente idénticos a nuestro organismo y tenemos, por lo tanto, condiciones de persistencia biológicas. El décimo apartado lo dedicaré a examinar la teoría del yo de Daniel Dennett, teoría que es compatible con las afirmaciones que habré defendido en el resto de apartados. Finalmente, en el apartado onceavo, expondré las principales conclusiones obtenidas en este trabajo.

1. La concepción reduccionista de Derek Parfit

Derek Parfit (2004) define la tesis reduccionista como aquella que sostiene que el hecho de que una persona persista a través del tiempo no consiste en otra cosa que en que existan un cuerpo y un cerebro y se produzcan determinados sucesos físicos y mentales relacionados unos con otros. Esta concepción niega, por lo tanto, que la identidad personal consista en algún otro hecho más profundo que vaya más allá de los anteriores. Es fácil darse cuenta de que para que esos hechos puedan constituir la identidad personal, es necesario que no la presupongan. Esto nos lleva a la segunda afirmación que Parfit atribuye a la teoría reduccionista, la de que el mundo puede ser descrito de manera impersonal. Es evidente que si esos hechos no presuponen la existencia de la persona cuya identidad constituyen, ha de ser posible describirlos sin hacer referencia a ninguna persona y, en consecuencia, si estos hechos son todo en lo que consiste la identidad personal, se tendrá que poder llevar a cabo una descripción impersonal del mundo que sea completa, es decir, que, de acuerdo con el reduccionismo, podemos describir completamente el mundo aunque nuestra descripción no diga nada sobre ninguna persona³. Esta última afirmación debe matizarse, pues, como señala el autor, sí que sería necesario hacer referencia a personas al describir el contenido de determinados sucesos mentales, como pensamientos⁴. Lo que sí necesita una descripción del mundo para ser impersonal es evitar afirmar la existencia de personas, es decir, que según la concepción reduccionista tiene que poder describirse completamente el mundo sin necesidad de afirmar que las personas existen. Una teoría de la

³ Como veremos más adelante, Parfit no considera que una persona sea simplemente el conjunto formado por el cerebro, el cuerpo y los sucesos físicos y mentales antes mencionados. El autor considera que la mejor versión del reduccionismo, la que más se acerca a nuestro concepto de persona, es aquella que afirma que somos entidades distintas a esos componentes pero que no existen separadamente de ellos. Podría defenderse que esta versión del reduccionismo no implica que el mundo pueda ser descrito de manera impersonal. Sin embargo, como afirma Parfit, parece plausible suponer que una descripción que afirma la existencia de todo lo que existe puede considerarse completa. Dicho de otro modo, parece sensato pensar que si una entidad no consiste en otra cosa que en la existencia de sus componentes y en el darse de ciertas relaciones entre ellos, la descripción de dichos componentes y de sus relaciones basta para haber descrito completamente tal entidad, aunque esta no sea nombrada en la descripción.

⁴ Si el fisicalismo fuera cierto, sería posible describir los estados mentales en términos puramente físicos y, presumiblemente, estos podrían ser descritos sin necesidad de hacer ningún tipo de referencia a personas, pero la concepción reduccionista de la identidad personal no necesita hacer esta afirmación.

identidad personal que niegue estas dos afirmaciones será no reduccionista. Según la opinión de Parfit, la única alternativa al reduccionismo consistirá en afirmar que la identidad personal consiste en algún tipo de hecho adicional profundo, como la existencia de un alma o un ego cartesiano que persiste a lo largo del tiempo y en cuya persistencia consiste la nuestra. Según el autor, la tesis de que somos algún tipo de entidad que existe separadamente expresa una idea que tenemos muy interiorizada, de modo que, aunque no seamos conscientes de ello, se encuentra presupuesta en muchas de nuestras intuiciones acerca de la identidad personal. Es por ello que Parfit dirige gran parte de sus argumentos contra este tipo de teoría para tratar de mostrar su implausibilidad.

Al tratar el tema de la identidad personal, el autor emplea una serie de experimentos mentales destinados a mostrar que nuestras creencias sobre lo que somos y sobre en qué consiste nuestra permanencia en el tiempo son equivocadas y deben ser revisadas. Aunque todas esas situaciones que imagina resultan imposibles actualmente y muchas de ellas podrían continuar siéndolo siempre, Parfit defiende la legitimidad de su uso aduciendo que dicha imposibilidad es meramente técnica y que tales experimentos mentales están destinados únicamente a sacar a la luz nuestras creencias sobre la identidad personal.

Dentro del reduccionismo, Parfit defiende el criterio psicológico. Este criterio tiene su origen en John Locke. Según Locke (1956), una persona actual es la misma que una persona pasada si recuerda sus acciones y experiencias. Como el propio Parfit señala, esta versión del criterio psicológico es inaceptable. Para empezar, se pierde la transitividad que supuestamente ha de tener una relación de identidad⁵. Además, Parfit (2004) señala que si el criterio de Locke fuera correcto, una persona no podría olvidar ninguna de sus experiencias pasadas, puesto que es el hecho de recordarlas lo que, según este punto de vista, la convierte en la misma persona que las tuvo. Parfit emplea una versión del criterio psicológico que evita estos problemas. En primer lugar, aunque reconoce que la memoria es una de las conexiones psicológicas más importantes por ser la que nos permite ser conscientes de nuestra propia existencia continua a lo largo del tiempo, el autor también da importancia al resto de conexiones psicológicas, como, por ejemplo, la que se da entre la formación de una intención y la acción de llevarla a cabo. Para evitar los problemas a que hemos visto que se enfrentaba la teoría de Locke, Parfit recurre a la noción de continuidad psicológica, que consiste en la existencia de cadenas parcialmente superpuestas de

⁵ Yo podría recordar mis experiencias de hace una semana pero no recordar las de hace un mes. Si resultara que la semana pasada sí que recordaba mis experiencias de hace un mes, nos encontraríamos con que yo soy la misma persona que mi “yo” de la semana pasada, esa persona es la misma que mi “yo” de hace un mes, pero mi “yo” actual no es la misma persona que mi “yo” de hace un mes, con lo cual hemos perdido la transitividad de la relación de identidad.

conexiones psicológicas directas⁶. De este modo, Parfit (2004) define la conexividad psicológica como “el tener lugar de conexiones psicológicas directas y concretas” mientras que la continuidad psicológica consistiría en “el tener lugar de cadenas parcialmente superpuestas de conexividad fuerte” (p. 381). Parece poco plausible la idea de que una sola conexión psicológica pueda bastar para convertir a una persona en numéricamente idéntica a otra persona pasada. Es por ello que el autor habla de conexividad fuerte. Con esto, Parfit pretende expresar que para que haya continuidad psicológica entre dos personas, es necesario que entre ellas exista un número suficiente de conexiones psicológicas directas, es decir, que el grado de conexividad psicológica existente entre ambas personas tiene que ser lo suficientemente elevado⁷. Con estos conceptos, el autor define una relación a la que se refiere como relación R, y que consiste en conexividad psicológica y/o continuidad psicológica con el tipo adecuado de causa⁸. Entonces, Parfit afirma que la identidad personal consiste precisamente en el darse de la relación R en forma no ramificada. Esta última condición de que la relación no tome una forma ramificada la incluye para evitar que puedan darse casos en que se pierda la transitividad de la relación de identidad. En caso de que existieran actualmente varias personas psicológicamente continuas con una persona pasada, es evidente que estas diferentes personas actuales no podrían ser numéricamente idénticas entre sí⁹, con lo cual nos encontraríamos con que una persona pasada sería numéricamente idéntica a dos personas actuales que no lo serían entre sí, perdiéndose así la transitividad de la relación.

Parfit afirma también que la unidad de conciencia y la unidad de las experiencias de una vida concreta no pueden explicarse afirmando que esas experiencias son tenidas por una misma persona. Estas unidades, según el autor, tienen que ser explicadas mediante la descripción de las relaciones que se dan entre dichas experiencias y entre ellas y el cerebro que las hace posibles. Estas experiencias y relaciones, como ya hemos dicho, podrán ser descritas sin hacer referencia a la persona a la que pertenecen. Aunque Parfit considere que las personas no son entidades que

⁶ Utilizando el anterior ejemplo de la memoria, si actualmente una persona A tiene recuerdos de las experiencias de una persona pasada B y esta última tiene recuerdos de las experiencias de una persona C anterior a ella, entonces aunque A no tenga recuerdos de ninguna de las experiencias de C, existirá continuidad de memoria entre C y A aunque no haya entre ellos ninguna conexión directa de memoria.

⁷ El autor considera que “hay conexividad suficiente si el número de conexiones directas, durante cada día, es *como mínimo la mitad* del número que se da, durante cada día, en la vida de casi toda persona real” (Parfit, 2004, p. 382). Aunque esta definición de conexividad fuerte puede parecer algo arbitraria, puesto que, como veremos más adelante, el autor no considera que la identidad personal tenga que ser siempre determinada, podemos entenderla como una simple aproximación. De hecho, el propio autor señala que sería necesario tener en cuenta que unos tipos de conexión psicológica son más importantes que otros.

⁸ En *Razones y personas*, Parfit no termina de ser claro sobre cuál ha de ser la clase correcta de causa. Parece defender que la clase correcta de causa es la causa normal y que si esta no se da, la identidad personal es, en el mejor de los casos, indeterminada. No obstante, en algunas partes del libro el autor parece opinar que la clase correcta de causa puede ser cualquier causa. En cuanto a la causa que ha de tener la relación que importa, Parfit sí que se pronuncia claramente en este último sentido.

⁹ Ello, como afirma el autor, nos llevaría a conclusiones absurdas como la de que una misma persona podría estar en varios lugares a la vez o la de que afirmaciones contradictorias podrían ser ciertas de una misma persona.

existan separadamente, no cree tampoco que seamos simplemente el conjunto de esas experiencias junto con el cerebro y el cuerpo de la persona en cuestión. En su opinión, somos entidades distintas de dichos componentes, aunque no seamos entidades que existan separadamente de ellos. Parfit emplea como ejemplo en muchas ocasiones las naciones y los clubes. Un club no es nada que exista separadamente de sus miembros y las reglas por las que se rige. No obstante, eso no significa que un club sea esos miembros y esas reglas. Del mismo modo, una nación no es tampoco algo que exista separadamente de sus habitantes y su territorio, pero ello no significa que una nación no sea una entidad distinta de esos ciudadanos y ese territorio.

Otra de las afirmaciones que defiende Parfit es que la identidad personal no es siempre determinada. Es decir, que puede haber casos en que la pregunta de si una persona pasada y una persona actual son la misma carezca de respuesta. De entre las afirmaciones que defiende Parfit, esta es probablemente una de las más contrarias a nuestras intuiciones. El hecho de que la identidad personal pueda ser indeterminada implica que hay ciertas situaciones en que no puede responderse a la pregunta de si una persona sobrevivirá a ciertos cambios. Desde este punto de vista, siguiendo uno de los ejemplos que emplea el autor, si yo fuera a sufrir un fuerte dolor tras ciertos cambios físicos y psicológicos, podría no haber respuesta a la pregunta de si seré yo quien sufrirá ese dolor. Estas afirmaciones contradicen claramente nuestras intuiciones. Como señala Parfit, estamos inclinados a creer que tiene que haber siempre una respuesta determinada a la pregunta de si vamos a sobrevivir o a la de si vamos a ser nosotros quienes tengamos una determinada experiencia. Sin embargo, la aparente implausibilidad de tales consecuencias queda atenuada a causa de la más importante afirmación de la teoría de Parfit.

La más importante de las afirmaciones de Derek Parfit es la de que la identidad personal no es lo que importa. Según el autor, lo que importa es que se dé la relación R, con cualquier causa¹⁰. De este modo, aunque en los casos en que la relación se dé en una forma ramificada sería indeterminado si hay identidad o no, sí que sería determinado que se da la relación que importa. Este punto de vista disminuye la implausibilidad de la anterior afirmación de Parfit. Aunque pueda no haber una respuesta a la pregunta de si, tras sufrir ciertos cambios, una persona va a sobrevivir, si se da la relación R esa persona tendrá algo tan bueno como la supervivencia corriente. Del mismo modo, aunque puede ser indeterminado si la persona que va a tener cierta experiencia en el futuro es o no una determinada persona actual, en caso de que se dé la relación R, lo que dicha persona podría esperar es lo mismo que si hubiera identidad. No obstante, la

¹⁰ Como veremos, Parfit también atribuye cierta importancia a las relaciones de continuidad y similitud física. Sin embargo, defenderé que la atribución de importancia a estas relaciones se basa en un error por parte del autor.

relación R es claramente una cuestión de grado, lo cual tiene ciertas consecuencias que examinaremos más adelante.

A continuación examinaremos los argumentos que emplea el autor para defender estas afirmaciones, así como las críticas que han recibido. Trataré de mostrar que muchas de las críticas que ha recibido la concepción reduccionista de Parfit son equivocadas, pero defenderé también que algunas de sus afirmaciones necesitan ser revisadas. Empezaremos centrándonos en su afirmación sobre la posibilidad de una descripción impersonal del mundo.

2. La objeción de la circularidad y la posibilidad de una descripción impersonal del mundo

Uno de los principales argumentos que suelen emplearse contra el criterio psicológico de identidad personal es la objeción de la circularidad¹¹. Según este argumento, atribuido normalmente a Joseph Butler, no es posible definir el concepto de identidad personal utilizando el concepto de memoria¹², puesto que el primero es epistemológicamente previo al segundo. Según este argumento, para poder verificar que alguien efectivamente recuerda una experiencia, necesitamos comprobar que fue él quien de hecho tuvo esa experiencia. Ello se debe a que forma parte del concepto de memoria el que uno sólo puede recordar experiencias que él mismo ha tenido. De este modo, la conexividad psicológica presupondría la identidad personal y, en consecuencia, lo mismo sucedería con la continuidad psicológica. Por lo tanto, si la conexividad se define utilizando conexiones psicológicas como la memoria, la relación R, como Parfit admite, presupondrá la identidad personal y no podrá formar parte de su definición.

Para resolver este problema, Parfit recurre al concepto, introducido originalmente por Shoemaker (1970), de cuasi-memoria. Este concepto es el mismo que el de memoria pero eliminando la restricción según la cual uno puede recordar únicamente las experiencias que él mismo ha tenido. Para tener un cuasi-recuerdo de una experiencia es necesario tener un recuerdo aparente de dicha experiencia, que esa experiencia haya tenido efectivamente lugar y que entre la experiencia y el recuerdo de ella exista una relación causal del tipo correcto¹³. De este modo, sería en principio posible cuasi-recordar las experiencias pasadas de otras personas y lo que llamamos memoria no sería otra cosa que cuasi-memoria de nuestras propias experiencias. Parfit aplica la misma estrategia al resto de conexiones psicológicas. Recurre, por ejemplo, a la noción de cuasi-intención. A diferencia de lo que presumiblemente sucede con las

¹¹ Al exponer la objeción de la circularidad sigo a Harold Noonan (1989).

¹² Aunque me refiero a la memoria, la objeción es en principio aplicable a los demás tipos de conexión psicológica.

¹³ Al principio, Parfit exige que la experiencia haya sido tenida por alguien, pero más adelante modifica la definición de cuasi-memoria para darle esta forma más impersonal.

intenciones, una cuasi-intención no necesita ser formada por la misma persona que la realizará. Dicho de otro modo, una persona puede formarse la cuasi-intención de que otra persona actúe de un modo determinado. Puede parecer extraño que se sugiera la posibilidad de “recordar” experiencias pasadas de otras personas, pero no parece que haya ninguna imposibilidad lógica en ello. Como Parfit señala, nuestros recuerdos de experiencias pasadas nos vienen junto con la creencia de que fuimos nosotros quienes vivimos tales experiencias. No obstante, como argumenta el autor, esto sucede porque estamos acostumbrados a recordar únicamente nuestras propias experiencias. Si sucediera a menudo que una persona cuasi-recordara experiencias ajenas, esa creencia dejaría de acompañar a los cuasi-recuerdos. En una situación así, como afirma el autor, no podríamos dar por sentado que fuimos nosotros quienes experimentamos aquello que parecemos recordar a no ser que contáramos con evidencia adicional de ello. De este modo, Parfit puede definir la conexividad psicológica, y por tanto también la continuidad psicológica y la relación R, utilizando el concepto de cuasi-memoria, junto con los del resto de cuasi-conexiones psicológicas, y evitar así la circularidad en que parecía incurrir el criterio psicológico.

Como ya hemos dicho, Parfit opina que la descripción de estas diferentes conexiones psicológicas¹⁴ y de las diferentes experiencias entre las que se dan puede llevarse a cabo sin afirmar que son tenidas por una persona. El autor considera una objeción a la posibilidad de una descripción impersonal planteada por Bernard Williams¹⁵. Según esta objeción, como señala Parfit, en caso de que se describan varios pensamientos será necesario aclarar de algún modo si estos tienen lugar en una misma vida o no. Aunque ello parece implicar la necesidad de emplear nombres personales, el propio Williams sugiere la posibilidad de que puedan encontrarse expresiones que resuelvan el problema sin tener que recurrir a sujetos de experiencias. Parfit (2004) sugiere que dos ejemplos de estas expresiones podrían ser “en la vida concreta que contiene el pensar del pensamiento que se expresa por la preferencia de esta sentencia, se piensa:...” y “en la vida concreta que ahora es causalmente dependiente del cuerpo A, se piensa:...” (p. 413). De este modo, como afirma Parfit, si la unidad de la vida de una persona y las conexiones psicológicas y sus causas pueden describirse de manera impersonal, será posible una descripción impersonal completa del mundo.

Centrémonos primero en ver cómo puede explicarse la unidad de conciencia en un momento determinado sin presuponer la identidad personal. Parfit considera que la unidad de conciencia se debe simplemente a la existencia de un estado de conciencia de múltiples experiencias. Lo

¹⁴ A partir de ahora, cuando hable de conexiones psicológicas, salvo que indique lo contrario, me estaré refiriendo a cuasi-conexiones psicológicas.

¹⁵ Como señala Parfit, Williams dirige esta objeción a Lichtenberg. Este último argumentó contra el *Cogito* de Descartes que este suponía la existencia del pensador y que debía ser substituido por una expresión impersonal.

que separa las experiencias conscientes que son tenidas en este momento por mí y las que son tenidas por otros es que existe un único estado de conciencia de todas mis experiencias. A esas experiencias puedo llamarlas mis experiencias conscientes presentes porque me encuentro en un estado mental que consiste precisamente en ser consciente de todas esas experiencias. Al decir que “me encuentro” en ese estado mental todavía estamos presuponiendo la identidad personal, pero la expresión puede reformularse fácilmente de modo que eso no suceda. Lo que unifica todas mis experiencias conscientes¹⁶ presentes es la existencia de una única experiencia o suceso mental que consiste precisamente en conciencia de todas esas experiencias. Para Parfit no hay ningún hecho profundo que explique la unidad de conciencia en un momento dado, simplemente hay conciencia de todas esas experiencias y eso es lo único en lo que consiste su unidad. Es un hecho que lo que todas esas experiencias tienen en común es que somos conscientes de ellas, por lo que la explicación de la unidad de conciencia que nos ofrece el reduccionismo resulta perfectamente plausible.

No hay mayor dificultad en explicar, sin presuponer tampoco la identidad personal, la unidad de las experiencias que tienen lugar dentro de una vida concreta. Como Parfit (2004) afirma, para explicar esta segunda unidad nos bastaría con “describir lo que, en momentos diferentes, se pensó, se sintió, se observó y se hizo, y la manera en que estos diversos sucesos estuvieron interrelacionados” (p. 451). Esas interrelaciones pueden describirse de manera impersonal recurriendo al concepto de cuasi-memoria y al resto de cuasi-conexiones psicológicas. Quedan todavía las relaciones causales que forman parte de las distintas conexiones psicológicas, pero tampoco parece haber problema en explicarlas de manera impersonal. Parece plausible suponer que el proceso causal que lleva a que se recuerde una determinada experiencia puede explicarse recurriendo a sucesos que tienen lugar en el cerebro, y no hay duda de que los procesos neurológicos pueden explicarse sin necesidad de presuponer la identidad personal. Parece sensato pensar que lo mismo sucede con el resto de conexiones psicológicas. Podríamos describir, por ejemplo, el suceso mental de formarse una intención y el de realizar dicha intención sin recurrir a la identidad personal. Puede parecer que para explicar la relación causal entre la formación de la intención y su realización tendríamos que afirmar que la causa de que la persona realice la intención es que fue ella quien se la formó, pero no necesitamos recurrir a la identidad personal de ese modo. Podríamos sencillamente explicar esa relación causal apelando, en caso de que la causa sea la habitual, a la continuidad física de cierto cerebro, cerebro que estaba relacionado de cierto modo con el suceso mental de formarse la intención y que está

¹⁶ Hablo de experiencias conscientes porque mis experiencias no conscientes, suponiendo que podamos llamar experiencias a los sucesos mentales no conscientes, no están unificadas de ese modo. A estas últimas también podemos llamarlas mías, pero qué es lo que las hace mías no puede explicarse, claro está, recurriendo a ningún estado de conciencia. La explicación de lo que hace mías a esos sucesos mentales consiste en la descripción de su relación con otros sucesos mentales, incluyendo su relación con aquellos sucesos mentales que sí sean conscientes.

relacionado de cierto modo con el suceso mental de realizarla¹⁷. Parfit considera que incluso las verdades subjetivas pueden ser descritas de manera impersonal. Un pensamiento puede, en su contenido, referirse a sí mismo sin que ello implique de ningún modo la existencia de un sujeto de experiencias. En opinión de Parfit, el concepto “yo”, del mismo modo que el resto de conceptos índicecos, puede explicarse recurriendo únicamente al uso autorreferencial del concepto “este”.

La afirmación del reduccionismo de que es posible una descripción impersonal del mundo ha sido objeto de múltiples críticas. Rupert Read (2015) argumenta que tiene un gran valor para nosotros ser un individuo y ser, concretamente, el individuo que somos. Según Read, una descripción impersonal del mundo omitiría eso. Tal descripción, según el autor, no incluye ese valor. Lo mismo podría pensarse, tal vez, del valor que tienen para nosotros otras personas precisamente por ser las personas que son. Podría argumentarse, como parece ser la intención de Read, que aunque la concepción reduccionista de Parfit no deje sin explicar ningún suceso mental ni físico, sí que omite precisamente ese valor, es decir, que explica el mundo de una manera que no puede incluir ese valor. Sin embargo, este argumento, según como lo interpretemos, o malentende la teoría de Parfit o se basa en suposiciones muy poco plausibles. El valor del que habla Read sí que estaría incluido en la descripción impersonal que propone Parfit. El valor que tiene para nosotros nuestra individualidad estaría incluido en la descripción del contenido de algunos pensamientos, y la descripción de las relaciones que existan entre estos pensamientos y otros sucesos mentales podría explicar la influencia que ese valor ejerce sobre nosotros. Lo mismo puede decirse, por supuesto, del valor que tenga para nosotros la individualidad de otras personas. Como afirma Parfit en múltiples ocasiones, los contenidos de los pensamientos pueden contener referencias a personas, incluyéndonos a nosotros mismos, sin que ello presuponga la existencia de ningún sujeto de experiencias. Tal vez Read crea que ese valor no puede estar sólo en el contenido de nuestras experiencias, sino que es en cierto modo objetivo. Si ese es el caso y el argumento de Read se basa en esa suposición, entonces se basa en una suposición muy implausible. El valor que tiene nuestra individualidad o la de otros para nosotros es claramente subjetivo y es natural esperar que ese valor esté únicamente en el contenido de nuestros estados mentales. Puede que Read crea que nuestra individualidad tiene valor no sólo para nosotros, tal vez crea que tiene valor objetivamente hablando. En mi opinión, dicha afirmación necesitaría argumentos en su favor, argumentos que Read, seguramente debido a que no es eso lo que pretende afirmar, no ofrece.

Ingmar Persson (2016) también argumenta contra la posibilidad de una descripción impersonal. En su opinión, “los reduccionistas no necesitan – e incluso no pueden – eliminar a las personas

¹⁷ Este tipo de explicación no presupone el fisicalismo. Podría defenderse, por ejemplo, un dualismo de propiedades como el de David Chalmers y dar, sin embargo, una explicación de este tipo.

[...] del rol de sujetos de la predicación mental¹⁸ (Persson, 2016, p. 158). Según Persson, suponer que podemos describir los sucesos físicos y mentales sin mencionar a las personas es entender a las personas como si fueran entidades que existen separadamente de sus cuerpos y cerebros. El autor afirma que, puesto que los estados mentales no pueden existir sin los estados cerebrales, los primeros son no derivativamente atribuibles a los últimos. Según Persson, por lo tanto, los reduccionistas deberían entender a las personas como sujetos de los estados mentales. Este argumento se debe, de nuevo, a una mala comprensión del reduccionismo. Precisamente porque el reduccionismo no trata a las personas como si fueran entidades que existan separadamente, necesita afirmar que es posible una descripción impersonal del mundo. Si una persona no consiste en nada más que en una serie de sucesos físicos y mentales y en la existencia de un cerebro y un cuerpo, describiendo estas cosas estaremos describiendo todo lo que hay aunque no mencionemos a las personas. Esto no impide que al describir esas cosas estemos implicando la existencia de personas, la descripción no deja de ser impersonal por ello, ya que es impersonal en la medida en que no contiene referencias a personas. Es importante señalar también que defender la posibilidad de una descripción impersonal del mundo no significa necesariamente negar la existencia de sujetos de experiencias. Parfit afirma en varias ocasiones que aunque sea posible una descripción impersonal, la versión del reduccionismo que él defiende no niega que las personas existamos ni que seamos sujetos de experiencias¹⁹.

Otro argumento contra esta posibilidad de una descripción impersonal es el ofrecido por Robert Ellis (2000). Ellis cree que para que fuera posible una descripción de ese tipo sería necesario que hubiera una correlación uno-a-uno absoluta entre los símbolos del lenguaje y la realidad. El autor argumenta que tal relación entre los símbolos del lenguaje y la realidad no tiene nada que ver con ningún lenguaje concebible. Sin embargo, este requisito no es realmente necesario para que una descripción impersonal sea posible. La idea de que este es un requisito necesario se debe a una mala interpretación de lo que Parfit quiere decir cuando habla de una descripción impersonal. Puede ayudarnos a entender cuál es el error de Ellis examinar el ejemplo que él emplea. El autor imagina a un neurocirujano describiendo los procesos cerebrales de Descartes de manera aparentemente impersonal. Ellis ve dos problemas en tal descripción. En primer lugar, aunque no se presupusiera la identidad de Descartes, se estarían empleando sujetos y predicados que presupondrían la identidad de unidades más pequeñas, como, por ejemplo, la del cerebro de Descartes. Además, Ellis argumenta que dicha descripción tendría que presuponer que Descartes y sus procesos cerebrales no forman parte del hablante, es decir, que son distintos a él. La primera de las dos observaciones del autor se basa en un error evidente. No hay problema en que la descripción presuponga la identidad de unidades más pequeñas, lo único que

¹⁸ La traducción es mía.

¹⁹ Yo, en cambio, defenderé que, aunque las personas sí que existimos, no somos sujetos de experiencias y que no existe nada real a lo que podamos referirnos en sentido estricto como sujeto de experiencias.

no puede presuponer si ha de seguir siendo impersonal es la identidad personal. Una explicación impersonal no necesita evitar hablar de identidad, sólo tiene que evitar hablar, cuando no esté describiendo el contenido de ningún pensamiento, de la identidad de las personas²⁰. El propio Parfit, al explicar cómo sería una descripción impersonal, recurre, por ejemplo, a la identidad del cerebro. La otra observación de Ellis, en cambio, es más interesante. Lo que el autor parece querer argumentar con esta segunda afirmación es que toda descripción presupone, en cierto modo, la identidad personal del hablante. Puede que Ellis tenga razón en este punto. Puede que la manera en que funcionan nuestra mente y nuestro lenguaje no nos permita referirnos al mundo sin presuponer nuestra propia existencia como una entidad separada de este. Personalmente, creo que esa opinión es errónea pero, en cualquier caso, no es incompatible con la posibilidad de una descripción impersonal del mundo que postula la concepción reduccionista. Podría ser que yo esté equivocado y Ellis tenga razón, pero ello no pondría en cuestión las afirmaciones de Parfit. Ellis está hablando de cómo funciona nuestra mente o de cómo funciona nuestro lenguaje. La concepción reduccionista de Parfit, en cambio, se limita a hacer afirmaciones sobre qué es lo que somos y sobre en qué consiste nuestra identidad a través del tiempo. Parfit no necesita afirmar que una persona puede de hecho dar una descripción completa del mundo en términos impersonales²¹, lo único que necesita afirmar es que una porción del mundo que contenga personas puede ser descrita, desde fuera de dicha porción del mundo, sin presuponer la identidad personal de ninguna de esas personas. Puede que la descripción no pueda evitar presuponer la identidad personal del hablante, pero, si eso fuera cierto, se trataría de un hecho sobre nuestro lenguaje o sobre el funcionamiento de nuestra mente, no de un hecho sobre lo que somos ni sobre aquello en que consiste nuestra identidad a través del tiempo. Podría defenderse que a todo hablante, por el hecho de formar parte del mundo, le es imposible realizar una descripción completa, incluso aunque no sea impersonal, del mundo, pero esto de ningún modo podría decirnos nada sobre la identidad personal, y lo mismo sucede con el argumento de Ellis.

Un argumento distinto contra la posibilidad de describir el mundo en términos impersonales es el que utiliza Lloyd Fields (1987). La objeción de Fields se dirige contra la posibilidad de definir una de las conexiones psicológicas de manera que no presuponga identidad personal. La conexión psicológica contra la que el autor dirige su argumento es la cuasi-intención. Fields imagina el caso de que una persona A pudiera leer la mente de otra persona B. Si por cualquier motivo A se sintiera siempre obligado a hacer lo que quisiera B, este último podría lograr que A

²⁰ Tal vez podría defenderse la posibilidad de una descripción completa del mundo que sólo presupusiera la identidad de partículas fundamentales, pero la concepción reduccionista de la identidad personal no necesita en absoluto realizar una afirmación tan fuerte.

²¹ Puede que Parfit sí que pretenda afirmar que una descripción impersonal es de hecho posible, pero en realidad no necesita hacer una afirmación tan fuerte.

actuara formándose las correspondientes intenciones. El autor argumenta que la definición que ofrece Parfit no permite distinguir un caso como este de un caso de cuasi-intención. Sin embargo, aunque puede que la definición de Parfit no termine de ser adecuada, hay una diferencia clara entre un caso y el otro. En un caso de cuasi-intención, la persona actuaría tras encontrar la intención en su mente, sintiendo la intención como propia. Aunque la diferencia entre los dos casos es muy difícil de precisar verbalmente, está claro que se trata de experiencias distintas, y todos conocemos esa diferencia. Todos somos capaces de concebir la diferencia entre cumplir los deseos de una persona tras leer su mente y realizar una intención. Que esa diferencia sea difícil de expresar en palabras no significa que no exista.

Brian Garrett (1991) ofrece otro argumento en contra de la posibilidad de realizar una descripción completa del mundo sin mencionar la existencia de personas concretas. El autor supone que en un mundo posible se realizara el trasplante de ambos hemisferios cerebrales y que en otro el de sólo uno de los dos. En este último, la persona resultante sería la misma que la original, mientras que en el primero no. Los hechos sobre dicha persona serían los mismos en ambos mundos, pero la identidad de la persona sería distinta. Sin embargo, lo necesario para que sea posible la descripción impersonal es que si todos los hechos sobre el mundo son los mismos, lo sean también todos los hechos sobre la identidad, y, en el caso que imagina Garrett, los hechos sobre ambos mundos son distintos. El autor contempla la posibilidad de que pudiera responderse de este modo y emplea otro argumento. Supongamos un mundo posible W_1 en que un objeto X está formado por las partes a, b y c . Supongamos otro mundo posible W_2 en que un objeto Y , que no es numéricamente idéntico a X , está formado por las partes a, b y d . Entonces el autor argumenta que si un objeto hubiera sido distinto en una sola de sus partes no por ello hubiera dejado de ser el mismo objeto²², de modo que podemos suponer otro mundo posible W_3 en que X estuviera formado por a, c y d y otro mundo posible W_4 en que Y estuviera hecho de a, c y d . En los mundos W_3 y W_4 todos los hechos pueden ser exactamente los mismos, pero en uno el objeto formado por a, c y d es X mientras que en el otro es Y . Por lo tanto, la identidad de esos objetos, y análogamente la de las personas, es un hecho que no puede reducirse a otros hechos más básicos del mundo y, en consecuencia, no es posible una descripción impersonal completa del mundo. Sin embargo, en cada uno de ambos mundos habría el mismo número de personas. En cada uno de ambos mundos, la persistencia de esas personas se daría del mismo modo, es decir, habría el mismo número de personas y estas trazarían las mismas rutas a través del tiempo. Los juicios de identidad que las personas de ambos mundos posibles llevarían a cabo serían los mismos. Las personas de uno y otro mundo que fueran cualitativamente idénticas entre sí tendrían el mismo nombre en ambos mundos, trazarían la misma ruta a través

²² En el caso de las personas podemos suponer una diferencia en una sola molécula o en una sola partícula.

del tiempo y los juicios de identidad que las personas de sus mundos harían sobre ellas serían los mismos. Las dos personas en cuestión que, a pesar de ser cualitativamente idénticas en ambos mundos, son distintas personas lo son sólo para nosotros, y lo son debido al punto de vista que adoptamos para describir cada mundo. A uno de ellos hemos llegado variando las características de una persona de cierto modo, mientras que al otro hemos llegado variando las características de otra persona de modo que el resultado sea el mismo. Cuando hacemos juicios de identidad entre mundos posibles, el resultado de dichos juicios depende del modo en que hemos empleado nuestros conceptos para llegar a dicho mundo. Aunque ambos mundos son idénticos, hemos llegado a cada uno de manera diferente, hay en cada uno una persona a la que hemos llegado utilizando el concepto de una persona distinta en cada caso. Nada de esto impide que sea posible una descripción impersonal del mundo. Si la concepción reduccionista es cierta, sencillamente ambos mundos posibles son en realidad el mismo. Simplemente los describimos de manera distinta en cada uno de ambos casos. La manera en que describimos el mundo posible en cuestión depende del modo en que nos hemos formado la idea de este. El reduccionismo es una teoría de la identidad personal a través del tiempo y no necesita hacer ninguna afirmación sobre cómo se realizan los juicios de identidad entre mundos posibles. A qué persona de nuestro mundo es idéntica cierta persona de cierto mundo posible no es un hecho sobre dicho mundo, es más bien un hecho sobre cómo describimos ese mundo. Para demostrar que no es posible una descripción impersonal del mundo no nos sirve de nada comparar distintos mundos posibles, sino que sería necesario mostrar, considerando un único mundo de manera aislada, que al describir impersonalmente dicho mundo no lo estamos describiendo de manera completa. Lo único que el argumento de Garrett demuestra es que no podemos realizar una descripción impersonal completa de la relación entre distintos mundos posibles, lo cual no tiene nada de misterioso, puesto que tenemos conceptos de las distintas personas y nuestros conceptos tienen una gran importancia en la descripción de dichas relaciones.

Parfit, además de mostrar cómo podría ser posible una descripción impersonal del mundo, ofrece argumentos positivos para tratar de demostrar que esa descripción sería efectivamente correcta. Puesto que algunos de estos argumentos le sirven también para socavar la idea de que somos entidades que existen separadamente, puede ser conveniente que antes de examinarlos analicemos la argumentación que realiza Parfit contra la tesis no reduccionista. En la siguiente sección analizaremos, pues, algunos de los argumentos que el autor emplea contra dicha tesis.

3. Refutando la tesis no reduccionista

Parfit considera que muchas de nuestras intuiciones con respecto a la identidad personal se deben a que creemos que somos entidades que existen separadamente²³. Es porque creemos que somos entidades de este tipo por lo que nuestra intuición nos dice que la identidad personal tiene que ser siempre determinada. La idea de que la unidad de conciencia y la unidad de las experiencias que tienen lugar en una vida concreta se explican atribuyendo dichas experiencias a la persona en cuestión también deriva de la tesis de que somos entidades que existen separadamente. Contra esta tesis, Parfit dirige algunos de sus argumentos. El autor objeta, empleando un argumento utilizado por Locke y Kant, que, incluso aunque tales entidades existieran y fuéramos directamente conscientes de su existencia, cosa que Parfit niega, no podríamos estar seguros de su persistencia. Dicha entidad podría ser substituida por otra que fuera psicológicamente continua con ella y no tendríamos ningún modo de darnos cuenta de ello. De hecho, podría existir una entidad distinta en cada instante y, siempre que se mantuviera la continuidad psicológica, permaneceríamos ignorantes de ello. Como afirma Parfit (2004), la conciencia de la existencia de entidades existiendo separadamente “no puede de hecho distinguirse de nuestra conciencia de la mera continuidad psicológica” (p. 410). Por lo tanto, no hay nada en nuestras experiencias que deba hacernos pensar que somos entidades que existen separadamente de nuestros cuerpos, nuestros cerebros y los sucesos físicos y mentales relacionados con ellos.

Parfit es consciente de que este argumento no demuestra que no seamos sujetos de experiencias que existen separadamente, puesto que si hubiera algún indicio de ello o si la existencia de tales entidades estuviera implicada por alguna de nuestras experiencias, tendríamos motivos para creer en su existencia. El autor, no obstante, considera que no tenemos tales motivos para pensar que seamos entidades de ese tipo. Parfit analiza, por ejemplo, el argumento de Descartes al respecto. Como señala Parfit, según la opinión de Descartes nuestras experiencias implican la existencia de entidades que existen separadamente. Con su famoso “pienso, luego existo”, Descartes creía estar demostrando la existencia de una substancia espiritual que sería el sujeto de dicho pensamiento. Como Parfit señala, además de suponer injustificadamente que un pensador tenía que ser un ego puro, Descartes cometió el error de dar por sentado que un pensamiento tenía que tener necesariamente un pensador. Parfit emplea contra Descartes el mismo argumento que Lichtenberg le dirigió, es decir, señala que la frase “pienso, luego existo” es engañosa y que debe ser substituida por otra como “se piensa: el pensamiento tiene lugar”²⁴. Parfit no se opone a la opinión de que los pensadores existen, pero si se reformula el *Cogito* de

²³ Esas entidades serían almas o egos cartesianos, es decir, sustancias puramente mentales.

²⁴ Ya hemos visto en la sección anterior la objeción de Williams a esta manera de reformular la frase de Descartes y las frases con que Parfit propone sustituirla.

Descartes de manera que no resulte engañoso, es fácil ver que no hay nada en él que demuestre que los pensadores sean entidades que existen separadamente.

Parfit no cree que la tesis no reduccionista sea inconcebible ni lógicamente imposible²⁵. En su opinión, dicha tesis podría haber resultado verdadera. Podría haber habido, según el autor, evidencia en favor de la teoría de que somos entidades que existen separadamente. Parfit pone algunos ejemplos de qué tipo de sucesos podrían considerarse, en caso de que se dieran, como evidencia en favor de la concepción no reduccionista. El autor considera que hubiéramos tenido motivos para creer que somos egos cartesianos si hubiera habido evidencia en favor de la reencarnación. Si hubiera resultado que la gente tuviera habitualmente cuasi-recuerdos de las vidas pasadas de personas que existieron antes que ellos y si la veracidad de esos recuerdos hubiera sido contrastada, por ejemplo, mediante hallazgos arqueológicos, seguramente la mejor explicación hubiera consistido en afirmar que un mismo ego cartesiano había sido el sujeto de las experiencias de ambas personas. Si además hubiera resultado que las enfermedades mentales o neurológicas no pudieran debilitar la continuidad psicológica y que esta última sólo pudiera, en tales casos, permanecer intacta o ser eliminada por completo, hubiéramos tenido evidencia en favor de que la identidad personal es siempre determinada y, por lo tanto, en favor de que somos un ego cartesiano que es “todo o nada”. Sin embargo, como señala Parfit, no tenemos evidencia de este tipo. Es más, el hecho de que en el segundo caso suceda justo lo contrario es un poderoso argumento en contra de la existencia de egos cartesianos.

Además de mostrar, como hemos visto, que no tenemos ningún tipo de evidencia en favor de la tesis no reduccionista, Parfit trata de mostrar que sí que tenemos mucha evidencia en contra de tal teoría. Como el autor afirma, además del hecho de que la continuidad psicológica sea una cuestión de grado, contamos con mucha evidencia en favor de la idea de que es el cerebro el portador de dicha continuidad psicológica. Parfit recurre, además, a ciertos experimentos mentales que pretenden mostrar la implausibilidad de la tesis no reduccionista. Aunque dichos experimentos mentales son técnicamente imposibles, no son lógicamente imposibles. Además, estos experimentos mentales se apoyan en ciertos hechos que han sucedido realmente y que ofrecen, por sí solos, una gran evidencia en contra de la idea de que somos egos cartesianos. Esos hechos y esos experimentos mentales ofrecen además evidencia en favor de la posibilidad,

²⁵ Bernard Williams, en cambio, sí que parece creer que la existencia de este tipo de entidad es inconcebible. Williams (1986c) señala que si dijéramos que yo podría no haber existido y que en mi lugar podría haber existido alguien que fuera exactamente como yo, no está claro que estaríamos diciendo. El autor afirma que si lo que se supone que se perdería en ese caso es mi centro cartesiano de conciencia, no está nada claro qué es lo que se estaría perdiendo. El autor también argumenta que si decimos que yo podría haber sido Napoleón, no podemos decir qué sería distinto en el mundo por el hecho de que yo fuera él. Según Williams, el hecho de que podamos imaginar esta clase de casos se debe a que experimentamos una ilusión. Como señala el autor, si yo me imagino siendo Napoleón, sólo me estoy imaginando interpretando el papel de Napoleón, cosa que no implica que sea una posibilidad lógica el que yo pudiera haber sido Napoleón.

que hemos examinado en la sección anterior, de una descripción impersonal del mundo. Los hechos a los que me refiero son ciertos casos de escisión cerebral en que se seccionó el cuerpo calloso que une ambos hemisferios. En la siguiente sección trataremos esos hechos y los experimentos mentales que Parfit construye a partir de ellos. Examinaré algunas objeciones de acuerdo con las cuales la interpretación que hace Parfit de tales casos es errónea y trataré de defender que dichos argumentos son equivocados y que las conclusiones que el autor extrae de tales situaciones se siguen realmente de ellas.

4. Cerebros divididos

Los primeros experimentos de bisección cerebral fueron llevados a cabo en gatos y monos (*vid.* Sperry, 1966). Sin embargo, dicha operación empezó, posteriormente, a realizarse en humanos con el objetivo de tratar casos de epilepsia severa que no respondían a otros tratamientos. Como señala Sperry (1966), en estas operaciones, para tratar de confinar los ataques a un único hemisferio, se seccionó el cuerpo calloso que permite la comunicación entre ambos²⁶. La operación resultó ser muy efectiva para tratar dichos casos de epilepsia, pero en situaciones experimentales con los pacientes se observó que había tendido efectos inesperados. Lo que los experimentadores observaron, según afirma Sperry, parece indicar la existencia de dos esferas de conciencia separadas, es decir, de dos mentes distintas. Como señala el autor, ninguno de los dos hemisferios parece ser consciente de lo que sucede en el otro.

Antes de examinar los experimentos que se han realizado en casos como estos y sus resultados, puede ser de utilidad conocer las capacidades de cada uno de los dos hemisferios cerebrales. El hemisferio izquierdo, también conocido como hemisferio mayor o dominante, es el que posee las capacidades verbales. Este hemisferio es capaz de hablar, manteniendo una conversación normal y puede también realizar otras actividades que involucran expresión por medio del lenguaje, como, por ejemplo, escribir. El hemisferio derecho, también conocido como hemisferio menor, no es capaz de hablar pero puede expresarse de manera no verbal (*vid.* Sperry, 1966). Tanto la cabeza como el cuello están conectados a ambos hemisferios, de modo que ambos pueden controlar esas zonas del cuerpo y recibir sensaciones de ellas. El tacto de cada una de las mitades del resto de zonas del cuerpo se encuentra relacionado con el hemisferio del lado contrario, de modo que al hemisferio izquierdo le llegan las sensaciones de la mitad derecha y al derecho las de la mitad izquierda. La mitad de cada retina está conectada con el hemisferio de su mismo lado, de modo que, puesto que la mitad izquierda de cada retina escanea la mitad derecha del campo visual y viceversa, cada hemisferio recibe la información correspondiente a la mitad contraria del campo visual. Cada oído transmite, en cierto grado, los

²⁶ Como señala Sperry (1966), también se seccionaron las comisuras anterior e hipocampal y la masa intermedia.

sonidos a ambos hemisferios. Por último, en lo referente al olfato sucede lo contrario que en los casos anteriores, cada lado envía las señales al hemisferio que se encuentra en dicho lado (*vid. Nagel, 1979*).

En los diferentes experimentos que se han realizado con los pacientes de este tipo de intervención, se ha procedido de manera que se pudiera hacer llegar determinada información a un único hemisferio y se ha juzgado la reacción de los dos hemisferios teniendo en cuenta las capacidades que cada uno de ellos controla. Sperry (1966) nos cuenta los resultados de un experimento en que se le tapaban los ojos a los pacientes y se les colocaban objetos en la mano izquierda, que es la que está conectada con el hemisferio menor. Aunque dicho hemisferio es incapaz de comunicarse verbalmente, los experimentadores podían cerciorarse de que reconocía de qué objeto se trataba pidiéndole que mostrara de qué modo había que utilizarlo o pidiéndole que después lo seleccionara, empleando el tacto o la vista, de entre un grupo de objetos. Como señala el autor, el sujeto no era capaz de expresar verbalmente de qué objeto se trataba. No obstante, en cuanto el hemisferio dominante obtenía información sobre cuál era el objeto en cuestión, por tocarlo con la otra mano o a causa de algún ruido característico, pasaba a ser capaz de responder correctamente. Obviamente, el experimento parece indicar que sólo el hemisferio derecho era consciente de la información enviada por las partes del cuerpo a que él estaba conectado. El autor reporta resultados semejantes en lo referente a la visión. Los sujetos no parecen ser capaces de expresar verbalmente la información de la mitad izquierda del campo visual, que sólo es recibida por el hemisferio menor. Sperry describe, por ejemplo, un experimento en que se mostró a los pacientes un color distinto en cada mitad del campo visual y se les preguntó si veían el mismo color en ambas mitades. El autor nos cuenta que los sujetos no fueron capaces de responder correctamente. No obstante, si ambos colores se muestran en una misma mitad del campo visual, los pacientes no tienen ningún problema en detectar que se trata de colores diferentes. Thomas Nagel (1979) describe también experimentos con olores llevados a cabo en este tipo de pacientes. El autor nos cuenta que cuando se enviaba un olor característico al hemisferio menor, el sujeto era capaz de seleccionar con la mano izquierda el objeto a que correspondía. No obstante, cuando se le pedía que lo hiciera, el paciente, a pesar de hacerlo, se quejaba de que no podía seleccionar el objeto porque no había olido nada. Nagel nos dice que si el olor que se le hacía llegar al hemisferio menor era desagradable, el paciente hacía gestos y exclamaciones guturales de disgusto a la vez que decía que no estaba oliendo nada.

El hemisferio menor, además de mostrarse capaz de manipular conceptos y realizar tareas sencillas que requieren ciertas capacidades mentales, parece ser capaz de mostrar emociones. En primer lugar, ya hemos visto que es capaz de mostrar desagrado cuando se le hace llegar un olor desagradable. Nagel (1979) nos cuenta que cuando en estos experimentos se coloca un objeto en

la mano izquierda del paciente, se le pregunta de qué objeto se trata y el sujeto responde erróneamente, el hemisferio menor suele provocar que frunza el ceño. En cambio, cuando la respuesta expresada es la correcta el sujeto suele sonreír. Estas reacciones se producen sin que los experimentadores hayan comunicado al sujeto si su respuesta era correcta o no, de modo que la única explicación es que el hemisferio menor se da cuenta del acierto o del error y reacciona en consecuencia. Lishman (1971) nos informa de que cuando el sujeto muestra una reacción emocional causada por un estímulo que solo llega al hemisferio menor, es incapaz de explicar cuál es la causa de su reacción.

También se han reportado situaciones que parecen indicar la existencia de conflictos entre la voluntad de ambos hemisferios. Sperry (1966) nos informa de que un paciente comunicó que cuando, al vestirse, se subía los pantalones, la mano izquierda trataba de bajárselos. El autor nos explica también que un paciente se quejó de que cuando intentaba que su mujer fuera con él para ayudarle en algo, la mano izquierda reaccionaba apartándola violentamente con un empujón. Nagel (1979) nos informa de una situación experimental en que ambos hemisferios entraron también en conflicto. A uno de estos pacientes se le colocó una pipa (“pipe”) en la mano izquierda y, tras retirársela, se le pidió que escribiera con la mano izquierda el nombre del objeto que había estado sosteniendo. El sujeto empezó escribiendo “PI” con mucha dificultad. Entonces, parece ser que el hemisferio dominante tomó el control y dedujo, a partir de lo que había escrito hasta el momento, que el objeto era un lápiz (“pencil”). El paciente empezó a escribir con mayor agilidad, convirtió la “I” en una “E” y escribió “PENCIL”. Tras ello, el hemisferio menor recuperó el control de nuevo, tachó las letras “ENCIL” e hizo un dibujo torpe de una pipa. Este caso, además de mostrar claramente que sólo el hemisferio menor era consciente de cuál era el objeto, parece indicar la existencia de un conflicto entre la voluntad de los dos hemisferios, cada uno de los cuales trataba de expresar la que creía que era la respuesta correcta. No obstante, Lishman (1971) no considera que estos casos demuestren la existencia de dos sistemas volitivos distintos en estos pacientes. El autor señala que lo que parece una voluntad separada por parte del hemisferio menor se manifiesta únicamente por medio del comportamiento motor de los pacientes, comportamiento que podría ser meramente automático. Lishman considera que lo más sensato es suponer, no la existencia de dos sistemas volitivos paralelos, sino que los procesos cognitivos de ambos hemisferios son capaces de acceder a cualesquiera que sean los sistemas neurales que controlan la volición.

Lo más sorprendente en este tipo de casos, sin embargo, no es lo que sucede durante las pruebas experimentales, sino lo que sucede durante la vida cotidiana de los pacientes. Mientras que en las situaciones experimentales en que se envían ciertos estímulos de manera que sólo lleguen a un hemisferio todo parece indicar la existencia, usando la expresión de Sperry, de dos esferas de

conciencia separadas, en la vida diaria de estos pacientes prácticamente nada parece indicarlo. Los pacientes no sufren grandes problemas durante su vida diaria, y no parecen ser conscientes de la falta de comunicación entre ambos hemisferios. En un primer momento, la intuición parece decirnos que, si les preguntáramos, estas personas nos informarían de que han perdido la visión de medio campo visual y la sensibilidad en la mitad del cuerpo. No obstante, nada de esto sucede. En ningún caso los pacientes son conscientes de haber perdido la mitad del campo visual. En cuanto a las sensaciones en la mitad izquierda del cuerpo, Sperry (1966) nos dice que los sujetos sólo se quejaron de cierto entumecimiento en la mano izquierda. Los aparentes problemas de conflicto entre la voluntad de ambos hemisferios, a pesar de lo inquietantes que puedan parecer, desaparecieron cuando hubo pasado un tiempo tras la operación. Los problemas de coordinación importantes se limitaron al periodo de tiempo cercano a la intervención, después los pacientes fueron capaces, como señala Nagel (1979), de llevar a cabo sus actividades cotidianas sin grandes dificultades, incluyendo aquellas que requieren cierta coordinación, como nadar o tocar el piano. Nagel también señala que ambos hemisferios se duermen y se despiertan a la vez, y que, en general, se dedican siempre a las mismas actividades y trabajan conjuntamente para llevarlas a cabo. De hecho, los experimentadores tienen que ser muy cuidadosos para evitar que la información llegue a ambos hemisferios. Sperry (1966) señala que, cuando en los experimentos se emplean estímulos visuales, las imágenes tienen que proyectarse durante el tiempo suficiente para que sean percibidas pero con la suficiente brevedad para impedir que un movimiento de los ojos permita al otro hemisferio recibir la información. El autor señala que en las pruebas que involucraban el uso del tacto, era necesario evitar sonidos característicos que pudieran alertar al hemisferio contrario de cuál era el objeto. Era necesario también, por supuesto, taparle los ojos al paciente y evitar que la mano contraria tocara el objeto, cosa que a menudo trataba de hacer. Elizabeth Schechter (2015) nos habla de un experimento en que se ponían en la mano izquierda de un paciente de este tipo objetos con forma de esfera, cubo o pirámide y se le pedía que dijera cuál de las tres era la forma de cada objeto. Para sorpresa de los experimentadores, el paciente obtuvo un gran porcentaje de aciertos. Descubrieron que cuando se le colocaba el objeto en la mano, el hemisferio menor provocaba que el sujeto mirara hacia un elemento de la habitación que tuviera la forma correspondiente, permitiendo así que el hemisferio dominante obtuviera la información que necesitaba. Cuando al paciente se le taparon los ojos, empezó a cometer los fallos que hemos visto que tenían lugar en otros casos. Nagel (1979) afirma que hay evidencia de que ambos hemisferios son capaces de comunicarse aprovechando el hecho de estar ambos conectados a los músculos de la cara y el cuello. Parece ser que cada hemisferio envía señales al otro produciendo movimientos en esos músculos, movimientos que el otro hemisferio es capaz de detectar. El autor nos dice también que en casos de personas nacidas sin el cuerpo calloso los resultados experimentales son mucho más cercanos a la normalidad, presumiblemente debido a que se han creado nuevas rutas de

comunicación entre ambos hemisferios. Los diferentes autores señalan, además, que, con el tiempo, los hemisferios empezaron a desarrollar control motor de ambos lados, cosa que dificultó saber qué hemisferio estaba actuando. Parece ser también que el hemisferio menor desarrolla la capacidad de leer a los pocos meses de la operación.

Por supuesto, la operación tuvo algunos efectos en la vida diaria de los pacientes, especialmente en el periodo cercano a la operación, pero Sperry (1966) los considera sorprendentemente leves teniendo en cuenta la magnitud de la intervención. El autor habla de cosas como una reducción de la meticulosidad y la iniciativa, un peor funcionamiento de la memoria a corto plazo y una peor coordinación entre ambas manos, pero señala que hubo mejoría con el paso del tiempo.

Derek Parfit (2004) comparte la opinión de Sperry de que estos resultados empíricos muestran la existencia de dos mentes en esos pacientes. Parfit imagina el caso de que ambos hemisferios pudieran volver a unirse. Según su opinión, en dicho caso recordaríamos lo que hubiéramos experimentado en cada una de las dos corrientes de conciencia y recordaríamos también que en cada una no fuimos conscientes de lo que experimentamos en la otra. Aunque por el momento no es posible volver a unir ambos hemisferios, Bernard Williams (1986b) nos refiere un caso que puede ayudarnos a hacernos una idea de lo que sucedería en tal situación. Williams describe un caso de personalidad múltiple. La paciente, a la que el autor se refiere como la señorita Beauchamp, tenía múltiples personalidades con distinto carácter y diferentes gustos y habilidades. En este caso, algunas de las personalidades eran conscientes de lo que hacían las demás y otras no. A pesar de que la discontinuidad de memoria no estuviera presente entre todas las personalidades, este caso parece indicar la existencia de múltiples personas de manera incluso más clara que los casos de escisión del cuerpo calloso, puesto que, además de las diferencias en el carácter y las habilidades, las distintas personalidades mostraban conflictos violentos entre sí. Pues bien, Williams señala que, cuando la paciente estuvo prácticamente curada, veía sus diferentes personalidades como si hubieran sido solamente cambios de humor y consideraba que había sido la misma persona todo el tiempo. Este caso, a pesar de las diferencias, parece apoyar la situación que Parfit supone que se daría en caso de que volvieran a unirse ambos hemisferios cerebrales.

A partir de estos casos, Parfit examina una situación imaginaria. El autor supone que se hubiera inventado un dispositivo capaz de interrumpir temporalmente la comunicación que ambos hemisferios llevan a cabo entre sí a través del cuerpo calloso. Parfit imagina que un dispositivo como este podría utilizarse, por ejemplo, de la siguiente manera. En un examen de física, uno podría tener dos ideas distintas sobre cómo tratar de resolver un problema. Si sólo tuviera tiempo para probar una de las dos soluciones, la persona en cuestión podría emplear el dispositivo para dividir su mente y probarlas ambas paralelamente, volviendo a unir ambos

hemisferios al final y escogiendo la solución correcta²⁷. El autor señala que en un caso como este, en que la división dura poco tiempo, es difícil creer que mientras dura haya dos personas distintas involucradas. Como señala el autor, cada una de esas personas creería ser la persona original y cuando se llevara a cabo la unificación, la persona recordaría haber sido ambas. Como afirma Parfit, también resultaría implausible la idea de que habría tres personas involucradas. Eso significaría suponer que habría una persona que dejaría de existir en el momento de la división, que entonces empezarían a existir dos personas distintas a la anterior, ambas creyendo ser la original, y que estas dos últimas desaparecerían con la reunificación, momento en el cual la persona original regresaría a la existencia. Lo más plausible es, pues, suponer que habría una única persona involucrada²⁸. Entonces, el autor afirma que si pensamos que en estos casos sólo existe una persona, tenemos que renunciar a la teoría de que la unidad de conciencia se explica por la propiedad, puesto que en estos casos habría dos unidades de conciencia que contendrían experiencias pertenecientes a una misma persona. Parfit afirma que si consideramos este caso imaginario en que la mente se divide y se reunifica al poco tiempo, las explicaciones, que hemos examinado antes, que él da a la unidad de conciencia y a la unidad de las experiencias que forman parte de la vida de una persona se convierten en la mejor explicación. También es obvio, como señala el autor, que los casos de mentes divididas representan una gran evidencia en contra de la tesis de que somos entidades que existen separadamente.

Parfit imagina otra situación en que el cerebro de una persona es dividido en dos y trasplantado al cuerpo de dos gemelos suyos idénticos a él. Este caso lo utiliza para defender la afirmación de que la identidad personal no es siempre determinada, por lo que volveremos a él en la sección correspondiente. De momento basta con señalar que aunque tal caso es imposible actualmente, los casos de separación de los hemisferios cerebrales que estamos considerando indican que tal imposibilidad es meramente técnica.

No obstante, no todos los autores consideran que de los casos de bisección cerebral puedan obtenerse las conclusiones que Parfit extrae. Grant Gillett (1986) cree que en tales casos no puede hablarse ni de dos mentes ni de dos esferas de conciencia separadas. El autor cree que lo que hay en esas situaciones es una única mente que no está funcionando correctamente. Gillett señala que cuando esas personas cometen errores no creen que el error haya sido de otra persona, sino más bien que han cometido un error que cuentan con el conocimiento para corregir. El autor señala, además, que las habilidades mostradas en estos casos por el hemisferio

²⁷ Hemos visto que el hemisferio menor probablemente no tendría las habilidades necesarias, pero se cree que cierto porcentaje de personas tiene las mismas habilidades en ambos hemisferios y Parfit supone que la persona en cuestión fuera uno de esos casos.

²⁸ Esto sería lo más sensato si no fuéramos reduccionistas, al tratar sobre la afirmación de Parfit de que la identidad personal no es siempre determinada volveremos sobre estos casos y veremos cuál es la mejor manera de entenderlos.

menor pueden deberse a reacciones automáticas y que no demuestran que dicho hemisferio sea consciente²⁹. Gillett considera que lo que les sucede a estos pacientes es que padecen cierto tipo de confusiones y afirma que, en ausencia de lenguaje, no está justificado atribuir conciencia al hemisferio derecho.

Thomas Nagel (1979), contra la existencia de dos mentes en estos pacientes, nos recuerda que en las situaciones cotidianas actúan como si fueran una sola persona. El autor añade, además, que si supusiéramos que existen dos mentes, tendríamos que concluir que existen dos mentes en todos los individuos, pero que no nos damos cuenta debido a la comunicación entre ambos hemisferios. Sin embargo, Nagel argumenta que en las situaciones experimentales parece implausible suponer que existe una única mente. Además de las tareas que ya hemos visto que el hemisferio menor puede desempeñar, el autor señala que estos pacientes son capaces de prestar atención, paralelamente con cada hemisferio, a dos tareas incompatibles entre sí. Puesto que las situaciones experimentales no provocan ningún tipo de cambio anatómico en el cerebro de los pacientes, Nagel descarta también la posibilidad de que exista una sola mente que se divida en dos al darse tales situaciones. La conclusión a la que llega el autor es que la noción de una mente individual debe ser rechazada. Nagel señala en apoyo de su tesis que, incluso en las situaciones experimentales, los pacientes actúan constantemente como si fueran una sola persona y que la división sólo se aprecia en sus reacciones a aquellos estímulos que llegan a un único hemisferio. Según el autor, lo que estos casos nos muestran es que el concepto de un sujeto individual de conciencia no tiene aplicación. Para Nagel, la unidad de conciencia es sólo una función de la integración de los contenidos del sistema que controla nuestro organismo. Lo único que sucede en esos casos es que la información que recibe el hemisferio menor no puede llegar al otro hemisferio, que es el que controla el lenguaje y su mitad del cuerpo. La pregunta de si en esos casos hay una sola conciencia o dos carece de sentido, al explicar el modo en que se integran los contenidos de los distintos procesos mentales ya estamos explicando todo lo que podríamos explicar. En mi opinión, las conclusiones de Nagel no son en absoluto incompatibles con el reduccionismo. Estas conclusiones apoyan, además, la idea de que en el caso imaginario de la división y la reunificación no podríamos decir en ningún sentido relevante que existen varias personas. Además, puesto que en el ejemplo en que los dos hemisferios son trasplantados a cuerpos distintos no podría haber comunicación ente ellos, las conclusiones de Nagel no son

²⁹ Esta idea podría verse apoyada por ciertos casos de ceguera cortical. Roger Penrose (2015) nos habla de un paciente, denominado D.B., al que se le extirpó parte del córtex visual. D.B. afirmaba que no veía nada en la parte correspondiente, pero si se situaba un objeto en dicha parte y se le pedía que imaginara cuál de entre varias formas tenía, acertaba en un elevado porcentaje de los casos, a pesar de que seguía sosteniendo que no veía nada. El autor concluye que en ese caso la información llegaba a un área no consciente del cerebro y señala que, tras ser entrenado, el paciente logró obtener conciencia real, aunque limitada, de dicha zona.

incompatibles con la idea de que habría dos mentes distintas³⁰. Por lo tanto, a pesar de que la teoría de Nagel, de ser cierta, convertiría en inadecuado el lenguaje con que Parfit trata estos casos, en mi opinión no impide que se sigan de ellos las conclusiones que él extrae, por no hablar de que el reduccionismo parece ser la concepción de la identidad personal más coherente con la idea de que no pueda hablarse de mentes individuales.

Elizabeth Schechter (2015) defiende la idea de que en tales casos existen dos esferas separadas de conciencia. La autora argumenta que hay evidencia que parece demostrar que la comunicación entre ambos hemisferios se lleva a cabo a través del exterior, de manera indirecta³¹. Para la autora, la apariencia de una sola mente en las situaciones ordinarias se debe a que ambos hemisferios comparten un mismo cuerpo y cooperan entre sí. Según argumenta, una comunicación que se lleva a cabo de manera externa no puede llevar a una sola mente unificada. La autora nos refiere un caso en que se tomó una imagen de la actividad cerebral de uno de estos pacientes antes y después de la operación. La imagen se tomó con el paciente anestesiado, de modo que no estaba llevando a cabo ningún tipo de tarea y la comunicación externa de ambos hemisferios no resultaba posible. Se observó una gran reducción de la actividad correlacionada de ambos hemisferios y los experimentadores llegaron a la conclusión de que tal resultado era compatible con un funcionamiento independiente de cada hemisferio con respecto al otro. Sperry (1966) también argumenta en favor de la existencia de dos mentes en estos pacientes. Compara el hecho de que los pacientes no sean conscientes de haber perdido la mitad del campo visual con situaciones similares observadas en víctimas de destrucción parcial de la retina. El autor señala también que la falta de capacidades lingüísticas en el hemisferio menor es comparable a la situación de pacientes que sufren afasia a causa de daños en los centros de habla y escritura. Nadie considera que tales pacientes carezcan de mente, y cuando se recuperan recuerdan haber sido conscientes durante el periodo en que no pudieron emplear el lenguaje para comunicarse.

Es cierto que los resultados que hemos examinado no demuestran de manera clara que existan voluntades en conflicto entre ambos hemisferios, pero Gazzaniga, Ledoux y Wilson (1977) nos hablan de un sujeto, al que llaman P.S., que mostró capacidades lingüísticas en ambos hemisferios³². Como señalan los autores, P.S. era capaz, con el hemisferio menor, de juntar letras para formar palabras y juntar palabras con sus antónimos o con palabras que rimaran con ellas. Roger Penrose (2015) nos cuenta que el hemisferio derecho de P.S. terminó aprendiendo a hablar. Sus dos hemisferios, señala el autor, manifestaron distintos gustos y deseos. Mientras

³⁰ Por supuesto, si aceptamos la idea de Nagel de que el concepto de mente individual carece de aplicación, la expresión “dos mentes distintas” sólo puede ser una forma de hablar.

³¹ El ejemplo del uso de los músculos del cuello y la cara que hemos visto antes sería un ejemplo.

³² Para el historial médico de dicho paciente, *vid.* Wilson, Reeves, Gazzaniga y Culver (1977).

que el hemisferio dominante quería ser dibujante, el menor deseaba ser piloto de carreras. Este último caso hace muy difícil dudar de la existencia de dos esferas separadas de conciencia en este tipo de pacientes. También resulta difícil creer que la comunicación externa entre los dos hemisferios pueda provocar unidad de conciencia. Pero aún sigue ahí la objeción de Nagel de que si concluimos que hay dos mentes en estos pacientes, tendremos que concluir que también las hay en sujetos normales, objeción que parece apoyar su teoría. Seguramente, cuando sepamos más sobre cómo funciona la conciencia sabremos si tenemos motivos para considerar que la bisección cerebral puede provocar que haya dos conciencias de las mismas experiencias en las situaciones de la vida diaria. Cuando tengamos los conocimientos necesarios podremos determinar si esos pacientes tienen dos mentes o si, por el contrario, Nagel tiene razón. Lo que sí parece estar claro, es que no se puede decir que estos pacientes tengan una sola mente. Por lo tanto, sea cual sea la verdad, parece que las conclusiones que Parfit extrae de estos casos se siguen efectivamente de ellos.

David Hershenov y Adam Taylor (2014) aceptan la existencia de dos esferas separadas de conciencia en estos casos pero defienden que ello no contradice la tesis de que somos entidades que existen separadamente. Los autores argumentan que la idea de que dichos casos demuestran la inexistencia del alma se basa en dos suposiciones erróneas. Uno de estos errores sería creer que el alma no depende del funcionamiento del cerebro y el otro consistiría en pensar que, puesto que el alma es simple, tiene que existir unidad en su pensamiento. Contra el segundo de los errores, los autores argumentan que nada impide que una entidad sea consciente de distintos pensamientos que no se encuentren interrelacionados entre sí, de modo que ambas corrientes de conciencia le pertenezcan pero que en cada una no sea consciente de la otra. Contra el primero de los dos errores arguyen que el alma necesita al cerebro para ser capaz de pensar y que, por lo tanto, los cambios en este afectan al pensamiento. Los autores ponen la visión como ejemplo. Del mismo modo que necesitamos los ojos para ver y que lesiones en los ojos afectan a nuestra visión, necesitamos el cerebro para pensar y tener experiencias y las lesiones cerebrales afectan a estos procesos. Sin embargo, podría argumentarse que si todas nuestras experiencias dependen únicamente del cerebro, el alma se convierte en innecesaria. Si todo lo que experimentamos depende de los procesos neurológicos y cualquier cambio en el cerebro puede provocar cualquier tipo de cambio en nuestro carácter o nuestros recuerdos, ¿para qué suponer la existencia de un alma? Tal como los autores ponen las cosas, podríamos estar compartiendo todos una única alma y no notaríamos ninguna diferencia. Además, los dos ejemplos imaginarios de Parfit hacen muy implausible la existencia de dicha alma. En el caso del trasplante, parece poco creíble que pueda haber dos personas compartiendo una única alma, mientras que en el caso de la división y la reunificación parece absurdo suponer que aparece una nueva alma que desaparece cuando se restablece la comunicación entre los hemisferios. Los

autores resuelven el problema diciendo que Dios crea una nueva alma en el momento en que los hemisferios se colocan en una situación en que no sea posible la unidad de agencia. Es de agradecer que Dios tenga la amabilidad de hacer que el número de almas se ajuste en todo momento a nuestras intuiciones, pero si esta solución no es *ad hoc.*, nada lo es. Además, podríamos preguntar qué sucedería si volviéramos a trasplantar ambos hemisferios al cuerpo original y los volviéramos a conectar. Esta situación sólo puede recibir una explicación satisfactoria si rechazamos definitivamente la idea de que somos entidades que existen separadamente.

Antes de pasar a examinar la afirmación de Parfit de que la identidad no es siempre determinada, puede ser interesante examinar cuál es el mejor criterio de identidad personal en caso de que aceptemos el reduccionismo. Esto es lo que haremos en la siguiente sección.

5. Bernard Williams: La importancia del cuerpo para la identidad personal

Hemos visto que, dentro del reduccionismo, Parfit defiende el criterio psicológico de identidad personal. Bernard Williams, en cambio, defiende que la continuidad del cuerpo es necesaria para que haya identidad. Williams emplea una serie de argumentos en contra del criterio psicológico y en defensa de la importancia del cuerpo para la identidad personal. Parfit dedica mucha atención a los argumentos de Williams y responde a la mayor parte de ellos. En esta sección examinaré los principales argumentos de Williams y las respuestas que Parfit les da y trataré de defender que la continuidad psicológica tiene mayor importancia que la continuidad corporal³³.

Williams (1986e) imagina una situación en que alguien supiera que iba a sentir un gran dolor al cabo de cierto tiempo. El autor señala que si esa persona fuera a permanecer consciente todo el tiempo y se le comunicara que antes de que sufriera ese dolor iba a experimentar ciertos cambios psicológicos que harían que dejara de ser psicológicamente continua con su yo actual, dicha persona, lejos de sentirse aliviada, consideraría todavía más preocupante su situación. Williams imagina diferentes cambios psicológicos que podría sufrir dicha persona. (1) A esa persona podría provocársele amnesia. (2) A esta pérdida de recuerdos podrían sumársele cambios en su carácter. (3) Podría suceder también que, además de lo anterior, tuviera recuerdos falsos. (4) Estos recuerdos falsos y los rasgos de carácter que dicha persona tuviera tras los

³³ En la sección siguiente, cuando examinemos la afirmación de Parfit de que la identidad personal no es siempre determinada, trataré de defender que en realidad nunca es determinada. Puede que esa conclusión haga parecer irrelevante la discusión de la presente sección, pero cuando discutamos la opinión de Parfit de que lo que importa no es la identidad personal sino la conexividad y/o la conectividad psicológicas, su importancia se hará evidente.

cambios podrían encajar, casualmente, con los de otra persona. (5) El hecho de que sus características psicológicas encajaran con las de otra persona podría no ser casualidad, podría haberse hecho que coincidieran transfiriendo dichas características desde el cerebro de la otra persona. (6) Por último, podría suceder que, además, a la otra persona se le hubieran aplicado los mismos cambios en sentido inverso, provocando que ambas intercambiaran sus rasgos psicológicos. Según argumenta el autor, el cambio (1) no da motivos para dejar de preocuparse por el dolor, y lo que se añade al pasar de cada cambio al siguiente no puede marcar tampoco una diferencia, de modo que hemos de suponer que, en el caso de que se llevara a cabo el cambio (6) desde el principio, ello tampoco nos daría motivos para dejar de preocuparnos por ese dolor futuro. Parfit (2004) señala incluso que si estuviéramos sintiendo ese dolor mientras los cambios se producen, no seríamos conscientes de dichos cambios. En una situación así, los cambios sólo supondrían una modificación en el modo en que actuaríamos cuando el dolor terminara, es decir, una modificación en nuestras disposiciones. Parfit responde al argumento de Williams señalando que si la persona permaneciera consciente todo el tiempo sí que habría cierto tipo de continuidad psicológica, habría continuidad de lo que se suele llamar presente especioso. Somos conscientes simultáneamente de las experiencias que tienen lugar en cierto intervalo de tiempo y ello crea una continuidad psicológica más fuerte que la provocada por el resto de conexiones psicológicas. También podríamos imaginar una situación en que cada célula del cuerpo de una persona fuera substituida por otra. Si la persona estuviera consciente durante todo el proceso y sintiendo un gran dolor, parece poco creíble la idea de que el cambio fuera a poner fin a su sufrimiento. De hecho, podría realizarse la substitución de modo que al final del proceso el cuerpo y el cerebro de esa persona fueran cualitativamente distintos, pero, aun así, la situación parece tener las mismas consecuencias que la anterior en que se eliminaba la continuidad psicológica. Tanto en este caso como en el anterior, lo que hace implausible la idea de que la persona tendría que esperar el fin de su sufrimiento es precisamente un determinado tipo de continuidad psicológica, por lo que, en realidad, el experimento mental de Williams apoya el criterio psicológico en lugar del corporal. Si entre dos momentos distintos del tiempo una persona ha permanecido consciente, siempre existe ese tipo de continuidad psicológica entre sus dos yoés. Si antes de eliminar su continuidad psicológica se dejara inconsciente a la persona en cuestión, sí que tendría razones para dejar de preocuparse por su dolor futuro. De hecho, puesto que la memoria es la conexión psicológica más importante, si se dejara inconsciente a dicha persona y se le practicara únicamente el cambio (1), tendría, por lo menos, motivos para estar menos preocupada por ese dolor futuro.

Parfit, a menudo, pone el ejemplo del teletransporte. Imagina que existiera una máquina que pudiera escanear la posición en que se encuentran nuestras partículas y transmitir la información a otro lugar para que se creara allí una réplica exacta tras la destrucción del original. La réplica

sería psicológicamente continua con el original, pero no habría continuidad física entre ambos. Por lo tanto, si el criterio psicológico es correcto, el teletransporte es una forma de viajar. Si, por el contrario, son los defensores del criterio físico los que tienen razón, el teletransporte es una forma de morir. Parfit imagina también el caso, al que se refiere como el caso de la línea secundaria³⁴, de que no se destruyera el original. Según la versión del criterio psicológico que defiende el autor, puesto que existirían dos personas psicológicamente continuas con el individuo que utilizó la máquina, ninguna de las dos sería el original. La exigencia de que la relación R se dé de forma no ramificada tiene como objetivo evitar otra de las objeciones planteadas por Williams (1986a). Dicha objeción se basa en la transitividad propia de la relación de identidad. La identidad es una relación uno-a-uno, de modo que ninguna relación que sea uno-a-muchos o muchos-a-muchos podrá ser un criterio adecuado de identidad. La continuidad psicológica y la conexividad psicológica no son relaciones uno-a-uno, de modo que la relación R tampoco lo es. Es por ello que Parfit introduce la condición de que dicha relación no tome una forma ramificada, convirtiendo así el criterio psicológico de identidad personal en una relación uno-a-uno. Contra la posibilidad de resolver el problema de este modo, Williams (1986a; 1986b; 1986d) argumenta que la diferencia entre que una persona sea o no idéntica a otra no puede deberse a un hecho arbitrario o trivial. Según el autor, es poco plausible que el que yo vaya a ser una persona futura o no pueda depender de un hecho externo. De acuerdo con el criterio de Parfit, en el caso de la línea secundaria yo no sería mi réplica debido a que no se destruyó el original. Si el original se destruyera, eso me convertiría instantáneamente en mi réplica. Lo que nos dice el argumento de Williams es que no puede ser satisfactorio un criterio que haga depender la identidad personal de un suceso como ese. Un suceso que no afecta en nada a mi réplica no puede marcar la diferencia entre que dicha réplica sea yo o no. Volvamos al ejemplo en que se trasplantaba cada hemisferio a un cuerpo idéntico distinto. En ese caso, según el criterio psicológico de Parfit, yo no sería ninguna de las dos personas resultantes, pero si uno de los dos trasplantes fallara, entonces la única persona resultante sí que sería yo. Lo que nos dice el argumento de Williams es que el hecho de que yo sea esa persona no puede depender de lo que le suceda a otro medio cerebro. Este último ejemplo pone de manifiesto que el criterio físico está expuesto al mismo problema. En el caso del trasplante de los dos hemisferios, la continuidad corporal se daría de forma ramificada, siendo, por lo tanto, una relación uno-a-muchos. Además, tanto Williams como Parfit señalan la posibilidad de que una persona se dividiera como una ameba. No obstante, Williams argumenta que, en el caso de la continuidad corporal, es posible, por lo menos, detectar la ramificación. Según el autor, puesto que para aplicar el criterio corporal necesitaríamos conocer la historia espaciotemporal del cuerpo del individuo en cuestión, el propio criterio nos permitiría llegar a la conclusión de que en ese caso

³⁴ Como veremos más adelante, en este caso Parfit imagina que la máquina dañara al original y le dejara sólo con cierto tiempo de vida.

no habría identidad. Además, Williams señala que si hay una división, no puede hablarse de continuidad física en sentido estricto. Lo que el autor señala con estos argumentos es que, a diferencia de lo que sucedía con la continuidad psicológica, la bifurcación no es un hecho externo o trivial. No obstante, Williams comete un error con estos argumentos. Es cierto que en el caso del trasplante de ambos hemisferios³⁵, el criterio físico nos permitiría detectar la bifurcación y responder que no hay identidad, pero si después uno de los trasplantes fracasara y el hemisferio correspondiente quedara destruido, tendríamos que decir que la única persona resultante es numéricamente idéntica a la original, y no veo que el criterio físico presente en esto ninguna ventaja en comparación con el criterio psicológico. Además, aunque la división sea un hecho importante para la continuidad física, lo que suceda después con la otra mitad ya no lo es y, por lo tanto, no puede marcar la diferencia entre que haya identidad o no³⁶. Parfit utiliza este último ejemplo para señalar que el criterio físico no cumple tampoco los requisitos que exige Williams. Lo que suceda después con uno de los hemisferios no es parte, como señala Parfit, de mi relación con el otro. Además, como argumenta Parfit, lo que le suceda a uno de los dos hemisferios es trivial, en el sentido que le da Williams a esta palabra, para el otro. Según Parfit, ningún criterio de identidad personal puede cumplir los requisitos que Williams exige.

Williams (1986b) emplea otros argumentos en defensa del criterio corporal. El autor argumenta que sólo en referencia al cuerpo podemos distinguir entre identidad numérica e identidad cualitativa³⁷. En mi opinión, este argumento es equivocado. Evidentemente, en un mismo instante del tiempo hay una diferencia clara entre que dos cuerpos sean idénticos y que sean el mismo, puesto que un mismo cuerpo no puede estar en varios lugares a la vez. Puesto que no podemos ver los pensamientos, podría parecer que no sucede lo mismo con la mente, pero también hay una diferencia entre que dos mentes sean la misma y que sean exactamente similares. Cuando nos preguntábamos si en la vida cotidiana de las personas con bisección cerebral había una sola mente o dos mentes teniendo las mismas experiencias, parece que estábamos realizando una pregunta que tendría sentido responder. Es cierto que, a diferencia de lo que sucede con los cuerpos, no podemos determinar la respuesta a simple vista, pero ello no significa que no haya una respuesta. Si nos referimos, en cambio, a dos personas que existan en dos momentos distintos del tiempo, entonces no parece estar claro que en el caso del cuerpo

³⁵ Supongo aquí que la continuidad del cerebro es suficiente para que se cumpla el criterio físico. Williams no aceptaría dicha suposición, pero el ejemplo puede modificarse fácilmente para evitar ese problema. Podría, por ejemplo, recurrirse al caso en que una persona se divide como una ameba.

³⁶ Williams tiene razón en que la división afecta a la continuidad física, pero podría decirse lo mismo de la destrucción de parte del cerebro y aun así consideramos que las personas que han sufrido lesiones cerebrales graves siguen siendo ellas.

³⁷ Parfit responde a esta objeción argumentando que si somos reduccionistas tenemos que admitir que en ciertos casos no hay diferencia entre la identidad numérica y la cualitativa. Sin embargo, aunque Parfit tiene razón, considero que Williams comete con su argumento un error aún más básico. Es por ello que, en lugar de centrarme en la respuesta de Parfit, ofrezco mi propia respuesta al argumento de Williams.

podamos distinguir entre identidad numérica y cualitativa. Tal vez podría decirse que dos cuerpos son numéricamente idénticos si hay continuidad física entre ellos, pero entonces podemos apelar a la continuidad psicológica en el caso de las mentes. Quizás podría objetarse que la continuidad psicológica puede darse de forma ramificada y que ello supondría un problema, pero ya hemos visto que lo mismo sucede con la continuidad física. Por lo tanto, si se considera que con respecto a los cuerpos podemos distinguir entre identidad numérica y cualitativa, lo mismo puede decirse de las mentes.

Williams (1986b) emplea otro argumento contra el criterio psicológico. El autor argumenta que un criterio de identidad personal ha de permitir juzgar si una persona es numéricamente idéntica a otra o no y afirma que ello sólo es posible recurriendo al cuerpo. Williams señala que para poder juzgar si hay continuidad psicológica, tenemos que saber lo que una persona ha hecho, y la única manera de saber lo que una persona ha hecho es recurriendo a testigos que tendrán que haber visto el cuerpo de dicha persona. Además, el autor afirma que si una persona cambiara constantemente de cuerpo sería totalmente imposible identificarla, ya que para identificar a una persona tenemos que recurrir al cuerpo. Sin embargo, en mi opinión, el problema de la identidad personal a través del tiempo no es un problema sobre qué método utilizar para identificar a las personas, sino más bien sobre en qué consiste ser la misma persona que una persona pasada. Si estuviéramos buscando un método para determinar si una persona actual es la misma que una persona pasada, probablemente el cuerpo tendría la importancia que le atribuye Williams, pero lo que queremos averiguar es en qué consiste que una persona presente sea la misma que una persona pasada. Para esta tarea no hay problema en suponer que contamos con pleno conocimiento de lo que sucede en la mente de las personas. Si queremos averiguar en qué consiste la identidad, no podemos descartar la posibilidad de que en ciertos casos la identidad no pueda ser verificada. Tal vez cuando sepamos más sobre el funcionamiento del cerebro tengamos alguna manera de determinar si una persona es psicológicamente continua con otra. Tal vez entonces la identificación ya no sea imposible en el caso del cambio constante de cuerpo que imagina Williams.

Williams (1986b) también argumenta que el cambio de cuerpo podría no ser lógicamente posible. El autor pone en duda que una persona pueda ser la misma si no es capaz, por ejemplo, de sonreír con la que era su sonrisa característica. Lo que parece estar argumentando Williams es que el cuerpo afecta a la mente de manera más profunda de lo que podría parecer. Puede que el autor tenga razón, pero ello no supone ningún problema para el criterio psicológico. Tal vez un cambio de cuerpo pueda ser capaz de acabar con la continuidad psicológica o de reducirla demasiado, pero no veo de qué modo puede impedir eso que la identidad personal consista en esa continuidad psicológica.

Volvamos al caso del teletransporte imaginado por Parfit. Supongamos que la máquina creara siempre una nueva réplica sin destruir el original. Si, cada cierto tiempo, el original volviera a utilizar la máquina, se llevaría la sensación de que lo único que hace la máquina es crear una réplica suya en algún otro lugar sin tener sobre él ningún efecto. Para él, la máquina sería una manera de crear réplicas suyas. En cambio, si cada réplica utilizara de nuevo la máquina, la última de las réplicas tendría la sensación de que la máquina le transporta a otro lugar y deja una copia suya en el lugar original. Para ella, la máquina sería una forma de viajar que tiene el efecto secundario de dejar un duplicado de la persona en el lugar de origen. ¿Qué impresión se llevaría una persona que hubiera sido el original en unas ocasiones y la réplica en otras³⁸? Dicha persona no tendría una impresión clara sobre esa máquina. Para esa persona, usar la máquina supondría no saber si se creará simplemente una réplica suya en otro lugar o si se encontrará de repente en otro sitio habiendo dejado un duplicado tras de sí. Por supuesto, todas las impresiones que acabo de describir no son más que ilusiones. Si el reduccionismo es cierto, cuando una persona fuera a utilizar la máquina, no tendría sentido que preguntara cuál de las dos cosas le sucederá, sencillamente habrá dos personas psicológicamente continuas con ella. La sensación que sentiría la persona de que al instante siguiente podría estar en un sitio o en otro no sería más que una ilusión. Del mismo modo, si fuera posible transferir la continuidad psicológica de un cerebro a otro, la persona en cuestión esperaría despertarse en un nuevo cuerpo. Es cierto que, del mismo modo que en el caso anterior, podría decirse que, en cierto sentido, eso sería una ilusión. Aun así, si hay un determinado tipo de relación que provoca en nosotros la sensación de identidad, es razonable pensar que si la identidad de hecho existe, tal relación tiene que tener un papel importante en ella.

En *La memoria de Shakespeare*, Borges (1983/2005) nos cuenta la historia de un hombre al que otro le ofrece la memoria de Shakespeare. Tras aceptar el ofrecimiento, el protagonista empieza a experimentar lo que podríamos llamar cuasi-recuerdos de la vida pasada de Shakespeare. Borges (1983/2005) describe del siguiente modo las experiencias del protagonista:

En la primera etapa de la aventura sentí la dicha de ser Shakespeare; en la postrera, la opresión y el terror. Al principio las dos memorias no mezclaban sus aguas. Con el tiempo, el gran río de Shakespeare amenazó, y casi anegó, mi modesto caudal. Advertí con temor que estaba olvidando la lengua de mis padres. Ya que la identidad personal se basa en la memoria, temí por mi razón. (p. 396)

Un poco más adelante, el autor nos dice por boca de su protagonista que “todas las cosas quieren perseverar en su ser, ha escrito Spinoza. La piedra quiere ser una piedra, el tigre un

³⁸ Me refiero a una persona que fuera, por ejemplo, la réplica de una persona que ya había usado la máquina en otras ocasiones y que tras esas ocasiones resultó ser el original.

tigre, yo quería volver a ser Hermann Soergel” (Borges, 1983/2005, p. 396). Es difícil dudar de que todos nos sentiríamos así en una situación semejante. Si tuviéramos una gran cantidad de cuasi-recuerdos de otra persona, sentiríamos que nuestra identidad se estaría viendo amenazada. La continuidad psicológica, especialmente la memoria, tiene una gran influencia en cómo vemos nuestra propia identidad, tiene una gran influencia en quién creemos que fuimos. Es difícil creer que algo que tiene tanta influencia en cómo vemos nuestra identidad pueda no ser esencial a ella. Esto no sucede con la continuidad física, por lo menos no de la manera tan fuerte en que sucede con la psicológica.

En la siguiente sección examinaremos la afirmación de Parfit de que la identidad personal no tiene por qué ser siempre determinada. Trataré de defender en dicha sección que la afirmación de Parfit tiene que llevarse más lejos, trataré de defender que la identidad personal nunca es determinada.

6. La vaguedad lingüística y la indeterminación de la identidad personal: los espectros psicológico, físico y combinado

Una de las afirmaciones más sorprendentes de Derek Parfit es la de que la identidad personal no tiene por qué ser siempre determinada. Según el autor, en ciertas situaciones no hay una respuesta a la pregunta de si una persona actual es o no la misma que una persona pasada. En dichas situaciones, la pregunta sería vacía. Las diferentes respuestas serían sólo distintas descripciones de una misma situación, no habría una que fuera la respuesta correcta.

Retomemos el ejemplo de los dos hemisferios cerebrales que se trasplantan a dos cuerpos idénticos. Parfit señala que, en este caso, en lo referente a la identidad personal, hay únicamente cuatro posibilidades. La persona no sobrevive, sobrevive como una u otra de las dos personas resultantes o sobrevive como ambas personas. El autor argumenta que ninguna de las opciones resulta satisfactoria. Si únicamente se trasplantara uno de los dos hemisferios y el otro fuera destruido, no hay duda de que la persona sobreviviría, y parece poco plausible que el trasplante exitoso del otro hemisferio pueda cambiar eso. Si el trasplante de uno de los hemisferios supone la supervivencia de la persona, es poco creíble que trasplantar también el otro suponga su muerte. Por lo tanto, decir que la persona no sobreviviría no parece ser la respuesta correcta. Tampoco parece razonable afirmar que sobreviviría como una sola de las dos personas resultantes. Contra esta posibilidad, Parfit argumenta que, puesto que ambas personas serían exactamente iguales³⁹, no hay nada que nos permita afirmar de manera plausible que una de las dos personas es la original y la otra no. Nos queda únicamente la posibilidad de que ambas

³⁹ Estamos suponiendo que se trata de un caso en que ambos hemisferios tienen exactamente las mismas capacidades.

personas resultantes sean numéricamente idénticas a la persona original. Contra esta última posibilidad, el autor argumenta que, a diferencia de lo que sucedía con el caso de la división y la reunificación, parece poco plausible que la persona pueda sobrevivir como ambas. Las dos personas resultantes podrían llevar vidas completamente distintas, podrían ir cambiando hasta que cada una tuviera una mente completamente diferente a la de la otra y podrían no volverse a encontrar nunca la una con la otra. Además, como señala el autor, suponer que ambas personas serían numéricamente idénticas a la original nos llevaría a conclusiones absurdas como la de que habría una misma persona en dos lugares distintos a la vez o a suposiciones implausibles como la de que ya existían dos personas, o incluso tres, antes de la división. Por lo tanto, ninguna de las posibilidades parece resultar satisfactoria. De hecho, como señala Parfit, aunque fuéramos capaces de llevar a cabo la operación, ello no nos permitiría conocer la respuesta a la pregunta. Podemos saber exactamente todo lo que sucedería en dicho caso y aun así ser incapaces de decir cuál de las cuatro posibilidades es la que describe correctamente la situación. El autor considera que esta situación es uno de los casos en que la pregunta sobre la identidad personal es vacía. En realidad, las cuatro opciones no son posibilidades diferentes, se trata únicamente de distintas descripciones de una misma situación. En opinión de Parfit, si sabemos que habría dos personas cada una de las cuales tendría un cuerpo idéntico a la original y uno de sus dos hemisferios, entonces lo sabemos todo. No hay nada que pueda hacer de una de las cuatro respuestas la única descripción correcta. Como señala el autor, si fuéramos entidades indivisibles que existieran separadamente, podría haber una respuesta correcta a la pregunta. Si eso fuera lo que somos, la persona original sobreviviría como aquella de las dos personas resultantes que se quedara con su ego cartesiano. Sin embargo, puesto que ya hemos visto que todo indica que no somos entidades de ese tipo, el reduccionismo es el que ofrece la mejor explicación a dicha situación. Parfit pone como ejemplo el caso de que un partido político se escindiera en dos. En dicho caso, la pregunta de cuál de los dos sería el partido original sería una pregunta vacía. Si diéramos una respuesta, no estaríamos afirmando nada sobre ningún profundo hecho metafísico, simplemente estaríamos escogiendo una de entre otras maneras de describir la situación. En un caso como el del partido político muchos aceptaríamos sin problema que la pregunta es vacía. Lo que afirma el reduccionismo es que la identidad personal es como la identidad de cosas como los clubs, las naciones o los partidos políticos. Sin embargo, aunque las distintas posibilidades sean sólo distintas maneras de describir una misma situación, Parfit considera que hay casos en que hay una que es la mejor descripción. En el caso del trasplante de los dos hemisferios cerebrales, el autor cree que la mejor descripción sería decir que ninguno de los dos es el original. Si decimos que ambas personas son el original, estamos violando la lógica de la identidad, puesto que la estamos convirtiendo en una relación no transitiva. Si decimos que una de las dos personas es el original, estamos haciendo una afirmación completamente arbitraria, puesto que ambas personas son idénticas y no hay ninguna

característica en ninguna de ellas que pudiera llevarnos a identificarla a ella con el original pero no a la otra. Por lo tanto, la mejor descripción consiste en afirmar que la persona original no sobreviviría y es por ello que Parfit exige, para que haya identidad personal, que la relación R se dé de forma no ramificada⁴⁰.

Para defender su afirmación de que la identidad personal no es siempre determinada, Parfit desarrolla tres casos imaginarios a los que se refiere como los espectros psicológico, físico y combinado. El autor comienza introduciendo el primero de estos experimentos mentales, el espectro psicológico. Este caso imaginario nos lo presenta como una modificación del experimento mental de Williams en que a una persona se le aplicaban distintos cambios psicológicos antes de que experimentara un fuerte dolor. Parfit imagina una serie de casos en que se aplicaran, de manera acumulativa, distintos cambios psicológicos. En los primeros casos de la serie, los cambios serían mínimos y la persona resultante sería claramente la misma que la original. En los últimos casos, los cambios psicológicos serían enormes y la persona resultante no sería la misma que la persona original. Entre los casos adyacentes de la serie, los cambios psicológicos no serían lo suficientemente distintos como para marcar la diferencia⁴¹. La pregunta, entonces, es qué pasaría en los casos intermedios. Según el autor, en los casos intermedios del espectro la pregunta de si la persona sobreviviría es vacía. La posibilidad de que la persona sobreviva y la de que no sobreviva no son en realidad dos posibilidades distintas, son sólo dos descripciones distintas de una misma situación. Si sabemos que la persona sufriría determinados cambios y que sería psicológicamente continua con la original sólo en un grado intermedio, sabemos todo lo que hay que saber sobre la situación, y al decidir si decimos que la persona ha sobrevivido o que no lo ha hecho sólo estamos escogiendo una forma de describir la situación. De hecho, Parfit considera que en los casos intermedios del espectro psicológico, a diferencia de lo que sucedía en el caso del trasplante de los dos hemisferios cerebrales, no hay una descripción que sea mejor que la otra. En estos casos, la identidad personal es indeterminada. Si sabemos exactamente lo que sucede en uno de esos casos, al preguntar si la persona es la misma que la original no estamos preguntando nada. La pregunta es vacía y, además, cada una de las dos descripciones es igual de buena que la otra.

⁴⁰ Podría objetarse que el hecho de hacer depender la identidad de un suceso externo y trivial convierte a esta en una mala descripción, pero, como ya hemos dicho, Parfit afirma que la identidad personal no es lo que importa, afirmación que permite que esta sea la mejor descripción a pesar de dicha objeción.

⁴¹ Si a una persona se le fueran aplicando uno a uno los distintos cambios, se podría argumentar que hay continuidad psicológica y que, por lo tanto, todas las personas de la serie son la misma. No obstante, la idea es que los distintos casos de la serie son distintas posibilidades, la persona en cuestión no recorrería la serie, sino que el conjunto de cambios en cuestión se aplicaría de golpe. De este modo, si se aplicaran los cambios de uno de los primeros casos, la persona sobreviviría, si se aplicaran los de uno de los últimos casos, la persona no sobreviviría y la diferencia entre aplicar los cambios de un caso y aplicar los del siguiente no es lo suficientemente grande como para marcar la diferencia entre que la persona sobreviva o no.

Los defensores del criterio físico podrían considerar que el argumento anterior, en lugar de demostrar que la identidad personal no es siempre determinada, demuestra que el criterio psicológico es incorrecto. Para responder a esta posibilidad, Parfit elabora otro experimento mental al que llama el espectro físico. El autor imagina una serie de casos en los cuales se reemplazan, de manera acumulativa, una a una las células del cerebro. Este caso imaginario tiene para el criterio físico las mismas implicaciones que el anterior tenía para el psicológico. Parfit considera la posibilidad de que se diga que hay una porción crítica del cerebro cuya continuidad sea necesaria para la identidad personal, pero descarta dicha posibilidad por ser poco plausible e imposible de comprobar.

El autor considera que alguien podría defender que tanto la continuidad física como la psicológica bastan, por sí solas, para que haya identidad personal. Para responder a esta posibilidad, Parfit propone un tercer experimento mental que combina los dos anteriores y al que se refiere como el espectro combinado. En este experimento mental, en cada caso de la serie se substituye una célula del cerebro y del cuerpo y se debilita ligeramente la continuidad psicológica. Como señala el autor, en este caso no podemos creer de ningún modo que las últimas personas del espectro serían numéricamente idénticas a la persona original. Parfit argumenta que si rechazamos el reduccionismo, nuestra única opción es afirmar que existe una línea divisoria estricta en el espectro combinado y que, por lo tanto, la diferencia entre sobrevivir o morir depende de unos cambios mínimos. Según el autor, si no somos entidades que existen separadamente no puede defenderse que nuestra identidad tiene que ser siempre determinada, puesto que si no somos ese tipo de entidades, no hay nada que pueda trazar una línea divisoria en el espectro combinado. Parfit considera la posibilidad de que se piense que, aunque no seamos entidades que existen separadamente, la identidad personal consiste en un hecho adicional, en lugar de consistir únicamente en las continuidades física o psicológica. Sin embargo, contra esta opinión el autor argumenta que si no somos entidades que existan separadamente, no hay nada en que pueda consistir ese hecho adicional y que pueda trazar una línea divisoria nítida en el espectro combinado.

Parfit utiliza nuevamente el caso del teletransporte como ejemplo. Después de utilizar la máquina, habría una persona que sería físicamente idéntica a la original y psicológicamente continua con ella. Dicha continuidad psicológica no tendría la continuidad física del cerebro como causa, es decir, no tendría su causa normal. Podríamos preguntarnos si esa persona y la original serían sólo cualitativamente idénticas o si serían también numéricamente idénticas, pero Parfit afirma que, en este caso, esas no son dos opciones diferentes. Si sabemos que la réplica es físicamente idéntica a la persona original, que es psicológicamente continua con ella y que esto último no se debe a la continuidad del cerebro, entonces conocemos todos los hechos. La

pregunta de si la réplica es numéricamente idéntica al original es vacía, las diferentes respuestas son sólo distintas descripciones de una misma situación⁴². Sucede lo mismo en los casos centrales del espectro combinado. En dichos casos, la identidad personal es indeterminada y las distintas respuestas a la pregunta por la identidad son sólo diferentes descripciones de la misma situación.

Geoffrey Madell (1985) considera que el argumento del espectro combinado no logra demostrar que la identidad personal pueda ser indeterminada. Madell considera que, aunque Parfit afirme que el espectro combinado es sólo técnicamente imposible, en realidad no podemos saber si realmente los estados psicológicos pueden ser transferidos de una persona a otra⁴³. En opinión del autor, aunque la situación que Parfit propone pudiera llevarse a cabo, no podemos estar seguros de que el resultado fuera el que propone Parfit. Tal vez cuando las células del cerebro fueran reemplazadas de manera que se modificaran las características psicológicas el resultado no fuera una mezcla de unas características y otras. Tal vez en los casos intermedios del espectro combinado la persona no fuera un caso intermedio de las personas de los casos primero y último. El autor propone que el resultado podría ser, por ejemplo, una persona con doble personalidad, las personalidades de las personas del principio y el final de la serie. Madell añade que, incluso en el caso de que los resultados fueran los que supone Parfit, ello no apoyaría su reduccionismo ni su afirmación de que la identidad personal no es siempre determinada. El autor afirma que cuando nos imaginamos a nosotros mismos sufriendo algún tipo de cambio psicológico, nos estamos imaginando que nos sucede a nosotros, y no a alguien que sea una combinación entre nosotros y otra persona. El autor concluye que no podemos concebir la idea “de que hay un mundo posible en que hay un estado de conciencia presente que es mío sólo en parte”⁴⁴ (Madell, 1985, p. 108). Sin embargo, Madell comete varios errores en sus objeciones al argumento de Parfit. En primer lugar, hay que señalar que la afirmación de Parfit de que la imposibilidad de su experimento mental es sólo técnica resulta enormemente plausible, y lo mismo sucede con los resultados que imagina. Si escogemos dos personas, podemos concebir a otra cuyas características psicológicas estén en un punto intermedio entre las de ambas. Es razonable suponer que existe algún estado cerebral posible que traería consigo las características

⁴² En *Razones y personas* el autor no termina de dejar claro si en los casos en que hay continuidad psicológica sin la causa normal la identidad personal es o no indeterminada. Al argumentar en favor de su afirmación de que la identidad personal no es lo que importa, parece suponer que si hay continuidad psicológica, hay identidad personal sea cual sea la causa. No obstante, en un trabajo posterior, Parfit (1986) parece dar por sentado que si la continuidad psicológica no tiene su causa normal la pregunta por la identidad personal es vacía. El presente ejemplo también parece ir en esa dirección.

⁴³ Parfit, al formular el experimento mental del espectro combinado, imagina que en los últimos casos la persona sería idéntica a una persona real distinta de la original. Al describir el experimento mental he omitido esta característica porque no es necesaria para demostrar que la identidad no es siempre determinada y, como vemos con la objeción de Madell, puede ser usada para socavar la plausibilidad de dicho experimento mental.

⁴⁴ La traducción es mía.

psicológicas de esta tercera persona que estamos imaginando. Si el cerebro de una persona hubiera sido parcialmente alterado para parecerse más al de otra, es más plausible suponer que el resultado sería un caso intermedio de la psicología de ambos que suponer que sería un caso de doble personalidad. No obstante, podemos modificar el experimento mental de Parfit afirmando que se actúa artificialmente sobre la actividad cerebral de la persona para asegurar que sus características psicológicas son un caso intermedio de las de los dos extremos. Puede que un caso así siga siendo siempre imposible, pero, si aceptamos que los estados mentales dependen de la actividad cerebral, la suposición de que tal imposibilidad es sólo técnica resulta muy razonable. En cuanto a la última de las objeciones de Madell, tiene razón al señalar que no podemos concebir, desde la perspectiva de primera persona, el caso de alguien que sea nosotros sólo en parte o cuya identidad numérica con nosotros sea indeterminada. Bernard Williams (1986e), al exponer el experimento mental en que una persona sufre ciertos cambios psicológicos antes de sufrir cierto dolor, nos dice lo siguiente:

Que se me diga que una situación futura es una situación fronteriza respecto a que sea yo el que sufra, y que es conceptualmente indecible si seré yo o no, es algo con lo que, parece, no puedo hacer nada; porque, en particular, no parece tener ninguna representación comprensible en mis expectativas ni en las emociones que las acompañan. (pp. 81-82)

Esto es así debido al modo en que anticipamos nuestras experiencias futuras. Como señala Judith Jarvis Thomson (1997), el motivo por el que creemos que no puede ser indeterminado si una persona ha sobrevivido o no es porque lo que estamos preguntando en realidad es qué está experimentando la persona original, y si es indeterminado si dicha persona existe o no, no puede haber ninguna respuesta que pueda parecernos satisfactoria. Cuando imaginamos lo que será para nosotros vivir una determinada experiencia, lo que hacemos es imaginarnos teniendo esa experiencia. Lo que hacemos es, por lo tanto, imaginar que esa experiencia nos sucede ahora. Sólo podemos imaginar que la experiencia nos sucede o que no nos sucede, no podemos imaginar que es indeterminado si la experiencia nos sucede o no. Podemos imaginar que tenemos una versión atenuada de dicha experiencia, pero eso no serviría. No serviría porque entonces estaríamos imaginando que tenemos una experiencia más débil, no que la experiencia le suceda a alguien cuya identidad con nosotros es indeterminada. No hay nada sorprendente en esta incapacidad para concebir la situación en que nuestra identidad con quien tiene la experiencia es indeterminada. Cuando anticipamos una experiencia futura, simplemente nos imaginamos teniendo la experiencia. Podemos tratar de imaginarnos teniendo esa experiencia más los cambios psicológicos en cuestión, pero ello, aunque supongamos que pudiéramos tener éxito, tampoco resolvería la cuestión, no es eso lo que Madell dice que no podemos concebir. Lo que se supone que no podemos concebir es la situación en que es indeterminado si alguien

que va a tener una experiencia es nosotros o no. Pero el problema no es concebirlo desde el punto de vista de tercera persona, puesto que para ello sólo tendríamos que imaginar una situación en que la experiencia es tenida por alguien que es en parte como nosotros y en parte como otra persona. El problema, como señala Madell, es concebir la situación desde la perspectiva de primera persona. Concebir la situación desde la perspectiva de primera persona consiste en anticipar las experiencias de aquella persona que vaya a ser numéricamente idéntica a nosotros. No podemos concebir desde la perspectiva de primera persona una época en la que nosotros ya no existamos, puesto que en esa época no hay nadie numéricamente idéntico a nosotros y, por lo tanto, no hay nada que anticipar. Cuando tratamos de concebir desde la perspectiva de primera persona una situación en que hay alguien cuya identidad con nosotros es indeterminada sucede algo distinto, lo que sucede es que no sabemos si tenemos que anticipar o no sus experiencias. Si concebir una situación futura desde la perspectiva de primera persona consiste en anticipar las experiencias de aquella persona que sea idéntica a nosotros y en una determinada situación es indeterminado si habrá o no una persona idéntica a nosotros, es natural que sea indeterminado si debemos anticipar sus experiencias o no. No hay nada sorprendente en que seamos incapaces de concebir la situación, que esto suceda es parte de aquello en lo que consiste que sea indeterminado si sobreviviremos o no. Volvamos por un momento al caso de la línea secundaria. En ese caso hay dos personas que son psicológicamente continuas con la original. En el caso de una de ellas, la continuidad psicológica tendría su causa normal, la continuidad física del cerebro. En el caso de la otra, no habría la causa normal. Antes de apretar el botón, la persona podría imaginarse como siendo cualquiera de las dos personas resultantes. Tras usar la máquina, ambas personas creerían ser la original. La pregunta de cuál de las dos personas es aquella cuyas experiencias debería anticipar la original es simplemente la pregunta de cuál de las dos es la original. La pregunta es vacía, conociendo los hechos que acabamos de describir lo sabemos todo. Decir que la persona original no puede concebir, desde la perspectiva de primera persona, su situación tras apretar el botón significa únicamente que la persona original no sabe cuál de las dos personas será ella. Puesto que es indeterminado cuál de las dos personas será ella, es natural que no lo sepa. No hay nada misterioso en su incapacidad para anticipar su situación tras la replicación.

Howard Curzer (1991) cree detectar una ambigüedad en la teoría de la identidad personal de Derek Parfit. Según el autor, con algunas de sus afirmaciones Parfit parece implicar que la identidad personal es siempre indeterminada, mientras que con otras parece querer decir que esta sólo es indeterminada en ciertos casos, como, por ejemplo, en los casos centrales del espectro combinado. En opinión de Curzer, la afirmación de que la identidad personal es siempre indeterminada se sigue del reduccionismo, y Parfit aplica esta teoría a todos los casos, no sólo a los casos centrales del espectro combinado. Según Curzer, si la identidad personal no

consiste en ningún hecho adicional, entonces la pregunta es vacía en todos los casos, puesto que si ya conocemos los hechos sobre la continuidad física y psicológica, responderla no nos daría información adicional. Sin embargo, Curzer comete aquí un error. Es cierto que, de acuerdo con el reduccionismo de Parfit, si conocemos todos los hechos sobre las continuidades física y psicológica, responder a la pregunta sobre la identidad no nos aportaría información nueva, pero eso no significa que la pregunta sea vacía en el sentido en que Parfit usa esta expresión. Según Parfit, en los primeros y los últimos casos del espectro combinado hay una única respuesta correcta a la pregunta por la identidad. Si conocemos todos los hechos físicos y mentales relacionados con esos casos, responder a la pregunta sobre la identidad no nos dará información nueva, responder a la pregunta sería redundante. Esto es así porque la respuesta a la pregunta ya se encuentra en esos hechos físicos y mentales que ya conocíamos. Pero nada de esto implica que la pregunta sea vacía. La pregunta de si en determinado lugar hay o no un bosque no es una pregunta vacía. Que en cierto lugar haya un bosque o no lo haya son dos alternativas claramente distintas. Si supiéramos cuántos árboles hay en ese lugar y cómo están distribuidos, responder a la pregunta de si hay o no un bosque no nos aportaría información nueva. No obstante, eso no significa que la pregunta sea vacía, significa simplemente que la respuesta ya está incluida en los hechos que conocemos. Decir que hay tal número de árboles en cierta zona y decir que hay un bosque son dos maneras distintas de responder a la pregunta de si hay o no un bosque. La pregunta no es vacía, sencillamente si hemos respondido de la primera manera resulta redundante responder también de la segunda. La opinión de Parfit parece ser que esto mismo es lo que sucede con la identidad personal. En un caso en que la identidad no es indeterminada, describir los distintos sucesos físicos y mentales es una manera de responder a la pregunta por la identidad. Responder también de manera directa diciendo si la persona es idéntica o no a la original no añade nada si ya conocemos dichos hechos, es redundante. No obstante, ello no significa que la pregunta sea vacía, significa que ya la habíamos respondido. En los casos intermedios del espectro combinado, en cambio, describir dichos hechos no responde a la pregunta. En esos casos, dichos hechos encajan con ambas respuestas. Las dos respuestas posibles son sólo dos maneras de describir dichos hechos. En esos casos, la pregunta sí que es vacía, no tiene respuesta. No hay, por lo tanto, ninguna ambigüedad en la teoría de Derek Parfit. En los casos intermedios del espectro combinado, la identidad personal es indeterminada. En los primeros y los últimos casos, en cambio, la identidad personal es, según Parfit, determinada. En mi opinión, sin embargo, Parfit está equivocado y la identidad personal es siempre indeterminada. Curzer considera implausibles ambas afirmaciones. Antes de defender que la identidad es siempre indeterminada examinaré las objeciones de Curzer contra la que en mi opinión es claramente la postura de Parfit, es decir, contra la idea de que, en el espectro combinado, sólo los casos intermedios son indeterminados.

Curzer señala que si aceptamos que hay una zona vacía en la parte central del espectro combinado, tenemos dos opciones. Podemos considerar que las fronteras de esa zona vacía son nítidas o podemos considerar que son borrosas. Curzer señala que los argumentos que Parfit emplea contra la posibilidad de que la identidad personal sea siempre determinada son también aplicables contra la presencia de fronteras nítidas en los extremos de la zona vacía. El autor argumenta entonces que la idea de que existe un área vacía en el espectro combinado delimitada por dos fronteras borrosas es incoherente. El argumento empleado por Curzer es el siguiente. Supongamos que A es una persona situada al principio del espectro combinado y que B es una persona situada en la zona borrosa que hay entre la zona vacía y una de las otras dos zonas. Puesto que no es cierto que B esté en la zona vacía, no es cierto tampoco que responder a la pregunta de si A y B son la misma persona no vaya a proporcionarnos información. Si esto último es así, B tiene que estar en el área de las personas que son numéricamente idénticas con A o en el área de las que no lo son. Entonces, no hay área borrosa y el área vacía tiene fronteras nítidas, con lo cual llegamos a una contradicción. Curzer considera que la mejor solución es considerar que en el espectro combinado hay sólo dos áreas, la de las personas idénticas a la original y la de las personas distintas a ella, separadas por una frontera borrosa. Según el autor, esta zona borrosa sería más pequeña que la zona vacía que postula Parfit, lo cual nos llevaría a la conclusión de que la pregunta por la identidad sólo puede ser vacía cuando la persona empieza a parecerse mucho a la persona del final de la serie. Curzer propone otro argumento que lleva a la misma conclusión. Supongamos de nuevo que el área vacía tiene fronteras borrosas. Entonces, si cogemos un punto de una de las dos áreas borrosas y preguntamos si ese caso se encuentra en la zona vacía, estaremos realizando una pregunta vacía. Entonces, decir que ese caso pertenece a la zona vacía y decir que no pertenece serán, de acuerdo con la definición que Parfit da de lo que es una pregunta vacía, sólo dos descripciones distintas de una misma situación. Por lo tanto, podríamos saberlo todo aunque no pudiéramos responder a la pregunta. Entonces, el autor argumenta que si por “todo” entendemos todas las propiedades que dicho caso comparte con la persona original, entonces dicho caso está en la zona vacía. Si, en cambio, “todo” incluye también la respuesta a la pregunta de si esa persona es la misma que la original, entonces dicho caso está en una de las dos zonas no vacías. Habíamos supuesto que ese caso estaba en una de las dos zonas borrosas, por lo que llegamos de nuevo a una contradicción.

Curzer comete varios errores en su argumentación. En primer lugar hay que señalar que si suponemos que en el espectro combinado hay únicamente dos zonas separadas por una frontera borrosa, nos encontramos con el mismo problema. Podemos preguntar si la zona borrosa está delimitada por fronteras nítidas o por fronteras borrosas. Si las fronteras son nítidas, podemos aplicar los argumentos de Parfit de los que habla Curzer. Si las fronteras son borrosas, nos encontramos con el mismo problema que detecta Curzer en la teoría de Parfit. De hecho, si las

fronteras son borrosas podríamos preguntarnos de nuevo si estas nuevas zonas borrosas están delimitadas por fronteras nítidas o borrosas, y así podríamos seguir indefinidamente. Este fenómeno es propio de todos los predicados vagos y se conoce como vaguedad de orden superior. Lo que Parfit pone de manifiesto al defender que la identidad no es siempre determinada es que el predicado “ser la misma persona que A” (donde A es una persona cualquiera) es vago. Si tenemos esto presente, es fácil darse cuenta de que la alternativa que propone Curzer no es realmente distinta de la de Parfit. Con cualquier predicado vago podemos optar por definir un nuevo predicado para la zona borrosa. Los predicados “alto” y “bajo”, por ejemplo, son predicados vagos. Los casos a que se aplican uno o el otro no están separados por una frontera nítida. Podemos optar por referirnos a los casos situados en la zona borrosa como medianos. Por supuesto, los casos medianos también estarán separados de los casos bajos y de los altos por fronteras borrosas. Algo similar es lo que hace Parfit. Este se refiere a la zona borrosa como zona vacía. En cierto modo, podríamos considerar que define un nuevo predicado para los casos situados entre los que son idénticos y los que son distintos con respecto a A. Define, en cierto sentido, un nuevo predicado para los casos cuya identidad con A es indeterminada. Eso, sin embargo, no es una alternativa distinta a lo que propone Curzer. Podemos definir un nuevo predicado o no hacerlo, pero esas sólo son dos maneras distintas de conceptualizar la misma situación. No estamos ante dos teorías distintas, sino ante dos maneras distintas de enunciar una misma teoría. El autor comete también el error de suponer que la zona borrosa entre dos predicados vagos tiene que ser necesariamente pequeña. Eso es un error, los casos de la zona borrosa serán todos aquellos a los que no sepamos cuál de los dos predicados aplicar. Dicha zona puede ser grande o pequeña y, de hecho, Parfit no se pronuncia en ningún momento sobre su tamaño. Ocupémonos ahora de la supuesta incoherencia que Curzer cree detectar en la teoría de Parfit. Centrémonos primero en el segundo de sus argumentos. El autor comete el error de suponer que si saberlo todo se refiere a conocer las relaciones físicas y psicológicas entre ambas personas, entonces nos encontramos en la zona vacía; mientras que si saberlo todo incluye también conocer la respuesta sobre la identidad, entonces estamos en una de las otras dos zonas. Esta suposición es errónea. Según la concepción reduccionista, si conocemos todas las relaciones físicas y psicológicas entre las dos personas en cuestión, entonces lo sabemos todo, estemos en la zona que estemos. Sencillamente, si estamos en una de las zonas no vacías, dichas relaciones contienen una respuesta a la pregunta por la identidad. Si nos encontramos en una de las zonas borrosas, será indeterminado si dichas relaciones contienen o no una respuesta, pero de ello no se sigue ninguna contradicción. Curzer comete en este argumento otro error que también comete en el primero. Puesto que es más fácil detectar dicho error en el primero de sus argumentos y puesto que aún tenemos que refutar dicho argumento, será mejor que nos centremos en él. Este argumento se basa en una falta de comprensión, por parte de Curzer, de lo que implican los órdenes superiores de vaguedad. El argumento se basa

en la suposición del autor de que si un caso está en el área borrosa, entonces no está en la zona vacía. Esa suposición es equivocada. Si un caso está en una de las áreas borrosas, es indeterminado si está en el área vacía o no. Si nos damos cuenta de esto, no llegamos a ninguna contradicción.

No obstante, he afirmado que, según mi opinión, Parfit está equivocado y la identidad personal es siempre indeterminada. Para ver por qué esto es así, necesitamos explorar más a fondo la relación entre la vaguedad lingüística y la identidad personal. He afirmado que lo que demuestra Parfit con su espectro combinado es que el predicado “ser la misma persona que A” es vago. A qué conclusiones nos conduzca esto dependerá de cuál creamos que es la concepción que mejor explica la vaguedad lingüística. La vaguedad lingüística se ha tratado de explicar de diferentes maneras, y cada una de ellas implica una respuesta distinta a la pregunta de qué casos del espectro combinado son indeterminados. La vaguedad lingüística es conocida por dar lugar a la paradoja sorites. Si utilizamos como ejemplo el predicado “bajo”, podemos imaginar una serie de personas ordenadas de menor a mayor altura tales que la primera persona sea claramente baja, la última sea claramente alta y entre una persona y la siguiente no haya la suficiente diferencia de altura como para marcar la diferencia entre ser bajo y dejar de serlo. Si la primera persona es baja y de una persona a la siguiente no hay una diferencia de altura suficiente para dejar de ser bajo, llegamos a la conclusión de que la última persona de la serie es baja, pero la última persona es alta y, por lo tanto, nos encontramos con una contradicción. El espectro combinado que plantea Parfit no es otra cosa que una serie sorites construida con el predicado “ser idéntico a la persona original”.

Peter Unger (1979b) defiende que la vaguedad lingüística demuestra que los conceptos vagos son incoherentes y que las propiedades que expresan no existen. Para Unger, la paradoja sorites, utilizando el ejemplo anterior, es una demostración de que no hay personas bajas. Unger emplea un argumento, al que se refiere como el sorites de la descomposición, en el que imagina que se fueran eliminando una a una las partículas de cualquier objeto. Puesto que una sola partícula no marca la diferencia, nos vemos obligados a concluir que cuando ya no quedan partículas el objeto sigue estando ahí y entonces, por reducción al absurdo, tenemos que concluir que la suposición de que el objeto existía era errónea. De este modo, Unger concluye que no existen las cosas ordinarias y que lo único que hay en realidad son aquellas partículas a que el sorites de la descomposición no puede aplicarse. En otro trabajo, Unger (1979a) utiliza el sorites de la descomposición para defender que las personas no existen. Por lo tanto, si la concepción de la vaguedad de Unger fuera cierta, deberíamos concluir que las personas no existen y que, por lo tanto, no tiene sentido plantear la cuestión de la identidad personal. El autor reconoce la aparente incoherencia de afirmar que él mismo no existe, pero atribuye la causa de tal

incoherencia a la inconsistencia de nuestros conceptos que, según él, queda demostrada por la paradoja sorites. Hilary Putnam (1983) considera que la posición mantenida por Unger es incoherente. Putnam argumenta que, puesto que Unger no puede evitar emplear multitud de términos vagos al enunciar su nihilismo ontológico, sus argumentos son aplicables también a su propia teoría. Si los conceptos vagos son inconsistentes, entonces, puesto que gran parte de los términos de nuestro lenguaje son vagos, no podemos decir nada que tenga sentido y Unger tampoco ha podido hacerlo al enunciar su postura. Por lo tanto, la teoría de Unger se autorrefuta. Tal vez Unger podría reconocer esta incoherencia, del mismo modo que reconoce la de afirmar que él mismo no existe, y atribuirlo a la inconsistencia de nuestros conceptos, pero ello no convertiría en menos problemático afirmar que lo que se está afirmando es incoherente. Hay muchas otras teorías que tratan de ofrecer una solución a la paradoja, y las consecuencias de todas ellas son mucho más plausibles que las del nihilismo ontológico de Unger.

Otra de las concepciones de la vaguedad que han sido propuestas es el epistemicismo. Según este punto de vista, uno de cuyos principales defensores es Timothy Williamson (1997), la vaguedad no es otra cosa que cierto tipo de ignorancia por nuestra parte. Los casos fronterizos son, según esta teoría, casos a los que no sabemos si el predicado se aplica o no. No hay, en opinión de los defensores de este punto de vista, ningún tipo de indeterminación, sólo la imposibilidad de saber si el predicado puede o no aplicarse. Podría pensarse que los defensores de esta teoría, en caso de que la aplicaran a la identidad personal, deberían pensar que somos entidades que existen separadamente, pero eso sería un error. Como señala Keith Hossack (2006), el tipo de ignorancia que el dualismo cartesiano atribuiría a los casos intermedios del espectro combinado no es el mismo que el que les atribuiría el epistemicismo. Como señala el autor, los dualistas cartesianos no creen que en el espectro combinado haya casos fronterizos a los que no sepamos si un concepto se aplica o no. Según los dualistas cartesianos, lo que no sabemos es si la persona resultante sigue teniendo el mismo ego cartesiano y si lo supiéramos, desaparecería nuestra ignorancia. Si aplicáramos el epistemicismo a la cuestión de la identidad personal podríamos, por ejemplo, sostener que el reduccionismo es cierto pero que, aun así, la identidad personal es siempre determinada. Si fuéramos epistemicistas deberíamos pensar que podemos saber todo lo que hay que saber sobre los casos centrales del espectro combinado y aun así ignorar qué respuesta deberíamos dar a la pregunta por la identidad. Esto no se debería a ningún tipo de indeterminación en la identidad personal, se debería a cierta ignorancia por nuestra parte sobre la aplicación del concepto de identidad⁴⁵. Sin embargo, en mi opinión, el epistemicismo resulta implausible. No parece estar claro a qué podría deberse esa ignorancia

⁴⁵ Parfit considera que si creemos que no somos entidades que existen separadamente, no podemos defender que la identidad personal ha de ser siempre determinada. Las presentes consideraciones demuestran que, contra la opinión de Parfit, dicha posición es defendible al menos en la medida en que el epistemicismo lo es.

que los epistemicistas postulan. No parece haber ningún tipo de información que pudiera llevarnos a saber si un predicado vago se aplica o no a un caso fronterizo. Parece más razonable creer que la causa de la vaguedad se encuentra en el significado de los términos vagos. Las teorías que buscan la causa de la vaguedad en el significado de los predicados vagos son las teorías semánticas de la vaguedad.

De entre estas teorías, la mayoría encajan con el punto de vista de Parfit de que sólo los casos intermedios del espectro combinado son indeterminados. Algunos autores consideran que la mejor opción es añadir un tercer valor de verdad al que se suelen referir como “indefinido” o “indeterminado”. Las frases que atribuyen un predicado vago a un caso fronterizo recibirían este tercer valor de verdad. Es fácil ver que esta concepción de la vaguedad apoya el punto de vista de Parfit de que en los casos intermedios del espectro la identidad personal es indeterminada. Los defensores de la teoría de los grados de verdad consideran que hay infinitos valores de verdad. Las frases que atribuyen un predicado vago a los primeros y los últimos casos de una serie sorites serían completamente verdaderas o completamente falsas, mientras que a las que los atribuyen a los casos intermedios les correspondería un grado intermedio de verdad. De nuevo, esta teoría apoyaría el punto de vista de Parfit y, además, encajaría plenamente con su afirmación de que la identidad personal es una cuestión de grado. Otras dos teorías que también encajan con la concepción de Parfit son el superevaluacionismo y el subevaluacionismo. Según la primera, los casos fronterizos de un predicado vago no son ni verdaderos ni falsos, mientras que según la segunda son verdaderos y falsos a la vez⁴⁶. Si alguna de estas teorías fuera cierta, tendría que dar un tratamiento adecuado a la vaguedad de orden superior, de modo que esta dejaría de ser un problema para la concepción de Parfit.

En mi opinión, sin embargo, la mejor explicación de la vaguedad lingüística es la ofrecida por el contextualismo no indécico⁴⁷. Una teoría de este tipo es la defendida por Delia Graff Fara

⁴⁶ Estas dos teorías emplean la noción de precisión admisible. Las distintas precisiones son las distintas maneras admisibles de hacer el predicado completamente preciso. Cada precisión asigna a cada caso fronterizo un valor de verdad determinado. Son admisibles aquellas que realicen esta asignación respetando las conexiones analíticas existentes entre dichos casos. Por ejemplo, no sería admisible una precisión del predicado “bajo” que tratara como determinadamente bajo a cierto caso y como determinadamente alto a otro cuya altura fuera menor. El superevaluacionismo define la verdad como superverdad (es decir, como verdad en todas las precisiones admisibles) y la falsedad como superfalsedad. Puesto que las frases que aplican el predicado a los casos fronterizos son verdaderas en unas precisiones y falsas en otras, para el superevaluacionismo no son ni verdaderas ni falsas. El subevaluacionismo, en cambio, define la verdad como subverdad (es decir, como verdad en al menos una precisión admisible) y la falsedad como subfalsedad. Por lo tanto, para el subevaluacionismo las frases que aplican un predicado vago a casos fronterizos son tanto verdaderas como falsas.

⁴⁷ No puedo examinar aquí los argumentos en favor y en contra de cada una de estas teorías, puesto que ello requeriría dedicar demasiado espacio al tema de la vaguedad lingüística. Sin embargo, ofreceré algunos argumentos en apoyo de las conclusiones que el contextualismo no indécico supone para la cuestión de la identidad personal. Diana Raffman (1994) desarrolla una teoría similar a la de Fara en favor de la cual ofrece poderosos argumentos.

(2000). De acuerdo con esta teoría⁴⁸, según recorremos la serie sorites va variando la extensión del predicado vago en cuestión. La propiedad expresada por el predicado es siempre la misma, pero se trata de una propiedad relativa al hablante⁴⁹. El valor de verdad de dicha proposición iría variando con el tiempo, a causa de los cambios psicológicos sufridos por el hablante. Los defensores del contextualismo no indécico suelen hablar del contexto interno o psicológico del hablante, que está en constante cambio y que determina cuál es la extensión de la propiedad expresada por cada predicado vago⁵⁰. Si aplicáramos el contextualismo no indécico a la cuestión de la identidad personal, tendríamos que concluir que el predicado “ser numéricamente idéntico a la persona A” expresa una propiedad relativa al contexto psicológico del hablante. En mi opinión, esto debe interpretarse en el sentido de que al decir que una persona es idéntica a otra estamos en parte expresando una impresión personal⁵¹, de modo que, aunque los primeros y los últimos casos del espectro suelen causarnos a todos la misma impresión, debemos concluir que, estrictamente hablando, la identidad personal es siempre indeterminada. La mayoría de partidarios del contextualismo no indécico no lo entienden de este modo. Estos autores considerarían que los únicos casos indeterminados del espectro combinado son los casos centrales. Esto es así porque, desde el punto de vista del contextualismo no indécico, se considera que una proposición es determinadamente verdadera o falsa cuando lo es en cualquier contexto psicológico, mientras que cuando es verdadera en unos contextos y falsa en otros es indeterminada. Sin embargo, un predicado vago sigue expresando una propiedad relativa al contexto psicológico del hablante y el hecho de que una proposición que emplee un predicado de ese tipo sea verdadera en todos los contextos psicológicos no es necesario sino contingente. No puede descartarse a priori, para ningún predicado vago, la existencia de un contexto psicológico en el cual una proposición que lo emplea y que se consideraba determinadamente verdadera sea falsa o a la inversa. El motivo por el que hay muchas proposiciones que, a pesar de contener predicados vagos, tienen el mismo valor de verdad en todos los contextos psicológicos es que, debido a que los distintos seres humanos tenemos una mente similar, los distintos contextos psicológicos son también similares. No obstante, en sentido estricto, lo más adecuado parece ser afirmar que las proposiciones cuya verdad depende de predicados vagos son siempre indeterminadas. Cuando afirmo que la identidad personal es siempre indeterminada

⁴⁸ Sigo, en la mayor parte de la explicación de la teoría de Fara, la exposición que de ella realiza Jason Stanley (2003).

⁴⁹ Según la teoría de Fara, la frase “la montaña es alta para una montaña” expresa que “la montaña tiene una altura significativamente mayor para el hablante que la altura típica de las montañas”.

⁵⁰ Según el contextualismo indécico, lo que va cambiando no es sólo la extensión de la propiedad expresada por los predicados vagos sino el propio contenido de dichos predicados. Sin embargo, en mi opinión, Stanley (2003) argumenta convincentemente en contra de este tipo de contextualismo.

⁵¹ En uno de sus artículos, Parfit (1971a) realiza unas afirmaciones que parecen estar cerca de esta idea. El autor afirma que si una persona no se identifica con su yo pasado, su afirmación de que no es esa persona pasada no puede ser literalmente falsa. Según Parfit, tales afirmaciones expresan en parte una falta de conexividad y en parte la actitud del hablante hacia dicho hecho.

lo digo en este sentido, es decir, en el sentido de que depende siempre del contexto psicológico del hablante o del pensador, a causa de lo cual podrían siempre, en principio, existir dos hablantes que realizaran juicios de identidad personal contrarios sin que estuviera equivocado ninguno de los dos. Esto no significa que podamos decidir libremente qué persona presente es idéntica a cierta persona pasada, significa simplemente que la verdad de un juicio de identidad depende en parte del contexto psicológico de quien lo realiza.

Curzer (1991) argumenta también contra la posibilidad de que la identidad personal sea indeterminada en todos los casos del espectro combinado. El autor argumenta que si la identidad es siempre indeterminada, entonces cualquier mínimo cambio que yo pueda sufrir convierte en vacía la pregunta de si la persona resultante sigue siendo yo. De hecho, suponer que la identidad personal es siempre indeterminada convierte también en vacía la pregunta de si yo soy Napoleón. Como señala el autor, desde este punto de vista lo único que hay sobre la cuestión es que yo comparto ciertas características con Napoleón y difiero de él en otras. Estas consecuencias pueden resultar algo chocantes, pero si se las considera detenidamente, no son tan implausibles como parecen. Según el punto de vista que estoy defendiendo, a diferencia de lo que sucede con el que defiende Parfit, la respuesta a la pregunta por la identidad expresa algo que no está contenido en los sucesos físicos y mentales relacionados con las personas involucradas, expresa la impresión que dichos sucesos nos causan. Por lo tanto, el hecho de que sea indeterminado si yo soy o no Napoleón sólo significa que no hay una respuesta a la pregunta que sea objetiva e independiente de cualquier hablante. La respuesta a la pregunta siempre expresará las impresiones del hablante. Por supuesto, la mente de todos nosotros es en cierto grado similar, de modo que muchos hechos nos causan a todos impresiones muy similares. Sería difícil encontrar a alguien que, tras examinar los sucesos físicos y mentales relacionados conmigo y compararlos con los relacionados con Napoleón, concluyera que ambos somos la misma persona. Puesto que tenemos razones para creer que dichos hechos no causarían tal impresión a nadie, tenemos motivos para creer que yo no soy Napoleón relativamente al contexto psicológico de nadie y, por lo tanto, si alguien dijera que yo soy Napoleón, sería razonable pensar que está diciendo algo falso. Puede parecer implausible la idea de que cualquier pequeño cambio que yo sufra convierta en vacía la pregunta de si la persona resultante es yo, pero, aunque se trate de una descripción infrecuente de los hechos, no hay motivos para creer que se trataría de una descripción falsa. Alguien, tras comparar los sucesos relacionados con mi yo pasado con aquellos relacionados con mi yo actual podría llevarse la impresión de que no somos la misma persona. Cuando Heráclito dijo que nadie puede bañarse dos veces en el mismo río, no parece que estuviera diciendo algo falso, más bien parece que estaba llevando a cabo una determinada descripción de los hechos. Somos conscientes de que los elementos que forman un río cambian constantemente, pero vemos cierta continuidad en él. Si somos

conscientes de estos hechos, decir que se trata siempre del mismo río o que, tras un instante, el río es otro distinto, no son más que dos formas distintas de describir la misma situación, dos descripciones distintas que reflejan las impresiones que dichos hechos nos causan. Lo mismo sucede con la identidad personal. Igual que en el caso del río, en las personas se van produciendo ciertos cambios a pesar de los cuales percibimos cierta continuidad. Podemos decir que la persona sigue siendo la misma o que no sigue siéndolo, pero estas no son dos posibilidades metafísicas distintas, son sólo dos descripciones distintas de la situación que reflejan la impresión que tal situación nos causa.

Las continuidades que podemos observar en una persona no son las mismas cuando se trata de otra persona que cuando se trata de nosotros. En lo referente a nosotros mismos, contamos con un punto de vista privilegiado sobre nuestra propia continuidad psicológica. Puede ser que Locke (1956) estuviera pensando en algo así cuando decía que la memoria es un criterio de identidad personal para uno mismo. Si una persona pierde totalmente la continuidad psicológica, podría resultar que quienes han sido testigos de su continuidad física le considerasen la misma persona pero que él, que es directamente consciente de haber perdido la continuidad psicológica, se considerase a sí mismo como una persona distinta. Bernard Williams (1986b) no considera que la idea de que la memoria es un criterio de identidad para uno mismo pueda tener sentido. El autor señala que cuando el criterio presumiblemente se tendría que aplicar es cuando la persona actual tiene un conjunto de recuerdos distintos a los que tenía antes. Sin embargo, la persona no puede tener ambos conjuntos a la vez, ya que si los tuviera o si los primeros estuvieran contenidos en los segundos, los dos conjuntos no serían diferentes en ningún sentido relevante. Puesto que la persona no puede tener ambos conjuntos de recuerdos a la vez, sólo puede recordar que tuvo recuerdos que ahora no tiene, y la única manera de comparar ambos recuerdos es recurriendo al testimonio de otras personas sobre su pasado, de modo que estaríamos recurriendo al criterio que emplean las otras personas. Sin embargo, aunque Williams tiene razón, el problema desaparece si tenemos en cuenta no solo la memoria sino también el resto de conexiones psicológicas. Una persona puede recordar haber tenido un carácter distinto, unos gustos distintos o unas intenciones distintas y puede juzgar, en base a estos datos, si la continuidad psicológica se da en un grado lo suficientemente elevado como para que haya identidad. Que la identidad personal sea siempre indeterminada, aun así, no impide que puedan existir respuestas que sean siempre incorrectas. Sería incorrecta cualquier respuesta que fuera autocontradictoria y lo serían también, por ejemplo, aquellas respuestas que violaran la lógica de la identidad⁵².

⁵² Puede parecer que la idea de que en ciertos casos hay una que es la mejor descripción o la de que puede haber descripciones inaceptables contradice la afirmación de que la pregunta por la identidad es vacía. En uno de sus artículos, Parfit (1993) considera esta objeción y responde que afirmar que una determinada

Podría pensarse que si lo único que está involucrado en nuestra identidad son ciertos sucesos físicos y mentales interrelacionados, deberíamos concluir que las personas no existen. Podría pensarse que lo único que existe son dichos sucesos físicos y mentales. Este es el punto de vista del reduccionismo eliminativo. Parfit (2003) considera esta posibilidad, que atribuye al budismo, pero la rechaza en favor de su versión del reduccionismo, a la que se refiere como reduccionismo constitutivo. El autor argumenta que la realidad con respecto a cosas como las naciones, los clubs y las personas se expresa mejor afirmando que tales cosas existen pero que consisten solo en aquellos elementos y sucesos que los constituyen. Un barco, por ejemplo, no es algo que exista separadamente de sus componentes, pero ello no debe llevarnos a afirmar que los barcos no existen. Si llamamos personas a ciertos elementos y a ciertos sucesos físicos y mentales relacionados con ellos, no hay problema en afirmar que las personas existen, siempre que por personas no entendamos algún tipo de entidad que existe separadamente.

La más importante de las afirmaciones de Parfit es la de que la identidad personal no es lo que importa. Según Parfit, lo que importa es el darse de la relación R, con cualquier causa. Dicha afirmación atenúa la implausibilidad de la indeterminación de la identidad personal. En la próxima sección me ocuparé de examinar esta parte de la teoría de Parfit.

7. La identidad personal no es lo que importa

Como ya hemos dicho, la más importante de las afirmaciones de Parfit es la de que la identidad personal no es lo que importa. Lo que en realidad tiene la importancia que solemos atribuir a la identidad es la relación R, conexividad y/o continuidad psicológica, con cualquier causa. Lo único que diferencia, según el criterio de Parfit, la identidad personal de la relación R es el hecho de que, en el primer caso, la relación sea uno-a-uno. El autor argumenta que si estamos relacionados, mediante la relación R, con otra persona, el hecho de que tal relación sea uno-a-uno (es decir, el hecho de que no estemos relacionados de ese modo con nadie más) no afecta a la naturaleza intrínseca de nuestra relación con dicha persona y, por lo tanto, no puede provocar que dicha relación deje de contener lo que importa. Como señala el autor, en el caso de la división ninguna de las personas resultantes sería la original, pero la relación de la persona original con cada una de ellas contendría lo que importa⁵³. El autor considera la posibilidad de

descripción es la mejor no supone afirmar que sea verdadera. Ninguna de las posibles descripciones se ajusta mejor a los hechos que el resto, simplemente hay algunas que realizan o implican suposiciones arbitrarias o inconsistentes, lo cual las convierte en una descripción peor. Volviendo a la idea de que la identidad es relativa al hablante, podemos decir que los hechos podrían causar cualquiera de las impresiones en el hablante, pero que ciertas impresiones se basarían en suposiciones arbitrarias o inconsistentes, suposiciones que sería racional revisar.

⁵³ En un trabajo anterior, Parfit (1971b) argumenta en favor de su idea de que la identidad personal no es lo que importa afirmando que puesto que las respuestas a ciertas preguntas importantes acostumbran a depender de las respuestas a las preguntas por la identidad y estas últimas preguntas no siempre tienen respuesta, tenemos que poner esas primeras cuestiones a parte de la cuestión de la identidad personal.

que se defiende que lo que importa es la continuidad física. El autor señala que el criterio físico, en su versión más plausible, sólo exige la continuidad de suficiente cerebro como para que haya continuidad psicológica. La continuidad del cerebro importa porque es la causa normal de la continuidad psicológica. El autor argumenta que si la continuidad del cerebro no llevara consigo la continuidad psicológica, no tendríamos motivo para atribuirle importancia, por lo que no puede defenderse que lo que importa es la continuidad física. Como mucho, podría defenderse que lo que importa es la continuidad psicológica con su causa normal, es decir, la continuidad psicológica causada por la continuidad de suficiente cerebro. Contra esta última posibilidad, Parfit argumenta que lo que importa no es la causa sino el efecto. Cuál sea la causa de que se dé la relación R no puede ser lo importante, puesto que ello no afecta al resultado. El resultado será, sea cual sea la causa, que habrá una persona que será psicológicamente continua con la original, y eso es todo lo que importa. Parfit reconoce que puede ser importante para nosotros tener un cuerpo igual que el que teníamos, pero, en ese caso, lo que es importante no es seguir teniendo el mismo cuerpo, sino tener un cuerpo cualitativamente idéntico al que teníamos. Por supuesto, como señala Parfit, si no importa la causa, tampoco puede importar que la causa sea fiable. Como afirma el autor, si la causa es fiable o no puede importar antes de ver el resultado, pero una vez hemos visto que hay continuidad psicológica no puede importar que la causa que la haya provocado fuera fiable o no. Por lo tanto, debemos concluir que lo que importa es la relación R con cualquier causa.

Volvamos al caso de la línea secundaria. En dicho caso, Parfit imagina que la máquina teletransportadora no destruyera a la persona original, sino que simplemente creara una réplica exacta en otro lugar. El autor añade que la máquina, inesperadamente, daña a la persona original, provocándole un problema cardíaco que la deja sólo con unas horas de vida. Entonces, la persona original habla con su réplica y esta le dice que continuará su vida donde ella la ha dejado. Según la teoría de Parfit, lo que la persona original tendría en dicho caso es casi tan bueno como la supervivencia corriente. Es cierto que la réplica no sería psicológicamente continua con la persona original en el momento de su muerte, sino que lo sería con dicha persona tal como era unas horas antes. Sin embargo, no habría ninguna diferencia importante entre tal situación y una en la que la persona fuera a perder los recuerdos de las últimas horas. A nadie nos gustaría perder los recuerdos de varias horas, pero difícilmente consideraríamos dicha situación tan mala como la muerte. Como afirma Parfit, la única diferencia entre el caso de la línea secundaria y esta última situación es que en el primero la vida de la persona original y la de la réplica se solapan brevemente en el tiempo. Esta diferencia no es lo suficientemente importante como para impedir que lo que la persona original tendría sea prácticamente tan bueno como la supervivencia corriente. El autor argumenta también que en el caso del teletransporte normal, en que la persona original es destruida, la relación que tendría la persona

original con su réplica sería tan buena como la supervivencia corriente. Puesto que la continuidad psicológica no tendría su causa normal, podríamos optar por decir que la persona original y su réplica no serían la misma persona. Podríamos optar también por decir que sí serían la misma persona. Estas no son dos posibilidades distintas, son sólo dos descripciones diferentes de una misma situación, y en dicha situación la relación entre ambas personas contiene todo lo que importa.

Según Parfit, el valor de la relación entre una persona actual y una persona futura depende tanto del grado de conexividad como del valor que para la persona actual tengan los rasgos físicos y psicológicos de la persona futura. El autor reconoce, además, que para ciertas personas sobrevivir con muchos cambios físicos puede no ser tan bueno como la supervivencia corriente. Muchas de las objeciones que otros autores dirigen contra la afirmación de Parfit de que la identidad personal no es lo que importa se centran en mostrar que existen muchas situaciones en que la relación R está presente pero en las cuales lo que la persona en cuestión tendría no sería tan bueno como la supervivencia corriente. Rupert Read (2015) argumenta que si Parfit no hubiera puesto fin al relato del caso de la línea secundaria en el momento en que lo hace, hubiera tenido que mostrar el modo en que la muerte de la persona original afectaría a su réplica, con lo cual se hubiera hecho evidente que lo que la persona tendría no sería tan bueno como la supervivencia corriente⁵⁴. Susan Wolf (1986) argumenta de manera similar, afirmando que hay casos en que el tipo de supervivencia que se tiene en el caso de la línea secundaria (es decir, sobrevivir teniendo también una réplica) sería peor que la supervivencia corriente.

Este tipo de argumentos se basan en un error que también es cometido por Parfit. Estas objeciones pasan por alto cuál es el tipo de importancia de que estamos hablando. En su respuesta a Wolf, Parfit (1986) parece darse cuenta de ello. Lo que el autor afirma en dicha ocasión es que lo que estamos preguntándonos es si sería racional preocuparnos por la persona resultante, no si la situación en cuestión tendría buenas o malas consecuencias para nosotros. No obstante, en la mayor parte de sus afirmaciones, Parfit comete el mismo error que Wolf. Como Parfit afirma al responder a la objeción de Wolf, lo que nos interesa saber no es si nuestra situación será tan buena como la supervivencia corriente, lo que nos interesa saber es si en dicha situación tendremos la relación que importa con la persona resultante. Si la máquina teletransportadora convirtiera en millonaria a la réplica de cada persona que la utiliza, nuestra situación tras utilizar la máquina sería mucho mejor que la supervivencia corriente, pero no en el sentido de que estamos hablando. Del mismo modo, si la máquina creara a las réplicas con un

⁵⁴ El autor también argumenta que cuando, tras la división, ambas personas empiezan a tener experiencias distintas, se convierten en dos personas diferentes. Sin embargo, como ya hemos visto, Parfit no niega eso. Lo que Parfit afirma es que la relación de la persona original con cada una de esas dos personas distintas contendría todo lo que importa.

brazo menos, nuestra situación tras usar la máquina sería mucho peor que la supervivencia corriente, pero, de nuevo, no en el sentido que nos interesa. Una manera de darnos cuenta de qué cambios serían relevantes para la cuestión que estamos tratando es ser neutrales con respecto a la pregunta de si la supervivencia es buena o mala. Podemos imaginar dos variantes de una determinada situación. En ambas se dan los mismos hechos, pero mientras que en una de ellas la persona en cuestión desea sobrevivir, en la otra, tal vez a causa de algún tipo de depresión, no lo desea. Aquellos hechos que importen en el sentido en que solemos suponer que la identidad importa serán positivos en la primera variante de la situación pero negativos en la segunda. Volvamos a los dos ejemplos anteriores. Supongamos que la máquina teletransportadora convierte en millonaria a toda réplica que crea. Si la persona desea sobrevivir, el hecho de que la persona se convierta en millonaria le parecerá todavía mejor. Si la persona, en cambio, no desea sobrevivir, el hecho de que se convierta en millonaria no empeorará la situación, sino que en cualquier caso la mejorará. Puesto que el hecho de convertirse en millonario no tiene valor positivo en una situación y negativo en la otra, podemos concluir que no importa en el sentido de que estamos tratando. Vayamos ahora al ejemplo en que la máquina crea a las réplicas con un brazo menos. Si la persona desea sobrevivir, el hecho de que sobreviva con un brazo menos empeorará su situación, su situación será peor que si hubiera sobrevivido con ambos brazos. Si, en cambio, la persona en cuestión no desea sobrevivir, el hecho de sobrevivir con un brazo menos no hará que su situación sea mejor, no hará que dicha persona esté menos preocupada por el hecho de sobrevivir. Si la persona sobrevive con un brazo menos, su situación será aun peor. El valor de perder un brazo es negativo en ambas variantes, de modo que debemos concluir que no importa en el sentido en que solemos creer que importa la identidad. Hemos visto que Parfit afirma que para ciertas personas puede ser importante tener un cuerpo muy similar al que tenían. Esto es cierto, pero no en el sentido del que estamos hablando. Si una persona desea sobrevivir y sobrevive con un cuerpo muy distinto al que tenía, puede que eso haga su situación peor que la supervivencia corriente. No obstante, si la persona no desea sobrevivir, el hecho de sobrevivir con un cuerpo muy diferente al suyo no hará que dicha persona se sienta aliviada, sino que hará que su situación sea todavía peor. Puesto que las diferencias físicas no invierten su valor en ambas variantes, podemos concluir que no importan en el sentido de que estamos hablando. El único hecho que invierte su valor en ambas variantes es el darse de la relación R. Si una persona desea sobrevivir y en el futuro existirá una persona psicológicamente continua con ella, la persona original podrá sentirse satisfecha. Si, en cambio, dicha persona no desea sobrevivir y entre ella y una persona futura se da la relación R, la persona en cuestión debería ver el darse de dicha relación como algo negativo. Puesto que la relación R es el único hecho cuyo valor se invierte en ambas variantes, podemos concluir que si hay alguna relación que importe en el sentido en que solemos creer que la identidad importa, dicha relación es la relación R. Hay muchas cosas

que importan, pero si hay alguna que importe en el sentido del que estamos hablando, se trata de la relación R.

Los argumentos anteriores sólo demuestran que si alguna relación importa, se trata de la relación R. Sin embargo, no demuestran que la relación R importe. Podría ser que ninguna relación importe. Podría ser que al demostrar que la identidad personal no es lo que importa hubiéramos demostrado que ninguna relación importa. Si esto fuera así, no tendríamos ningún motivo para preocuparnos por ninguna persona futura o, mejor dicho, no tendríamos ninguna razón para preocuparnos de ninguna persona futura por las mismas razones y del mismo modo que nos preocupamos por nosotros mismos. Parfit se refiere a este punto de vista como la tesis radical. Puede pensarse que si la identidad personal no consiste en otra cosa que en el darse de ciertas relaciones entre ciertos sucesos presentes y ciertos sucesos futuros, entonces no tenemos ningún motivo para preocuparnos por nuestro propio futuro. Si el que una persona futura vaya a ser yo sólo significa que sus características físicas y psicológicas tendrán cierta relación con las mías, no está claro que hay en el hecho de que esa persona futura sea yo que deba llevarme a preocuparme por lo que le suceda. Parfit señala que el hecho de que estemos o no preocupados por nuestro futuro es una cuestión distinta de la de si tenemos razones para estar preocupados por dicho futuro. La teoría de la evolución, como afirma Parfit, nos ofrece una buena explicación de por qué estaríamos preocupados por nuestro propio futuro aunque no tuviéramos razones para ello. En favor de la afirmación de que deberíamos preocuparnos por la persona futura con la que tengamos la relación R, Parfit argumenta que nuestra relación con dicha persona sería muy íntima. Según el autor, podríamos preocuparnos por dicha persona del mismo modo en que nos preocuparíamos por un hijo. También podríamos preocuparnos por esa persona debido al interés que tenemos en que nuestras intenciones se realicen y nuestros deseos se lleven a cabo. Sin embargo, Parfit reconoce que podría objetarse que dicha preocupación no es el mismo tipo de preocupación que sentimos por nosotros mismos. Podría objetarse, como señala el autor, que si tal persona fuera a sentir dolor no deberíamos anticipar dicho dolor del mismo modo en que lo anticiparíamos si lo fuéramos a sentir nosotros. El autor concluye que la tesis radical es defendible, pero afirma que también es defendible la tesis más moderada según la cual la relación R nos da razones para estar especialmente interesados por la persona futura con la cual estemos relacionados de ese modo. Parfit afirma que no conoce ningún argumento que pueda permitirnos optar por una u otra tesis.

En primer lugar, debemos dejar claro cuál es el sentido en el que estamos investigando si tenemos razones o no para preocuparnos por nuestro futuro. A la mayoría de nosotros nos bastarían las razones que Parfit da para preocuparnos por nuestro yo futuro, pero esas no son las razones que estamos investigando si se dan o no. No resulta inconcebible que pudiera existir

una persona que se preocupara únicamente por sí misma. Del mismo modo, no resulta inconcebible la existencia de una persona que sólo se preocupe por su propio presente. Lo que queremos averiguar es si dicha persona, a causa del hecho de estar preocupada por su propio presente, estaría racionalmente obligada a preocuparse por su propio futuro. Lo que estamos investigando es si la preocupación por el propio presente implica necesariamente preocupación por el propio futuro, es decir, queremos saber si la actitud de alguien que se preocupara por su propio presente pero no por su propio futuro sería incoherente. Es fácil darse cuenta de que eso sería así si fuéramos entidades que existen separadamente. Si hubiera una entidad que persistiera a lo largo del tiempo y que fuera el sujeto de nuestras experiencias y si nosotros fuéramos tal entidad, la preocupación por el propio presente implicaría necesariamente preocupación por el propio futuro. Si fuéramos entidades de este tipo y nos preocupáramos por nosotros mismos ahora, deberíamos preocuparnos también por nosotros mismos en el futuro, puesto que, ahora y en el futuro, la entidad en cuestión sería la misma. Si nos preocupara dicha entidad ahora, debería preocuparnos dicha entidad en el futuro, puesto que la entidad presente y la entidad futura serían la misma. Si, por el contrario, no somos entidades que existan separadamente, las cosas no están tan claras. ¿De qué modo nuestra preocupación por nuestro yo actual podría implicar preocupación por nuestro yo futuro? Podríamos pensar que nuestra preocupación por nuestro yo actual nos da motivos para preocuparnos por nosotros mismos durante todo el intervalo temporal que dura el presente especioso. Entonces, puesto que mientras permanecemos conscientes habría continuidad del presente especioso, podríamos pensar que tenemos razones para preocuparnos por nuestro yo futuro hasta donde alcance dicha continuidad, es decir, hasta donde haya conciencia ininterrumpida. Sin embargo, esto es una ilusión. Que haya tal continuidad no implica de ningún modo que yo ahora y yo en el instante siguiente seamos una misma entidad en ningún sentido que lleve a que la preocupación por la primera entidad implique preocupación por la segunda. Puede que al ser conscientes del presente especioso seamos conscientes de cierto intervalo de tiempo, pero, puesto que no podemos ser conscientes de lo que aún no ha sucedido, dicho intervalo de tiempo no se extiende hacia el futuro. Puede que tengamos la sensación de que el presente especioso se extiende en cierto modo hacia el futuro, pero ello es sólo una ilusión. Únicamente podemos estar percibiendo simultáneamente sucesos que ya han tenido lugar, de lo cual no podemos extraer ninguna razón para preocuparnos por nuestro propio futuro. Lo único que el presente especioso nos muestra es que el pasado inmediato nos afecta de una manera especialmente relevante, pero, puesto que ya sabíamos que nuestro pasado nos afecta, esto no nos permite demostrar nada que no pudiéramos demostrar ya.

Sin embargo, sí que tenemos una razón, aunque indirecta, para preocuparnos por nuestro propio futuro si nos preocupamos por nuestro propio presente. Como señala Parfit (1982), si no

tuviéramos continuidad psicológica, especialmente en lo referente a la memoria y la intención, no podríamos ni pensar, ni planificar, ni actuar, ni, en general, hacer prácticamente nada. Podemos llevar esta afirmación de Parfit un poco más lejos. Podemos decir que si no nos preocupáramos por nuestros yoes futuros, tampoco podríamos hacer ninguna de esas cosas. Si pensáramos en nosotros mismos como siendo sólo nuestro yo actual, nuestra vida duraría, para nosotros, un único instante, un instante durante el cual no podríamos hacer nada. Para poder vivir satisfactoriamente el momento presente, necesitamos preocuparnos por nuestro propio futuro. Puede que nuestra preocupación por nuestro futuro se deba a una ilusión, pero necesitamos seguir sufriendo dicha ilusión para que nuestro presente tenga valor. Por lo tanto, si nos preocupa nuestro yo actual, queremos mantener tal ilusión y, si mantenemos tal ilusión, nos preocuparemos por nuestros yoes futuros. De este modo indirecto, la preocupación por nuestro yo actual justifica la preocupación por nuestro futuro. Puede que la preocupación por nuestro futuro sea una actitud irracional, pero, empleando la expresión que Parfit utiliza al hablar de otros temas, podemos decir que en este caso tenemos razones para adoptar una actitud irracional. Sin embargo, la adopción de esta actitud irracional no supone ningún tipo de autoengaño. Afortunadamente, aunque sepamos, a nivel teórico, que no es racional preocuparnos por nuestro futuro, ello no provoca fácilmente un cambio en nuestra actitud hacia dicho futuro. Dicho cambio de actitud requeriría un gran esfuerzo, un esfuerzo que tenemos buenas razones para no llevar a cabo. Por supuesto, todo esto no significa que aceptar el reduccionismo no pueda modificar nuestra actitud hasta cierto punto. Parfit opina que, al ser la relación R una cuestión de grado, tenemos más motivos para preocuparnos por nuestro futuro cercano que las que tenemos para preocuparnos por nuestro futuro lejano. Parfit también afirma que, puesto que para el reduccionismo el que alguien muera sólo significa que en el futuro no habrá ningún suceso que esté relacionado de ciertas formas con ciertos sucesos actuales, aceptar el reduccionismo debería provocar que viéramos la muerte como algo menos terrible. Mi anterior argumento no implica que tengamos razones para evitar estos cambios en nuestra actitud. Por último, hay que señalar que el anterior argumento no nos permite descartar la posibilidad de que tengamos razones directas para preocuparnos por nuestro propio futuro. Tal vez haya algún argumento que nos permita demostrar la existencia de tales razones. Parfit declara no haber encontrado ese argumento. Yo tampoco he podido encontrar dicho argumento, pero por lo menos tenemos razones indirectas para preocuparnos por nuestro propio futuro.

Antes de pasar a la siguiente sección, puede ser conveniente examinar una última objeción a la idea de que la identidad personal no es lo que importa. David Lewis (1983) defiende que la identidad personal consiste precisamente en la relación R, coincidiendo, por lo tanto, con la relación que importa. En opinión de Lewis, la identidad personal es una relación entre personas continuas que induce otra relación, a la que el autor se refiere como relación I, entre las etapas

de la persona continua en cuestión. Lewis está de acuerdo con Parfit en que la relación de identidad y la relación R son relaciones distintas, pero opina que esta última es la misma que la relación I. El autor afirma que hay identidad entre dos personas continuas si se da la relación I entre sus etapas, y defiende que, en cualquier caso posible, entre todo par de etapas I-relacionadas se da también la relación R y a la inversa. Lewis entiende una persona continua como una sucesión de etapas unidas por la relación I, concretamente como una sucesión máxima de etapas relacionadas de ese modo. Si dos etapas se encuentran I-relacionadas, las personas continuas a las que pertenece cada etapa serán una y la misma persona y si dos personas continuas son la misma, sus etapas estarán relacionadas de ese modo. El autor afirma que en los casos normales la relación I hereda las características formales de la relación de identidad, pero argumenta que puede haber casos en que una misma etapa pertenezca a más de una persona continua y señala que en tales casos la relación I diferirá, en sus características, de la relación de identidad. Puesto que con la relación R sucede lo mismo, Lewis no ve problema en afirmar que esta y la relación I son una y la misma. De este modo, para Lewis, dos etapas se encuentran I-relacionadas, o R-relacionadas, en caso de que exista alguna persona continua a que ambas pertenezcan, cosa que no impide que ambas pertenezcan también a otras personas continuas. La relación R difiere de la de identidad en que no es transitiva. Los casos en que la relación R es intransitiva son los casos de fisión y fusión, es decir, los casos en que una persona se divide en dos personas distintas y los casos en que dos personas distintas pasan a formar una sola persona. Según Lewis, lo que sucede en dichos casos es que hay un solapamiento de personas continuas. En tales casos nos encontramos ante ciertas etapas que pertenecen a más de una persona continua. Cuando tiene lugar un caso de fisión, lo que sucede es que dos personas continuas que se encontraban solapadas toman caminos distintos y dejan de compartir etapas. Cuando tiene lugar un caso de fusión, dos personas continuas se solapan y pasan, por lo tanto, a compartir sus etapas. Eso implica que en dichos casos hay más de una persona donde parece haber sólo una. Para evitar la aparente implausibilidad de esta conclusión, Lewis argumenta que no contamos a las personas por la identidad, sino por la indiscernibilidad de sus etapas, de modo que, cuando nos encontramos con dos etapas idénticas las contamos como si se tratara de una sola persona. Del mismo modo que Parfit, Lewis opina que la identidad personal puede ser indeterminada. El autor afirma que si la relación R se da en un grado reducido, puede ser indeterminado si un cierto conjunto de etapas forman una única persona continua o no.

La teoría de Lewis tiene la virtud de preservar la intuición de que la identidad personal es lo que importa. A pesar de ello, tiene algunos defectos que la hacen inferior a la teoría de Parfit. En primer lugar, es innecesariamente más compleja en el sentido de que postula, en cada instante particular, la existencia de más personas de las que implica la teoría de Parfit. Además, en caso de que algunas de nuestras etapas pertenecieran a más de una persona concreta, no podríamos

saber cuál de esas personas somos ni por el futuro de cuál de ellas hemos de preocuparnos. De hecho, si nos preocupáramos por el futuro de la persona correcta, la otra persona continua se estaría preocupando por el futuro de la persona equivocada. Tal vez Lewis respondería que aunque lo correcto sea decir que hay dos personas continuas, lo único que hay en el momento en que estas se solapan es la etapa que comparten y que debemos preocuparnos por todas las personas continuas que comparten dicha etapa porque lo que somos en ese instante es la etapa en cuestión. Pero si Lewis opinara esto, su teoría sería la misma que la de Parfit expresada de otro modo, y de un modo mucho menos natural. Otro problema de la teoría de Lewis se hace evidente cuando consideramos el caso de la fisión. Como hemos dicho, según el autor, cuando una persona se divide en varias lo que sucede es que había, desde el principio, varias personas. Supongamos que existiera una máquina teletransportadora que, en lugar de crear una sola réplica, creara dos de ellas en dos lugares distintos. Según la teoría de Lewis, si yo usara la máquina, ello significaría que hasta entonces existían varias personas continuas solapadas. El hecho de que donde yo estoy haya una o varias personas depende de si en el futuro utilizaré o no dicha máquina. Parece extraño que la cuestión de cuántas personas hay ahora en cierto lugar dependa de sucesos futuros y no pueda ser respondida examinando únicamente los hechos presentes⁵⁵. En mi opinión, estos problemas de la teoría de Lewis bastan para preferir la idea de Parfit de que la identidad personal no es lo que importa.

La relación que importa, por lo tanto, será aquella que cree la ilusión que nos lleva a preocuparnos por nuestro propio futuro. En esta sección y en la sección quinta he defendido que dicha relación es la relación R. Sin embargo, creo que la relación R tiene varios defectos y necesita ser revisada. Esto es lo que defenderé en la sección siguiente.

8. Similitud psicológica: eliminando la causalidad de la relación R

Imaginemos que en algún lugar del universo, por pura casualidad, ciertas partículas se reunieran formando una réplica exacta de Descartes, que fuera psicológicamente continua con Descartes tal como era cuando murió. Supongamos que la reunión dura lo suficiente como para que esa réplica sea consciente de su propia existencia durante unos instantes, tras los cuales las partículas se dispersan de nuevo y la réplica vuelve a la inexistencia. No es necesario que nos preguntemos si algo así podría o no suceder. Preguntémonos, en cambio, si, en caso de que Descartes hubiera creído que algo así iba a tener lugar, hubiera tenido que pensar en dicho suceso como en su resurrección. Por supuesto, Descartes no hubiera visto tal suceso de ese

⁵⁵ Sin embargo, Lewis no consideraría adecuada esta última objeción, puesto que era partidario de la teoría de las partes temporales o cuadridimensionalismo, según la cual un objeto está compuesto por sus distintas etapas temporales del mismo modo que lo está por distintas partes situadas en diferentes lugares del espacio.

modo, puesto que creía que somos entidades que existen separadamente. Lo que nos estamos preguntando es si, de haber sabido la verdad, sea la que sea, sobre la identidad personal, hubiera estado racionalmente obligado o no a preocuparse de dicha réplica como de sí mismo. Si la teoría de Parfit es cierta, la respuesta es no, puesto que, según él, lo que importa es que se dé la relación R con cualquier causa. Puesto que lo que se da en este caso es la relación R sin ninguna causa, las condiciones de Parfit no se cumplen⁵⁶. Sin embargo, para defender que lo que importa es la relación R con cualquier causa, Parfit argumentó que no podía importar cuál era la causa, sino que lo importante tenía que ser el resultado. Cuál sea la causa no parece poder afectar a la importancia intrínseca que tenga la relación existente entre las dos personas en cuestión. Si lo que importa no es cuál sea la causa, sino cuál sea el resultado, ¿por qué iba a importar que haya o no una causa?

Ingmar Persson (2016) argumenta de un modo similar en favor de la necesidad de eliminar la causalidad de la relación R. El autor imagina que una persona se desintegrara y que, casualmente, otras partículas se reunieran formando una réplica exacta de la persona tal como era en el momento de la desintegración y situada exactamente en el mismo lugar dónde la persona original desapareció. La reunión de las partículas podría tener lugar un instante después de la desintegración, de modo que nadie se percatara de ninguno de los dos sucesos. La réplica sería psicológicamente continua con la persona original. De hecho, la réplica creería ser el original. La única diferencia entre la relación que se daría entre la persona original y su réplica en este caso y la que se daba en los casos que hemos considerado en las secciones anteriores es que aquí el darse de la relación R no tendría ninguna causa, es decir, que no habría ningún nexo causal entre las características psicológicas de la persona original y las de la réplica. Como señala el autor, no parece haber ningún motivo por el que fuera a ser peor para la persona original ser reemplazado de este modo que serlo de una manera que involucrara un nexo causal entre ella y la réplica. Por lo tanto, “lo que importa [...] es simplemente que haya la similitud psicológica que normalmente involucra la conexividad y/o continuidad psicológica”⁵⁷ (Persson, 2016, p. 164). Noonan (1989) también argumenta que Parfit, para ser coherente, debería eliminar la exigencia de que haya una causa. El autor argumenta que lo que distingue la relación R sin ninguna causa de la relación R con cualquier causa es una diferencia trivial y señala que Parfit, al argumentar en favor de su idea de que la identidad personal no es lo que importa, afirma que lo que importa no puede depender de ningún hecho trivial. Noonan también señala que los argumentos que usa Parfit para defender que el tipo adecuado de causa puede ser cualquier causa pueden también utilizarse para defender que no es necesario que haya ninguna

⁵⁶ Podría pensarse que el azar debería contar como una causa, pero lo que Parfit exige es que exista una relación causal entre la persona pasada y aquella persona presente con la que está R-relacionada, exigencia que, por supuesto, el azar no cumple.

⁵⁷ La traducción es mía.

causa. Noonan considera que la conclusión de que la relación R puede no tener causa y seguir siendo lo que importa es absurda. No obstante, si el reduccionismo es cierto y la identidad personal no consiste en otra cosa que en el darse de ciertas relaciones entre determinados sucesos pasados y presentes, no parece tan implausible que lo importante sean dichas relaciones independientemente de cuales sean sus causas e independientemente incluso de que tengan alguna causa. Si la identidad personal no consiste en ningún hecho adicional profundo y lo único que hay que saber sobre el tema es que se dan ciertos sucesos interrelacionados de ciertas formas, entonces si hay algo que importe, parece plausible suponer que las relaciones causales que provocan que se den dichas relaciones no forman parte de ello. Lo que importa son los resultados. Si se da el resultado adecuado, lo más habitual es que haya habido una causa, pero eso no debe hacernos pensar que lo que importa es esa causa. Lo que importa es, sin duda, el resultado, de modo que Persson tiene buenos motivos para afirmar que lo que importa es la relación R independientemente de que haya causa o no, es decir, que tiene buenas razones para afirmar que lo que importa es lo que él llama similitud psicológica.

Centrémonos por un momento en el tipo de conexión psicológica más importante, la cuasi-memoria. Habíamos dicho que para tener un cuasi-recuerdo de una experiencia pasada es necesario tener un recuerdo aparente de dicha experiencia, que la experiencia en cuestión haya tenido realmente lugar y que entre la experiencia y el recuerdo de la misma exista una relación causal del tipo apropiado. Es evidente que las consideraciones anteriores eliminan el tercer requisito. Para que podamos hablar de continuidad y/o conexividad psicológica sin ninguna causa es necesario eliminar la causalidad de la definición de las conexiones psicológicas. En el caso de la cuasi-memoria tendremos, por lo tanto, que decir que para que haya un cuasi-recuerdo de una experiencia pasada basta con que exista un recuerdo aparente de dicha experiencia y con que dicha experiencia haya tenido realmente lugar⁵⁸. Puede que esta ya parezca una definición adecuada, pero en realidad sigue exigiendo demasiado. Es necesario revisar todavía el segundo de sus requisitos. No podemos exigir que la experiencia haya tenido realmente lugar.

Estamos acostumbrados a pensar en la memoria como algo completamente fiable que nos aporta conocimiento seguro del pasado. Pensamos en nuestros recuerdos como si fueran reproducciones de experiencias pasadas que hubieran sido almacenadas en nuestro cerebro. Sin embargo, la realidad es muy distinta. Los recuerdos parecen ser, más bien, una reconstrucción de experiencias pasadas, reconstrucción a la que se incorpora información obtenida tiempo después de vivir la experiencia en cuestión. Kathryn Braun-LaTour, Michael LaTour, Jacqueline

⁵⁸ Podría objetarse que si se elimina la causalidad de la cuasi-memoria, esta deja de ser conocimiento, pero no hay ningún motivo para pensar que el conocimiento tiene que estar involucrado en la relación que importa.

Pickrell y Elizabeth Loftus (2004) llevaron a cabo un interesante experimento⁵⁹ que muestra lo maleable que es la memoria. A los 66 participantes se les asignó al azar una de entre dos condiciones. A cada uno de los grupos se le dio un anuncio diferente de Disneyland. Cada uno de los anuncios realizaba una descripción de una experiencia típica en el parque y llamaba al lector a recordar la suya. Sin embargo, mientras que uno de los dos anuncios describía el evento de darle la mano a Mickey Mouse, el otro describía el de darle la mano a Bugs Bunny, que es un personaje de la Warner Bros y que, por lo tanto, es poco probable que uno se lo encuentre en Disneyland. Cada uno de los dos anuncios contenía además una imagen del personaje del que hablaba. El objetivo era comprobar si el anuncio era capaz de modificar los recuerdos que los participantes tenían de su experiencia en el parque. En los cuestionarios que se les realizó se observó que entre quienes habían leído el anuncio con información falsa hubo más personas que afirmaron que Bugs Bunny había formado parte de su experiencia pasada en Disneyland, concretamente un 22% frente al 7% de quienes recibieron el anuncio que no contenía información falsa⁶⁰. Este experimento y muchos otros con los que se han obtenido resultados que van en la misma dirección llevan a la conclusión de que la forma correcta de ver nuestros recuerdos es como “reconstrucciones de fragmentos y piezas de información que hemos obtenido a lo largo del tiempo”⁶¹ (Braun-Latour *et al.*, 2004, p. 7). Los autores afirman que aunque muchas veces estas reproducciones son muy similares a las experiencias originales, en otras ocasiones el recuerdo es muy distinto a lo que realmente sucedió. En el experimento que acabamos de ver está involucrada únicamente la memoria a largo plazo, pero existen numerosos experimentos que han mostrado lo fácil que resulta que se produzcan recuerdos falsos de eventos cercanos en el tiempo (*vid.* Drivdahl, Zaragoza y Learned, 2009). De hecho, hay incluso evidencia de que los procesos no conscientes bastan para la creación de recuerdos falsos (*vid.* Cotel, Gallo y Seamon, 2008). Algunos autores explican la creación de falsos recuerdos afirmando que no recordamos cuál es la fuente de nuestros recuerdos sino que lo inferimos de sus características, de modo que en ciertas ocasiones resulta posible confundir la fuente y creer que la de eventos posteriores, como sugerencias, es la misma que la del recuerdo original (Drivdahl *et al.*, 2009). Si la memoria no es la reproducción exacta de la experiencia original sino una reconstrucción de la misma realizada a partir de la información almacenada, carece de sentido incluir en la definición de recuerdo el requisito de que la experiencia haya tenido lugar realmente. ¿Qué significa que la experiencia haya tenido lugar? Parece razonable pensar que en los casos en que el recuerdo no sea falso lo único que podremos afirmar es que ha tenido lugar una experiencia muy parecida. Si lo que sucede cuando recordamos algo es que nuestra mente

⁵⁹ Experimentos prácticamente idénticos ya habían sido llevados a cabo anteriormente por otros autores.

⁶⁰ Hay que tener en cuenta, además, que algunos de quienes recibieron el anuncio con información falsa se dieron cuenta de que Bugs Bunny no era un personaje de la Disney. Evidentemente estas personas no se refirieron luego a Bugs Bunny como formando parte de su experiencia pasada.

⁶¹ La traducción es mía.

crea una experiencia a partir de la información que tiene disponible⁶², no tiene sentido preguntar si esa misma experiencia ha tenido lugar o no. Siendo esto así, no podemos exigir, para que haya un cuasi-recuerdo, que la experiencia haya tenido realmente lugar, sólo podemos exigir que haya tenido lugar una experiencia lo suficientemente parecida.

Alguien podría tal vez pensar que en los casos en que no hay falsos recuerdos la experiencia que se recuerda sí que es exactamente como la que tuvo lugar. No obstante, eso no tiene sentido sea cual sea el mecanismo por el que recordamos nuestras experiencias. De lo que no hay duda es de que la experiencia de recordar algo es una experiencia distinta a aquella que se recuerda. Para que se pudiera exigir que la experiencia haya tenido lugar realmente tendría que ser posible separar de la experiencia de recordar algo que fuera exactamente idéntico a la experiencia recordada. No veo de qué modo podría ser posible algo así. Lo mismo sucede con el resto de conexiones psicológicas. La experiencia de realizar una intención no es la misma que la experiencia de formarse dicha intención y no parece que haya nada en la primera de lo que se pueda decir que es exactamente idéntico a algo de la segunda. Podría pensarse que las cosas son distintas con los rasgos de carácter, pero dichos rasgos no son otra cosa que disposiciones a actuar de ciertas maneras, y nuestras disposiciones están siempre en constante cambio, aunque se trate de cambios de poca importancia. Cuando hablamos de conexiones psicológicas, por lo tanto, no sólo no podemos exigir que haya un nexo causal, tampoco podemos exigir que la característica psicológica en cuestión sea la misma en ambos casos. Únicamente podemos exigir similitud. Volviendo al ejemplo de la cuasi-memoria, la manera más adecuada de definirla sería afirmando que hay un cuasi-recuerdo de una experiencia si hay un recuerdo aparente de dicha experiencia y si tuvo efectivamente lugar una experiencia lo suficientemente similar al recuerdo que se tiene de ella. De esta manera, podemos emplear la expresión “similitud psicológica” en un sentido mucho más literal que aquél en el que lo hace Persson. Podemos hablar de conexiones psicológicas en este nuevo sentido y, puesto que Parfit define la continuidad psicológica en términos de la conexividad, podemos mantener la definición que Parfit da de dicho concepto. Puesto que nuestra conexividad y/o continuidad psicológica es muy distinta a aquella en la que pensaba Parfit, parece conveniente dejar de hablar de la relación R y utilizar la expresión de Persson. Diremos que, por lo tanto, lo que importa es que haya similitud psicológica, es decir, conexividad (en el nuevo sentido que le hemos dado a esta palabra) y/o continuidad psicológica con cualquier causa o sin ninguna causa. Es evidente que, puesto que cuando creíamos ver la relación R en realidad sólo estábamos viendo similitud psicológica, los mismos argumentos que nos llevaron a la conclusión de que la relación R era lo que importa

⁶² Incluso en individuos con una memoria muy superior a lo normal se ha observado que se pueden inducir falsos recuerdos, lo cual debe hacernos pensar que los mecanismos mentales que llevan a la obtención de los recuerdos en tales personas son los mismos que en las personas con una memoria normal (*vid. Patihis et al., 2013*).

deben llevarnos ahora a pensar que lo que importa es que haya similitud psicológica. Es la similitud psicológica lo que provoca esa ilusión de continuidad que nos lleva a preocuparnos por nuestro propio futuro, es la similitud psicológica aquello que nos lleva en la mayoría de casos a juzgar que hay identidad, de modo que es esta similitud la relación que hace que sea racional preocuparse por una persona futura.

En *Razones y personas*, Parfit ve la tesis de que somos entidades que existen separadamente como la única alternativa importante al reduccionismo. Sin embargo, posteriormente surgió una teoría de la identidad personal que ha cobrado gran importancia y no puede ignorarse. Se trata de la teoría animalista de la identidad personal. En la próxima sección examinaré esta teoría y trataré de defender que el reduccionismo sigue siendo la mejor concepción de la identidad personal de que disponemos.

9. La concepción animalista de la identidad personal

En el mismo lugar en que se encuentra cada uno de nosotros hay un animal de la especie *Homo Sapiens*. Donde yo estoy ahora hay también un animal humano que está compuesto exactamente de la misma materia que yo y que realiza exactamente las mismas acciones que yo realizo. Parece bastante razonable suponer que si existe un animal que es indistinguible de mí y que hace exactamente lo mismo que yo hago, yo tengo que ser ese animal. Esta es la afirmación en que consiste la teoría animalista de la identidad personal. Según este punto de vista, cada uno de nosotros es numéricamente idéntico a un animal humano y nuestras condiciones de persistencia son, por lo tanto, las de dicho animal. Cuáles sean dichas condiciones de persistencia es una cuestión que el animalismo deja sin resolver y a la que los distintos autores partidarios de dicha concepción ofrecen respuestas distintas. La concepción animalista parece, en principio, la más cercana al sentido común. Al fin y al cabo, todos hemos oído decir alguna vez que somos animales racionales, más concretamente animales de la especie *Homo Sapiens*. Además, si no somos ningún tipo de alma inmaterial, ¿qué podemos ser sino cierto tipo de organismos? Y si pretendemos que además del animal existe una entidad distinta que es la persona y que tiene condiciones de persistencia distintas, entonces, como señala Eric Olson (1994), estamos afirmando que pueden existir dos objetos distintos que estén constituidos al mismo tiempo por las mismas partes. Sin embargo, esta concepción, al menos en la versión de ella que mantienen la mayoría de sus defensores, resulta incompatible con el criterio psicológico de identidad personal⁶³, que es el que hemos visto que defendía Parfit y que antes del surgimiento del animalismo era el criterio mayoritariamente aceptado.

⁶³ Incluso una versión del animalismo que sostenga que el cerebro es el organismo o que en un caso de trasplante cerebral el organismo iría junto con el cerebro sería incompatible con el criterio psicológico. Al

El principal argumento que los animalistas dirigen contra los partidarios del criterio psicológico es el que se conoce como argumento o problema del animal pensante. El animal que ocupa en todo momento el mismo espacio que yo ocupó y que realiza siempre las mismas acciones que yo llevo a cabo tiene, del mismo modo que yo, un cerebro. Puesto que no parece haber nada que impida que el animal utilice dicho cerebro para pensar, tendremos que decir también que el animal piensa en todo momento los mismos pensamientos que yo. Por lo tanto, si yo no soy ese animal, hay en todo momento dos pensadores. Si eso es así, nos encontramos con un problema epistémico. No tengo manera de saber si soy la persona o el animal. Puesto que hay dos pensadores pensando cada pensamiento, cuando pienso que soy la persona el animal está pensando exactamente lo mismo y, puesto que ambos coincidimos tanto en la materia que nos compone como en las acciones que realizamos, no hay ninguna diferencia entre ambos y, a causa de ello, cualquier razón que pueda yo tener para pensar que soy la persona en lugar del animal es también una razón que tiene este último para pensar lo mismo. Por lo tanto, incluso en el caso de que no seamos animales, no podemos tener ningún tipo de razón que justifique la creencia de que no lo somos. Puesto que no podemos tener ninguna razón para pensar que el animalismo es equivocado, debemos concluir que es la concepción correcta (*cfr.* Olson, 2004).

Uno de los principales defensores del animalismo es Eric Olson. La concepción animalista de Olson se limita a la afirmación de que somos numéricamente idénticos a un determinado animal, nuestro organismo. El autor se mantiene neutral sobre si nuestra identidad con dicho animal es esencial o contingente y mantiene abierta la posibilidad de que existan personas que no sean animales⁶⁴. Además, Olson no tiene problema en aceptar que la posesión de ciertas características mentales mínimas sea un requisito para que una determinada entidad pueda ser considerada una persona, por lo que admite la posibilidad de que existan animales humanos que no sean personas, como sería el caso de los fetos. Por lo tanto, podemos dejar de ser personas sin dejar por ello de existir⁶⁵ pero lo que, de acuerdo con Olson, no puede suceder es que dejemos de ser animales y sigamos existiendo. La concepción de Olson descarta también la posibilidad de que un animal deje de ser orgánico. Si cierta parte de un animal es substituida por un substituto inorgánico, dicha parte deja de pertenecer al organismo (*vid.* Olson, 2003). Este

fin y al cabo, el cerebro podría perder la continuidad psicológica sin dejar de existir. Para mantener el criterio psicológico tendría que exigirse que, para seguir siendo el mismo organismo, el cerebro tuviera que mantener la continuidad psicológica, pero seguramente ningún animalista estaría dispuesto a aceptar dicha condición.

⁶⁴ Parece que Olson aceptaría como persona a cualquier entidad pensante, como sería el caso, si tales cosas existieran o llegaran a existir, de robots inteligentes o seres inmateriales conscientes. Aunque esos seres serían personas, no serían animales, y sus condiciones de persistencia no tendrían por qué ser las mismas que las nuestras.

⁶⁵ Como señala Tamar Szabó Gendler (2009), la concepción animalista de Olson es compatible con la existencia de una relación práctica en la que consista ser la misma persona, relación que podría estar regida por el criterio psicológico pero que, al contrario de lo que sucede con la identidad personal tal como se la entiende habitualmente, no sería una relación metafísica.

punto de vista, a pesar de su concordancia con el sentido común, tiene que hacer frente a múltiples objeciones⁶⁶.

Supongamos que el cerebro de una persona es trasplantado a otro cuerpo. El receptor del trasplante será psicológicamente continuo con el donante y, por lo tanto, creará ser él y tendrá todos sus recuerdos. Según el animalismo de Olson, sin embargo, lo que está sucediendo en este caso no es que estemos dándole al donante un nuevo cuerpo, lo que está sucediendo es que el receptor recibe un nuevo cerebro, del mismo modo que en otro tipo de trasplante recibiría un nuevo corazón o un nuevo riñón. La persona a la que se le extrae el cerebro no viajaría con este, sino que lo perdería. La persona que lo recibe, en cambio, estaría obteniendo un nuevo cerebro, lo cual tendría como consecuencia la pérdida de sus recuerdos y la obtención de los del donante. En este tipo de casos, la intuición nos lleva a conclusiones opuestas a las que nos ofrece el animalismo. La mayoría creeríamos que la persona cambiaría de cuerpo junto con el cerebro. Cuando Parfit (2004) se ocupa del criterio físico de identidad personal, considera como la versión más plausible de dicha teoría la que afirma que basta con la persistencia de la suficiente cantidad de cerebro. La teoría de Olson es cercana al criterio físico⁶⁷, pero en la versión que exige continuidad del cuerpo en su conjunto, versión que Parfit consideraba muy implausible. Uno de los principales motivos por los que esa versión del criterio físico parecía tan implausible es el tipo de intuiciones que tenemos en casos como el del trasplante cerebral. En opinión de Noonan (1989), la intuición del trasplante es una de las objeciones más importantes a que tiene que enfrentarse el animalismo. Sin embargo, Olson (2016) argumenta que la intuición nos dice que la persona va con el cerebro porque creemos que al trasladarlo estamos moviendo un ser inteligente de un cuerpo a otro, cosa que, según él, no sucede. El autor señala también que el caso en que se trasplanta cada mitad del cerebro a un cuerpo diferente supone para los defensores del criterio psicológico un problema similar al que el caso del trasplante supone para los animalistas. Además, en opinión de Olson, el argumento del animal pensante neutraliza el problema que parece suponer el caso del trasplante. Según los animalistas, como hemos visto, este argumento lleva a la conclusión de que el único ser pensante involucrado es el organismo, de modo que al trasplantar el cerebro no estamos trasplantando a ningún ser pensante, sólo estamos trasplantando el órgano que permite que el organismo piense.

⁶⁶ Al analizar las distintas objeciones al animalismo, me centraré especialmente en la versión de esta concepción que defiende Olson.

⁶⁷ Olson exige, para que haya identidad personal, que el organismo continúe desempeñando sus actividades vitales, es decir, que siga vivo. Otros defensores del animalismo, como David Mackie (1999), consideran que el organismo continua existiendo después de morir y que su existencia no termina hasta que la materia que lo forma deja de estar organizada de la manera adecuada. Este último punto de vista coincidiría completamente con la versión del criterio físico que exige continuidad del cuerpo y no del cerebro.

Hemos visto que la mayoría de versiones del animalismo consideran que para que haya identidad personal nuestro cuerpo tiene que seguir funcionando como el organismo de un animal. Otra de las objeciones que suelen dirigirse contra el animalismo explota la diferencia entre estas condiciones de persistencia y las del cuerpo. Cuando un animal muere, su cuerpo continúa existiendo. Lo que queda tras la muerte del animal parece ser algo que ya existía cuando el animal estaba vivo. No parece plausible que la muerte de un organismo suponga la creación de un nuevo tipo de objeto. Entonces nos encontramos con un problema similar al que el argumento del animal pensante supone para los defensores del criterio psicológico. Si el cuerpo del animal ya existía mientras este estaba vivo, entonces parece razonable pensar que estaba compuesto de la misma materia que este. El cuerpo del animal estuvo en todo momento en el mismo lugar que él, llevando a cabo las mismas acciones que este. Dicho cuerpo tenía un cerebro, y no parece haber nada que pudiera impedirle utilizarlo para pensar. Por lo tanto, el cuerpo pensaba en todo momento los mismos pensamientos que el animal. Si realmente existen a la vez tanto el cuerpo como el animal, no tenemos ninguna manera de saber cuál de los dos somos. Cualquier motivo que el animal pueda tener para pensar que él es el animal es también un motivo que el cuerpo tiene para pensar lo mismo. Si el animalismo fuera cierto, no podríamos saberlo. Del mismo modo que el argumento del animal pensante nos llevaba a pensar que éramos el animal, este argumento debería conducirnos a la conclusión de que somos nuestros cuerpos. Este argumento no supone ningún problema para la teoría animalista de David Mackie (1999). Según su punto de vista, el animal es numéricamente idéntico a su cuerpo. En opinión del autor, del mismo modo que cuando un gato o una mariposa muere consideramos que lo que queda es un gato muerto o una mariposa muerta, cuando muere una persona esta no deja de existir, simplemente deja de estar viva. En opinión de Mackie, ningún organismo deja de existir mientras sus partes sigan estando adecuadamente organizadas. Para la teoría de Olson, en cambio, este argumento sí que supone un problema. Puesto que parece ser paralelo al argumento del animal pensante, en principio cualquier manera de esquivarlo acabaría también con la mejor razón que tenemos para creer en la verdad del animalismo. Olson (2004) rechaza el punto de vista de Mackie. Según el autor, las condiciones de persistencia de un organismo no pueden ser las de un cuerpo porque lo que hace que un organismo sea un organismo es precisamente el hecho de que esté vivo. Olson considera también la posibilidad de responder que, estrictamente hablando, no existen cuerpos, sino sólo partículas agrupadas de una determinada manera. Sin embargo, el autor señala que si se acepta esta respuesta, los partidarios del criterio psicológico podrían argumentar del mismo modo y resolver el problema del animal pensante afirmando que no existen animales, sino sólo ciertas partículas agrupadas de una determinada manera cuya organización se mantiene más o menos estable. Olson opta por responder al argumento limitándose a aceptar que cuando el organismo muere un nuevo objeto empieza a existir. En opinión del autor, dado que las condiciones de persistencia de cuerpos y

organismos son muy distintas, no hay nada misterioso en el hecho de que un cadáver empiece a existir en el momento en que un organismo muere. Sin embargo, no deja de haber algo implausible en la respuesta de Olson. Aquello que existe cuando el organismo deja de funcionar parece ser algo que ya existía mientras aún lo hacía. Entre el organismo y el cadáver que queda tras su muerte existe el tipo de continuidad física que suele llevarnos a considerar que un objeto continúa existiendo. Como señala Mackie (1999), cuando una máquina se estropea y deja de funcionar no consideramos que deje de ser la misma máquina ni que un nuevo objeto la substituya. No parece haber ningún motivo para suponer que un organismo que no está en funcionamiento sea un nuevo tipo de objeto que antes no existía. Si el animalismo quiere evitar este problema, su única opción es aceptar la inverosímil afirmación de que, cuando morimos, en lugar de dejar de existir pasamos a ser un cadáver.

Uno de los problemas más graves a que tiene que enfrentarse el animalismo es el que se conoce como problema de la persona remanente, problema que fue inicialmente planteado por Mark Johnston. Supongamos que el cerebro de una persona es separado de su cuerpo y mantenido con vida artificialmente. Supongamos también que se dan los medios a dicho cerebro para que pueda comunicarse con el exterior recibiendo información a través de sensores que simulen sentidos como la vista y el oído y enviando respuestas a través de algún tipo de dispositivo que le permitiera hablar. Es de suponer que en tal estado el cerebro sería capaz de pensar, del mismo modo que era capaz de hacerlo cuando estaba conectado a su cuerpo. Entonces, la pregunta obvia es quién sería el sujeto de dichos pensamientos. Tanto si suponemos que el sujeto de esos pensamientos es el cerebro como si suponemos que es algún otro tipo de entidad, el caso es que tal sujeto encaja con la definición de Locke según la cual una persona es un ser consciente. Entonces, el animalismo se encuentra con el problema de tener que explicar de dónde ha salido dicha persona. Según la concepción animalista, antes de la extracción del cerebro existía una persona que era numéricamente idéntica al animal y que, por lo tanto, no podía seguir existiendo si dejaba de hacerlo el organismo ni podía tampoco pasar a existir separadamente de este. Tras la extracción del cerebro, el cuerpo podría ser mantenido con vida artificialmente, de modo que el organismo, aunque sin cerebro, continuaría existiendo. Por otro lado, nos encontraríamos con que existe una persona distinta de ese organismo que piensa determinados pensamientos y que, aparentemente, ha empezado a existir en el momento en que el cerebro ha sido extraído. ¿Cuál es la relación de dicha persona con la persona que antes era el organismo? La respuesta más obvia parece ser que ambas eran la misma persona y que al separar el cerebro del organismo la persona ha sido extraída con él. Sin embargo, los partidarios del animalismo no pueden aceptar esta última afirmación. Si cierto tiempo después volviéramos a introducir el cerebro en el que fue su cuerpo y volviéramos a conectarlo de la manera adecuada con este, nos encontraríamos con que volvemos a tener el organismo original en el mismo estado en que lo teníamos antes. Si

al principio el organismo era la única persona existente y el único sujeto de los pensamientos de dicho cerebro, lo razonable sería afirmar que eso vuelve a ser así. ¿Qué ha sucedido entonces con la persona remanente? De nuevo, la mejor respuesta parece ser incompatible con el animalismo. Lo más razonable parece ser suponer que hemos trasladado a la persona remanente junto con el cerebro y que es la persona que ahora controla de nuevo el que al principio era su cuerpo. Como señala Johnston (2016), para llevarnos a la conclusión de que el animalismo es erróneo el argumento de la persona remanente sólo necesita dos afirmaciones adicionales que resultan muy plausibles. La primera de estas afirmaciones es que cortar tejido no puede crear una nueva persona, a no ser que dicho tejido impida de algún modo el adecuado funcionamiento del cerebro. La segunda es que juntar tejido no puede resultar en la destrucción de ninguna persona, a no ser que se haga de algún modo que imposibilite el adecuado funcionamiento del cerebro. Existe la posibilidad de afirmar que antes de la extracción ya existían dos personas, pero si esto es así, nos encontramos de nuevo con un problema análogo al del animal pensante. Olson (2016) analiza algunas de las principales respuestas que los animalistas han ofrecido a este problema y, aunque no parece considerarla una respuesta completamente satisfactoria, se decanta por aquella según la cual el cerebro pasa a ser una persona en el momento en que es desconectado del organismo. Por lo tanto, el autor rechaza la afirmación de que cortar tejido no puede provocar la creación de una nueva persona. Puede parecer que esta afirmación es extremadamente implausible, pero Olson señala que las principales alternativas al animalismo tienen que hacer frente a un problema análogo al de la persona remanente. Los autores que se oponen al animalismo suelen afirmar que en lugar de ser numéricamente idénticos a nuestro organismo estamos constituidos por él. Según este punto de vista, somos entidades distintas a nuestro organismo pero nuestra existencia no consiste en otra cosa que en la de dicho organismo. El autor señala que, de acuerdo con este punto de vista, cuando el cerebro aún está conectado al resto del cuerpo no constituye a la persona, sino que es el organismo entero el que lo hace. Si cuando desconectamos el cerebro la persona es extraída junto con él, entonces el cerebro pasa repentinamente a constituir una persona, es decir, pasa a constituir un ser pensante. En opinión de Olson, el que pueda hacerse que el cerebro constituya un ser pensante sólo con desconectarlo del cuerpo es tan misterioso como el que se pueda crear mediante la misma acción un nuevo ser pensante. Sin embargo, Johnston (2016) señala que, cuando aún permanece en el cuerpo, el cerebro constituye parcialmente a la persona y argumenta que no hay nada implausible en el hecho de que una vez separado del cuerpo pase a constituir la por completo. Al fin y al cabo, del mismo modo que sucede con cualquier tipo de objeto o entidad, la constitución de una persona puede ser modificada si se le añade o se le extrae materia. Rory Madden (2016) ofrece una respuesta similar al problema de la persona remanente. En su opinión no hay nada inverosímil en afirmar que cortar tejido puede crear una nueva entidad. El autor señala que cuando creamos una estatua a partir de un bloque de mármol también estamos creando una

nueva entidad sólo con eliminar materia. Madden también argumenta que cuando cortamos una parte de una planta y la replantamos estamos obteniendo una nueva planta, aunque los procesos que tengan lugar en ella sean los mismos que se daban cuando sólo era una parte de la planta original. Sin embargo, en el caso del cerebro las cosas son más complicadas. Supuestamente, antes de la operación el organismo es el sujeto de experiencias. Cuando el cerebro es extraído y mantenido con vida artificialmente, las experiencias a que da lugar surgen a causa de los mismos procesos que las producían antes. Resulta muy difícil de creer que, tras la operación, el sujeto de esas experiencias pueda ser una entidad pensante completamente nueva que además desaparecería en cuanto el cerebro volviera a conectarse al organismo. Olson (2016) señala también la posibilidad de argumentar que el cerebro continúa siendo parte del organismo aunque lo desconectemos de él, pero, como él mismo advierte, si conectáramos un nuevo cerebro a dicho cuerpo, nos encontraríamos con que el organismo tendría dos cerebros a la vez o, en caso de que se pretenda que al conectar un nuevo cerebro el original deja de pertenecerle, tendríamos de nuevo una persona remanente cuyo origen no podemos explicar. El autor considera también la propuesta de van Inwagen según la cual los conjuntos de partículas sólo constituyen un objeto si sus actividades dan lugar a un organismo vivo. Según este punto de vista, el único tipo de entidad que puede ser considerada como un sujeto de experiencias es el organismo en su conjunto. El cerebro desconectado del cuerpo no constituye, según esta opinión, el tipo de unidad necesaria para que pueda ser considerado un sujeto de pensamientos. Lo único que tendríamos en dicho caso sería un conjunto de partículas distribuidas de manera que tengan la forma de un cerebro, pero no habría ningún objeto ni entidad propiamente dichos. Sin embargo, resulta difícil de creer que un cerebro en el que se dieran los procesos habituales dejaría de servir de base a pensamientos únicamente por haber sido desconectado del organismo. Si dicho cerebro pudiera recibir suficiente información del exterior no hay motivo para pensar que en él dejarían de producirse pensamientos. Podría decirse que los pensamientos que podrían tener lugar en dicho cerebro carecerían de pensador, pero, como argumenta Johnston (2016), si los pensamientos que tienen lugar en un cerebro desconectado del organismo pueden tener lugar sin un pensador, no hay motivo para pensar que los pensamientos que tienen los organismos normales necesiten uno. Además, contra la propuesta de van Inwagen, Johnston señala que aunque la vida resulte espectacular no hay ningún motivo para pensar que los únicos objetos complejos que pueden existir son los animales, de modo que el partidario del criterio psicológico podría dar al problema del animal pensante una respuesta análoga y afirmar que no existen organismos, sino sólo partículas agrupadas en forma de organismos. Otra posibilidad es afirmar que al extraer el cerebro del cuerpo extraemos el organismo junto con él. Hay varias maneras en que esta afirmación podría hacerse. Podría

afirmarse que el organismo es el cerebro desde un primer momento. Si un partidario del animalismo afirmase esto, estaría diciendo que somos nuestro cerebro⁶⁸. Sin embargo, esta idea es muy poco plausible. Como se señala a menudo, resulta difícil de creer que cuando hablamos de nosotros mismos estemos hablando de nuestro cerebro, puesto que acostumbramos a atribuirnos propiedades que este no posee. Más implausible todavía resulta la afirmación de que el organismo es el cerebro. Un cerebro, a pesar de la importancia que tienen sus funciones para el correcto funcionamiento del organismo, es simplemente un órgano, del mismo modo que lo son el riñón o el hígado. Como señala Johnston (2016), la manera en que el organismo está vivo es distinta a la manera en que lo está el cerebro y la manera en que cada uno deja de estarlo es también diferente, por lo que no pueden ser numéricamente idénticos uno con el otro. Otra opción sería afirmar que en un primer momento el organismo está constituido por el conjunto del cuerpo y que, cuando el cerebro es desconectado de él, pasa a estar completamente constituido por este último. Sin embargo, el cuerpo vacío que queda tras la extracción parece un mejor candidato para ser el organismo. El cerebro no es más que un órgano, mientras que el cuerpo que deja tras de sí es un organismo al que le falta un órgano. Si el animalismo optara por este tipo de respuesta podríamos preguntarle por qué es especial el cerebro. Si extrajéramos cualquier otro órgano, como el hígado o uno de los riñones, no consideraríamos que él pase a ser el organismo. El motivo por el que eso no parece tan extraño en el cerebro es que dicho órgano es el portador de la continuidad psicológica. Sin embargo, en opinión de los partidarios del animalismo, la continuidad psicológica es irrelevante para la identidad personal. ¿Por qué, entonces, el hecho de que un órgano sea el portador de dicha continuidad iba a convertirle en el organismo tras desconectarlo del resto del cuerpo? Stephan Blatti (2016) trata de responder al argumento de manera distinta. En su opinión, no hay ninguna persona remanente. El autor argumenta que el comportamiento que pudiéramos observar en el cerebro separado del cuerpo no se parecería lo suficiente al nuestro como para que pudiéramos tener evidencia suficiente en favor de la idea de que este estaría teniendo pensamientos y experiencias. Sin embargo, aunque actualmente no sea técnicamente posible, no hay ningún motivo para pensar que no podría hacerse llegar al cerebro información del exterior a través de dispositivos capaces de emular los sentidos. En dicho caso, no parece haber motivo para suponer que no se darían en él los mismos procesos que se dan en cualquier cerebro, procesos que presumiblemente resultarían en experiencias conscientes. Si el cerebro volviera a conectarse al organismo, es de suponer que la persona recordaría las experiencias que vivió antes de la reconexión. Parece muy difícil negar la existencia de la persona remanente, y esta supone uno de los principales problemas para la concepción animalista.

⁶⁸ Parfit (2004) argumenta contra la opinión de Nagel de que somos nuestros cerebros. Sin embargo, la teoría de Nagel que examina Parfit no era animalista y las objeciones que este dirige contra ella, aunque logran socavar los argumentos que Nagel adujo en su favor, no muestran su falsedad si tenemos en cuenta el argumento del animal pensante.

En uno de sus últimos trabajos sobre la identidad personal, Derek Parfit (2012) argumenta en contra del animalismo. El autor empieza apelando a la intuición del trasplante que hemos examinado antes, tras lo cual expone el problema de la persona remanente. Para responder a aquellos animalistas según los cuales el organismo va con el cerebro, al exponer este argumento Parfit plantea, en lugar de la extracción de todo el cerebro, la de una porción de este que baste para que se mantenga la conciencia. Un problema para el animalismo relacionado con el de la persona remanente que Parfit también examina es el conocido como problema de las partes pensantes. Puesto que el cerebro, o la parte consciente de este, podría pensar si estuviera desconectado del cuerpo y fuera mantenido con vida artificialmente, es razonable suponer que también puede hacerlo mientras está conectado con el resto del organismo. Entonces nos encontramos con un problema epistémico similar al que plantea el argumento del animal pensante. Como ya hemos comentado antes, si cierta parte de nosotros es capaz de pensar, no tenemos manera de saber si somos el organismo o la parte pensante de este. Puesto que cualquier pensamiento es tenido por ambos, cualquier razón que el organismo pudiera tener para creer que él es el organismo es una razón que la parte pensante tendría para pensar lo mismo. Parfit considera que la mejor opción para resolver todos estos problemas a que tiene que hacer frente el animalismo es considerar que somos numéricamente idénticos a las partes pensantes. El autor señala que la parte pensante es un candidato mucho mejor que el organismo para ser el sujeto de las experiencias y argumenta que si parece absurdo que dicho sujeto sea una porción del cerebro, puede responderse que la parte pensante no es esa porción cerebral, sino que está relacionada con ella de la misma manera en que el animal lo está con el cuerpo. Según el punto de vista de Parfit, por lo tanto, la persona sería la parte del organismo que piensa. A la objeción que hemos comentado antes según la cual nos atribuimos propiedades que nuestros cerebros o nuestras partes pensantes no poseen, el autor responde que nos atribuimos propiedades que pertenecen a nuestro cuerpo porque nos encontramos en una relación muy íntima con él. Parfit opina que el pronombre “yo” es ambiguo y que unas veces lo usamos para referirnos al animal y otras para referirnos a la persona. En opinión del autor, los casos imaginarios demuestran que es el sentido interno del pronombre al que le damos más importancia. En cuanto al problema del animal pensante, Parfit argumenta que no hay dos procesos cerebrales distintos y que el único pensador es la parte pensante. Aunque el pronombre “yo” sea ambiguo, el autor señala que podemos definir un nuevo pronombre para referirnos sólo al animal y otro para referirnos a la parte pensante, de modo que desaparece cualquier posible problema epistémico.

Contra el problema de las partes pensantes, Stephan Blatti (2016) argumenta que no podemos atribuir pensamientos ni otras propiedades mentales a una única parte del organismo. En opinión del autor, cada órgano lleva a cabo su función en relación con el resto del organismo, y, del mismo modo que, siguiendo el ejemplo de Blatti, no tendría sentido atribuir

independientemente al corazón la acción de circular la sangre, lo correcto sería afirmar que es el animal el que piensa por medio del funcionamiento de su parte pensante. Este argumento se derrumba si se admite la posibilidad de que los pensamientos continúen teniendo lugar una vez extraída la parte pensante. Como hemos visto, Blatti no cree que eso sea posible. La plausibilidad del ataque del autor al problema de las partes pensantes dependerá, por lo tanto, de la de su argumento contra el problema de la persona remanente, argumento que ya hemos analizado y que, en mi opinión, resulta poco convincente. Rory Madden (2016) trata de resolver el problema de manera distinta. Según el autor, el sujeto consciente será aquella entidad en cuya coordinación consista la función de las partes pensantes. Madden señala que la función de las partes pensantes es coordinar los inputs y outputs del organismo en su conjunto, y no únicamente los del cerebro o los de una porción de este. Es cierto que si consideramos el cerebro aisladamente, podemos decir que este se encarga de coordinar sus propios inputs y outputs, pero, puesto que el cerebro contribuye a garantizar la supervivencia del organismo en su conjunto, el autor concluye que su función es coordinar a este último. Madden considera la posibilidad de que se objete que las actividades de las partes pensantes también contribuyen a la persistencia del cerebro, pero responde que lo que explica la existencia de dichas partes es el papel que juegan en la coordinación del organismo en su totalidad, por lo que su función es coordinar a este último. Como hemos visto antes, el autor responde al problema de la persona remanente afirmando que el cerebro se convierte en un nuevo ser pensante una vez es extraído del cuerpo. Es de suponer que Madden considera que, una vez realizada la operación, la función de las partes pensantes pasa a ser coordinar únicamente el cerebro. Sin embargo, la presencia de las partes pensantes no se explicaría ya, en este caso, por dicha función. Además, si pudiéramos suplir adecuadamente la información que el cerebro solía recibir por medio del organismo, los procesos que se darían en este serían los mismos que tenían lugar antes. ¿Cómo puede ser posible que la función del cerebro haya cambiado si lo que sucede en él continúa siendo lo mismo? Supongamos que podemos conectar el cerebro a un cuerpo robótico que simule el funcionamiento del organismo. ¿Quién sería ahora el sujeto de experiencias? La presencia de las partes pensantes no se explica, de acuerdo con lo dicho por Madden, ni por su función de coordinar los inputs y outputs del cerebro ni por la de coordinar los del cuerpo artificial, de modo que si los argumentos del autor son correctos, nos quedaríamos sin ningún criterio para decidir cuál es el sujeto consciente en dicha situación. Además, también podríamos poner en duda la afirmación de que la presencia de las partes pensantes se explique por su contribución a la persistencia del animal. Podríamos decir, por ejemplo, que la existencia de las partes pensantes del cerebro se explican por su contribución a la supervivencia de dicho órgano. Es cierto que la persistencia del órgano es importante precisamente por contribuir a la de todo el organismo, pero eso no parece motivo suficiente para que estemos obligados a definir la función de una pequeña parte refiriéndonos a la totalidad del organismo.

Eric Olson (2015) respondió al artículo de Parfit centrándose más en argumentar contra el punto de vista de este que en responder a sus objeciones contra el animalismo. Al principio de su artículo, Parfit (2012) expone su criterio de identidad personal. De acuerdo con dicho criterio, si hay continuidad psicológica que se da de forma no ramificada y con su causa normal, la persona resultante es numéricamente idéntica a la original. Si no hay continuidad psicológica ni existencia continua de suficiente cantidad del mismo cerebro, la persona resultante no es la misma que la original. En el resto de casos, según Parfit, la identidad personal será indeterminada y la pregunta de si la persona resultante es la misma que la original será, por lo tanto, vacía⁶⁹. Olson empieza su crítica señalando la existencia de una inconsistencia en el criterio de identidad de Parfit. Como señala el autor, si se trasplantara cada hemisferio cerebral a un cuerpo distinto y se modificara uno de ellos para que perdiera la continuidad psicológica, nos encontraríamos con que una de las dos personas resultantes sería numéricamente idéntica a la original mientras que sería indeterminado si la otra lo sería o no. Si tenemos una persona que es la misma que la original, debería ser determinado que no existe ninguna otra persona que lo sea. Tal vez Parfit podría responder que si ya hay una persona determinadamente idéntica a la original, cualquier otra persona es distinta a ella, pero Olson podría responder que se trata de una respuesta *ad hoc*. Probablemente Olson tenga razón en su crítica al criterio de Parfit, crítica que no supone ningún problema contra la idea que he defendido antes de que la identidad personal es siempre indeterminada. A continuación, Olson se centra en refutar la teoría de que somos la parte pensante del organismo. Como señala Olson, según dicha teoría sólo la parte pensante puede pensar no derivativamente, mientras que cualquier entidad que posea partes que no piensen lo hará de forma derivativa. Contra este punto de vista, el autor argumenta que, si se acepta dicha teoría, habrá de aceptarse una afirmación análoga para cualquier otra actividad. Tendremos que decir, por ejemplo, que el animal no camina, o que lo hace sólo de forma derivativa, siendo la parte caminante la única que lleva a cabo dicha acción. De manera análoga, habrá que afirmar que el animal ni come, ni respira, ni habla, ni realiza ninguna otra acción de la que se ocupe una parte específica del organismo. Como señala el autor, si se aceptan estas afirmaciones, tendremos, en lugar de un único ser que hace muchas cosas, una gran cantidad de entidades cada una de las cuáles lleva a cabo una única acción. Nosotros, concretamente, no haríamos otra cosa que pensar. Es más, Olson argumenta que, puesto que las diferentes partes del cerebro están dedicadas a la realización de distintas funciones, no existiría ni siquiera una única entidad pensante, sino múltiples entidades cada una de las cuales desempeña una sola de las funciones cerebrales. No hay duda de que las conclusiones a las que nos lleva dicha teoría resultan muy poco creíbles. Sin embargo, Olson argumenta que no puede hablarse de una única

⁶⁹ En *Razones y personas*, Parfit parecía opinar que si no hay continuidad psicológica, no hay identidad. Aquí el autor afirma que si hay continuidad de suficiente cerebro, la identidad es, en ausencia de continuidad psicológica, indeterminada.

parte que lleve a cabo cada función. Al fin y al cabo, en el caso del pensamiento son muchas las partes del organismo que lo hacen posible, permitiendo que la sangre pueda llegar al cerebro o que dicha sangre contenga aquello que este necesita. Además, como señala el autor, el enfoque de Parfit requiere que podamos decir cuál es exactamente la parte del organismo que se encarga de pensar. Según Olson, por mucho que conociéramos todos los hechos no podríamos determinar cuáles son exactamente las moléculas involucradas en la actividad de pensar, de modo que no podemos decir cuál es exactamente la parte del cerebro a la que se supone que somos numéricamente idénticos. El autor señala, además, que si somos una porción de nuestro cerebro, no podemos tener condiciones de persistencia psicológicas, puesto que dicha porción existía antes de que en ella tuviera lugar actividad mental. Olson considera la posible respuesta de que somos partes pensantes en funcionamiento, pero señala que si un cerebro y un cerebro en funcionamiento son entidades distintas, entonces pueden existir infinitas entidades distintas compuestas de la misma materia, como, por ejemplo, cerebros despiertos o cerebros que piensan en un tema determinado. Si esto fuera así, en el momento en que todas esas entidades coinciden no podríamos saber cuál de ellas somos y nos encontraríamos, por lo tanto, con un problema epistémico mucho peor que el que plantea el argumento del animal pensante. La teoría que Parfit propone para resolver el problema de las partes pensantes y el del animal pensante se enfrenta, por lo tanto, a más problemas que aquellos con que se encuentra el animalismo, por lo que será necesario dar con otra alternativa. Antes de examinar las distintas soluciones que los detractores de la concepción animalista han ofrecido al problema del animal pensante analizaremos una última objeción a este punto de vista.

Lynne Rudder Baker (2016) utiliza el hecho de que los organismos tengan que estar compuestos de materia orgánica para tratar de refutar la concepción animalista. La autora señala que, de acuerdo con el punto de vista animalista, si se cambia alguna parte del organismo por un sustituto inorgánico, este no será parte del animal. Entonces, Baker argumenta que podría realizarse la sustitución de manera que ya no pudiera hablarse de una entidad orgánica, tal vez porque sus funciones tuvieran que pasar necesariamente por la parte robótica o porque esta fuera la que mantuviera conectadas las distintas partes del organismo de modo que, sin ella, nos viéramos obligados a hablar de varias entidades orgánicas distintas. En un caso como ese, si el animal en cuestión fuera un ser humano, el animalismo nos llevaría a la implausible conclusión de que la persona, a pesar de las apariencias, habría dejado de existir. Si, en cambio, consideramos que la persona está constituida por el organismo, podríamos decir, como señala Baker, que la persona pasaría a estar constituida tanto por partes orgánicas como por partes inorgánicas.

A pesar de la existencia de estas importantes objeciones contra el animalismo, el problema del animal pensante sigue siendo un muy buen argumento en favor suyo. Los detractores de esta teoría han propuesto distintas soluciones a dicho problema. Johnston (2016) argumenta que, aunque somos animales, podemos dejar de serlo sin que ello suponga el fin de nuestra existencia. En su opinión, el argumento del animal pensante demuestra únicamente que somos animales, no que no podamos dejar de serlo. El autor señala que si el argumento fuera correcto, podría usarse también para demostrar, por ejemplo, que un adolescente no puede dejar de serlo sin dejar de existir, cosa que es evidentemente falsa. En opinión de Johnston, por lo tanto, el ser animales sería simplemente una fase que en los casos actuales dura toda la vida de una persona, pero que podría terminar sin que ello supusiera la muerte. Para que pudiera establecerse que el animalismo es la teoría correcta faltaría demostrar, como afirma el autor, que si un ser es un animal, no puede dejar de serlo. Hay, sin embargo, un problema en el argumento de Johnston. Olson (2004) emplea otro argumento en favor de la concepción animalista que pone de manifiesto el error que comete Johnston con su objeción. Nuestro organismo empieza a existir como un feto que carece de las propiedades mentales que lo convertirían en una persona. Ninguno de nosotros somos psicológicamente continuos con el feto que fue nuestro organismo, de modo que, de acuerdo con el criterio psicológico, no somos él. Sin embargo, el feto no deja de existir, sino que continúa desarrollándose hasta adoptar la forma definitiva de nuestro organismo. Si nosotros no somos ese feto pero él sigue existiendo como nuestro organismo, entonces existen dos seres, nosotros y el organismo. Olson también emplea a menudo el ejemplo de aquellas personas que por algún tipo de enfermedad terminan su vida como un vegetal carente de propiedades mentales. De acuerdo con la definición de Locke, el vegetal no sería una persona, pero parece sensato afirmar que continúa siendo el organismo. Por lo tanto, si en todo momento hay un único ser, parece que el hecho de ser una persona es una fase por la que pasa el organismo. No podemos, entonces, afirmar que el hecho de ser un organismo es también una fase por la que pasa la persona, puesto que entonces nos encontraríamos con que no hay nada que seamos esencialmente. Si quiere mantenerse el criterio psicológico, hay que aceptar que hay dos entidades, el organismo y la persona, de modo que el problema del animal pensante sigue estando sin resolver. Noonan (1989) trata de resolver el problema afirmando que tanto el animal como la persona piensan los mismos pensamientos pero el referente del pronombre “yo” es siempre la persona. Sin embargo, es difícil ver por qué el referente del pronombre es la persona y no el animal. Parece que quien utiliza el pronombre “yo” lo emplea para referirse a sí mismo y que, por lo tanto, si el animal usa el pronombre, él es su referente. Si hay algo en su significado que impida que el animal lo use para pensar en sí mismo, podríamos emplear una nueva palabra “yo*” cuyo referente sea siempre aquel ser que la profiera o la piense. En este caso, si pensamos “yo* soy la persona”, no podemos saber si la frase es cierta o no. El problema epistémico que plantea el argumento del animal pensante sigue en pie. Baker

(2016) propone una solución más compleja. En su opinión, la persona piensa los pensamientos de manera no derivativa mientras que el animal que la constituye lo hace derivativamente. El único de los dos que tiene una perspectiva de primera persona que le permite emplear la palabra “yo” con sentido es la persona, el organismo sólo tiene esas propiedades mentales derivativamente. De este modo, en sentido estricto, hay una sola persona y un único sujeto de experiencias que tiene los pensamientos en cuestión. Es difícil ver, sin embargo, por qué es la persona la que tiene los pensamientos de manera no derivativa. Cuando hemos examinado el argumento de las partes pensantes, la intuición parecía decirnos que era la parte pensante la que tenía los pensamientos de manera no derivativa, en lugar del animal de que formaba parte. ¿Realmente en este otro caso la intuición nos lleva en sentido inverso? La opinión de Baker es que el organismo piensa sólo en virtud de constituir una persona, pero, ¿por qué no decir que es la persona la que piensa derivativamente en virtud de estar constituida por un organismo que tiene un cerebro y que lo utiliza para pensar? Si realmente existen dos entidades, la persona y el animal, ¿qué hay que justifique la afirmación de que el sujeto de experiencias es una en lugar de la otra? Sydney Shoemaker (2016) trata de ofrecer una explicación de por qué la persona puede pensar pero el animal no. El autor considera que las personas estamos, en los casos habituales, constituidas por un animal, pero opina que dicho animal carece de propiedades mentales. Shoemaker argumenta que una propiedad es individuada por el papel causal que juega y argumenta que si la etapa anterior de una propiedad mental pertenece a un animal distinto de aquel al que pertenece la etapa posterior, entonces la propiedad mental en cuestión pertenecerá no al animal sino a la persona, que es una entidad que puede pasar de estar constituida por un animal a estarlo por otro y a cuya historia pertenecen los roles funcionales de las propiedades mentales de dichos animales. Podemos utilizar como ejemplo el caso del teletransporte. En dicho caso, según el animalismo, la persona resultante es distinta de la original. Esto es así porque, como señala Olson (2002), la réplica sería un nuevo organismo y no podemos dejar atrás nuestro organismo si somos numéricamente idénticos a él. De acuerdo con el punto de vista de Shoemaker, en cambio, la réplica sí que sería la persona original. Esta habría pasado de un organismo a otro y es a ella, y no a ninguno de los dos animales implicados, a quien podemos atribuir las propiedades mentales. Esto último es así porque el papel causal de dichas propiedades puede estar entre ambos animales, teniendo, por así decir, su input en uno de ellos y su output en el otro. Es por ello que el sujeto de esas propiedades sería, según el autor, la entidad que persiste tras el teletransporte y no el organismo que la constituye en cada momento. Sin embargo, la solución de Shoemaker presenta, del mismo modo que las anteriores, algunos problemas. Supongamos que en el momento de crear la réplica no se creara una sino varias. Tendríamos en ese caso varias personas distintas. ¿A cuál de ellas es a la que pertenecían las propiedades mentales que exhibía el organismo original? En mi opinión, lo que sucede es que hemos replicado las propiedades mentales en cuestión. No podemos decir, en sentido estricto,

que haya una persona que continúa teniendo esas propiedades. No tiene sentido afirmar que el rol causal de las propiedades mentales de la persona constituida por el organismo original se extiende más allá del momento del teletransporte. Una propiedad mental puede tener muchos efectos sin que estos formen parte de su rol funcional. Por ejemplo, una persona podría decidir actuar de una manera o de otra a causa de haber observado la presencia de cierta propiedad mental en otra persona. En un caso como este, no pretenderíamos que el papel causal de dicha propiedad se extendería de una persona a la otra. En el caso del teletransporte, la propiedad mental del organismo original tiene efectos en el organismo resultante, pero esto es así porque se ha creado una réplica exacta de ese organismo con el objetivo, entre otros, de que tenga propiedades mentales exactamente idénticas. Pretender que eso justifica que extendamos el papel causal de las propiedades en cuestión a ambos organismos es tan absurdo como pretenderlo en el ejemplo de que hemos hablado unas líneas más arriba. No podemos, por lo tanto, recurrir al rol funcional de las propiedades mentales para mostrar que, en caso de que el animal y la persona sean entidades distintas, es la última la única a la que podemos atribuir las. Si no tenemos un muy buen motivo para afirmar que los animales no pueden ser sujetos de pensamiento, entonces, como afirma Olson (2002), estamos defendiendo la extremadamente implausible idea de que existen dos entidades que están compuestas exactamente de la misma materia, que se encuentran exactamente en el mismo entorno y que, a pesar de ello, tienen distintas propiedades. Si queremos, pues, resolver el problema del animal pensante, tendremos que probar por otro camino. Antes de tratar de encontrar una solución adecuada al problema, examinaremos brevemente otro problema que afecta a cualquier teoría según la cual los sujetos de experiencias sean seres materiales. Dicho problema afecta tanto a la concepción animalista como a la idea de que las personas son entidades distintas a los animales pero constituidas por ellos. Tras exponer el problema trataré de ofrecer una solución que resuelva tanto este último como el del animal pensante.

Hemos visto antes que Peter Unger defendía el punto de vista de que las personas no existen. Sin embargo, posteriormente cambió de opinión y pasó a sostener la tesis de que somos almas inmateriales. En favor de este punto de vista, Unger (2007) utiliza un argumento al que se refiere como el problema de los muchos (“the problem of the many”)⁷⁰. El autor expone el problema utilizando las nubes como ejemplo. Las nubes están formadas por conjuntos de gotas de agua microscópicas en suspensión. La densidad de la nube va reduciéndose de manera gradual a medida que nos alejamos de su centro y, aunque desde la distancia sus fronteras parezcan fijas, es imposible determinar cuál es el conjunto exacto de las gotas que forman parte de una nube concreta. Hay una gran cantidad de conjuntos de gotas que son candidatos

⁷⁰ En su primera exposición del problema, Unger (1980) lo utilizó para defender su nihilismo ontológico, aunque posteriormente lo vio como un argumento en favor de la tesis de que somos entidades que existen separadamente.

igualmente buenos a ser la nube. Si consideramos que uno de esos conjuntos es una nube, tenemos los mismos motivos para afirmar que cualquiera de los otros lo es. Por lo tanto, si consideramos que existe una nube en un determinado lugar, tenemos que aceptar que existen muchas. Como señala Unger, el mismo argumento puede aplicarse a cualquier tipo de entidad material. Sencillamente, en lugar de hablar de gotas de agua en suspensión, tenemos que hablar de partículas o moléculas. No es posible, por ejemplo, determinar el conjunto exacto de partículas que forman una determinada mesa. Hay muchos conjuntos de partículas que son candidatos igualmente buenos a ser la mesa, de modo que, si afirmamos que hay un conjunto concreto que sea la mesa, tenemos que aceptar que hay muchos conjuntos que lo son, de modo que si tenemos una mesa, entonces tenemos muchas. No es difícil ver que, como argumenta el autor, si un sujeto de experiencias es una entidad material, entonces el mismo problema puede aplicarse a los sujetos de experiencias y, por lo tanto, si ciertos pensamientos tienen un sujeto, hay muchos sujetos pensando exactamente los mismos pensamientos. En opinión de Unger, es posible aceptar que donde parece haber un solo objeto material hay en realidad muchos, pero no es defendible la idea de que existen muchos sujetos de experiencias involucrados en exactamente los mismos procesos mentales. El argumento de Unger, además, plantea un problema epistémico similar al que supone el problema del animal pensante. Si existen billones de sujetos de experiencias que piensan en todo momento los mismos pensamientos, no podemos saber cuál de ellos somos⁷¹. Cualquier motivo que podamos tener para pensar que somos uno de ellos en concreto es un motivo que el resto tienen para pensar lo mismo. Puesto que el argumento de Unger es aplicable a cualquier entidad material y puesto que según el animalismo somos organismos y, por lo tanto, entidades materiales, dicha teoría no logra evitar el problema. Si rechazamos el animalismo nos encontramos con el problema de tener que aceptar la existencia de dos entidades que piensan en todo momento los mismos pensamientos, pero el problema de los muchos nos lleva a tener que aceptar, en caso de que el animalismo sea cierto o de que sea correcta cualquier otra teoría que postule la existencia de pensadores materiales, la existencia de billones de sujetos de experiencias que experimentan siempre los mismos pensamientos. La solución que Unger propone al problema consiste en afirmar que los sujetos de experiencias no son entidades materiales, sino almas inmateriales. De este modo, es posible aceptar la existencia de muchos sistemas nerviosos parcialmente solapados relacionados con un único sujeto de experiencias inmaterial. ¿Debemos, entonces, aceptar un dualismo de substancias como propone Unger? Hemos visto que existe una gran evidencia en contra de la tesis de que somos entidades que existen separadamente. Si no queremos aceptar que seamos almas inmateriales ni que existan múltiples sujetos de experiencias involucrados en los mismos

⁷¹ Sin embargo, en el caso del argumento del animal pensante, el problema epistémico implica que, en caso de que el animalismo fuera falso, no podríamos saberlo. En el caso del problema epistémico que supone el argumento de Unger no sucede nada análogo.

procesos mentales, la única solución que nos queda es rechazar la existencia de sujetos de experiencias.

El punto de vista de Parfit acerca de la existencia de sujetos de experiencias es algo confuso. Como señala Shoemaker (1997), Parfit parece sugerir, al afirmar que las personas son sujetos de experiencias debido a la manera en que hablamos, que las personas son construcciones lógicas. Además, como afirma Shoemaker, el hecho de que Parfit afirme que es posible una descripción impersonal del mundo parece implicar que las experiencias no requieren, por sí mismas, la existencia de ningún sujeto. De hecho, la concepción reduccionista no parece ser compatible con que existan, en ningún sentido interesante, sujetos de experiencias. De acuerdo con esta teoría, como ya hemos visto, la existencia de una persona sólo consiste en la existencia de un cuerpo, un cerebro y una serie de sucesos físicos y mentales interrelacionados. Los sucesos mentales son los sucesos cerebrales o se encuentran relacionados con estos de algún modo. ¿En qué sentido puede decirse que la persona sea el sujeto de los sucesos mentales? Más bien parecen ser un resultado de su actividad. Supongamos que los sucesos mentales pueden identificarse con ciertos sucesos cerebrales. Entonces, esos sucesos serían algo que tiene lugar en el cerebro, no algo que el cerebro experimente. Si los sucesos mentales son distintos de los sucesos cerebrales pero están relacionados de algún modo con estos, tampoco está claro que podamos decir que sean experimentados por el cerebro. Parece más adecuado afirmar que los sucesos mentales tienen lugar en la persona o que tienen lugar como consecuencia del funcionamiento de su cerebro. En el caso de que fuéramos almas inmateriales, tendría sentido decir que somos sujetos de experiencias, tendría sentido la idea de que los sucesos mentales son algo que le sucede al alma. Sin embargo, si somos algún tipo de sustancia material y especialmente si nuestra existencia es reducible de la manera que postula el reduccionismo, la idea de un sujeto de experiencias parece extraña. Podemos llamar sujeto de experiencias a una entidad sólo porque en ella ocurren sucesos mentales o porque estos están causados por ella, pero este parece ser un sentido de la expresión muy distinto al que esta tendría si la aplicáramos a un alma inmaterial.

David Hume (1984) rechazó la existencia del sujeto de experiencias. Argumentaba que no percibimos ningún yo que sea dicho sujeto, sino únicamente percepciones particulares que no requieren la existencia de ningún pensador. En su opinión, en la sucesión de sucesos u objetos interrelacionados imaginamos la existencia de un sujeto continuo. Según Hume, no somos más que haces de percepciones, no existe ningún ser o entidad que persista a lo largo del tiempo. Sin embargo, hay algo que Hume no tuvo en cuenta. Parece difícil evitar la impresión de que las experiencias de que somos conscientes nos suceden a nosotros. Esto no se debe únicamente a las características de nuestro lenguaje. Podemos cambiar la expresión “me duele” por “hay

dolor”, pero con ello no conseguiremos dejar de atribuirnos dicha experiencia a nosotros mismos. Esa atribución de las experiencias a un sujeto que somos nosotros no es sólo una particularidad de nuestro lenguaje, sino que parece ser más bien algo que acompaña a las propias experiencias. Aun así, no parece haber nada en dichas experiencias que implique la existencia de un sujeto, y desde luego dicho sujeto no es algo que percibamos. Parece apropiada la idea que Shoemaker atribuye a Parfit de que las personas somos construcciones lógicas a las que atribuimos las experiencias y pensamientos de que somos conscientes. Parece que de algún modo construimos en nuestra mente un sujeto al que atribuimos los sucesos mentales. Aunque ello sea así, eso no significa que tal sujeto exista en sentido ontológico.

¿Qué es entonces el yo? ¿Qué es esa entidad que aparentemente persiste a lo largo del tiempo y que, si los defensores del criterio psicológico están en lo cierto, podría pasar de un organismo a otro? Si realmente queremos resolver el problema del animal pensante tendremos que dar con una explicación de qué es ese yo, una explicación que no implique la existencia de un sujeto de experiencias y que atribuya a dicho yo unas condiciones de persistencia distintas a las de los organismos. En la próxima sección analizaré una teoría del yo que tiene estas consecuencias y que es coherente con la idea que he defendido de que, en el caso de las personas, la identidad numérica es una propiedad relativa al hablante.

10. Daniel Dennett: El yo como producto biológico

Daniel Dennett (1995) defiende una teoría del yo según la cual este no es más que un producto biológico. Como señala el autor, todos los animales se construyen un yo biológico. Cualquier organismo se distingue a sí mismo del resto de elementos de su entorno. Gracias a eso, cuando un animal siente hambre no se le ocurre comerse ninguna parte de su cuerpo. Los humanos, además del yo biológico, nos construimos también un yo psicológico al que atribuimos distintas propiedades. Este yo es, de acuerdo con Dennett, un producto biológico del mismo modo que lo son la tela de la araña o las presas que construyen los castores. Del mismo modo que este tipo de elementos facilitan la supervivencia de los organismos en cuestión proporcionándoles protección o facilitándoles la obtención de alimento, el yo psicológico ayuda a los organismos humanos en su lucha por la supervivencia. En opinión del autor, el lenguaje tiene un papel muy importante en la construcción de este yo. Del siguiente modo explica Dennett (1995) cómo crea todo organismo humano este yo:

A partir de su cerebro teje una tela de palabras y de actos, y, como las demás criaturas, no tiene por qué saber qué está haciendo; sólo lo hace. Esa tela lo protege, como la concha del caracol, le proporciona el sustento, como la tela de la araña, y favorece sus perspectivas para el sexo, como la glorieta del pájaro jardinero. (p. 426)

Como vemos, Dennett señala que, del mismo modo que el resto de animales llevan a cabo sus construcciones ciegamente, sin ser conscientes de lo que están haciendo ni de su propósito, nosotros construimos nuestro yo de manera automática, sin darnos cuenta de ello. Este yo no es, según el autor, algo que podamos construir gracias a nuestra conciencia, sino que esta es más bien el producto de dicha construcción. Nuestro yo, por lo tanto, va mucho más allá del yo biológico existente en la gran mayoría de especies. Para explicar en qué consiste esta diferencia, esta particularidad del ser humano, Dennett (1995) nos dice:

El cangrejo ermitaño ha sido diseñado a fin de asegurar que se procure una concha. Su organización, podríamos decir, *implica* la concha, y, por tanto, en un sentido *muy* amplio, representa tácitamente al cangrejo como poseedor de una concha, pero el cangrejo, en ningún sentido estricto del término, se representa a sí mismo como poseedor de una concha. De ningún modo es un candidato para la autorrepresentación. ¿Para quién *se representaría a sí mismo* y por qué? No necesita recordarse a sí mismo este aspecto de su naturaleza, ya que su diseño innato se ocupa del asunto, y no hay nadie más en los alrededores que esté interesado.

[...]

Nosotros, por el contrario, estamos casi constantemente ocupados en presentarnos a nosotros mismos a los demás, o a nosotros, y, por tanto, en *representarnos* a nosotros mismos, con el lenguaje y el gesto, externo e interno. (p. 428)

El autor considera que lo que explica esta diferencia entre los seres humanos y otras especies es nuestra propia conducta. La diferencia entre nuestro entorno y el de otros organismos por medio de la cual el autor explica esta particularidad del ser humano es la presencia de palabras, las cuales “son unos poderosos elementos de nuestro entorno que incorporamos fácilmente, ingiriéndolas y excretándolas, tejiéndolas como telas de araña hasta construir secuencias de *narraciones* autoprotectoras” (Dennett, 1995, p. 428). En opinión del autor, construimos, sin darnos cuenta, nuestro yo por medio de estas narraciones. Nuestra técnica de autoprotección consistiría en contar a otros y a nosotros mismos historias sobre quiénes somos, de modo que tanto nosotros como los demás podamos imaginar la existencia de un sujeto al que podamos atribuir ciertas propiedades, tanto mentales como físicas. Dennett se refiere a ese agente unificado que suponemos como un centro de gravedad narrativa. Este centro de gravedad narrativa sería el yo, es decir, seríamos nosotros, y “como el yo biológico, este yo narrativo o psicológico es otra abstracción, no una cosa en el cerebro, pero, con todo, es un atraedor de propiedades muy robusto y casi tangible...” (Dennett, 1995, p. 429). Podría decirse que, en cierto modo, en el ser humano, el yo biológico y el yo psicológico forman una única abstracción, o tal vez sea mejor afirmar que, en cierto sentido, atribuimos el yo biológico al yo

psicológico. Como afirma el autor, al hablar de nuestro cuerpo o nuestro organismo no estamos queriendo decir que este se pertenezca a sí mismo, sino que atribuimos su propiedad al yo psicológico. Podríamos decir que creamos una abstracción que es nuestro yo psicológico y que la consideramos constituida por aquellos elementos que pertenecen a nuestro yo biológico. Consideramos todas las partes de nuestro cuerpo como parte de nosotros, pero estas pueden dejar de serlo. Si una parte de nuestro organismo es separada de este, dejamos de verla como perteneciente a nuestro yo psicológico. Del mismo modo, nuestro yo puede seguir existiendo tras la desaparición de todo nuestro organismo. Esto es así porque la persistencia del yo se debe a “la estabilidad que le da el tejido de creencias que lo conforman, y cuando estas creencias pasan, pasa el yo, permanente o temporalmente” (Dennett, 1995, p. 434). El autor se muestra de acuerdo con la analogía que Parfit emplea entre las personas y los clubs. Del mismo modo que el hecho de que un club siga siendo el mismo depende de que identifiquemos el club pasado con el club presente, el yo persiste si nos identificamos y nos identifican con la persona futura en cuestión. De este modo explica Dennett (1995) lo que es ese yo y por qué es tan importante:

Un yo [...] no es un viejo punto matemático, sino una abstracción que se define por la multitud de atribuciones e interpretaciones (incluidas las autoatribuciones y las autorrepresentaciones) que han compuesto la biografía del cuerpo viviente del cual es su centro de gravedad narrativa. Como tal, juega un papel particularmente importante en la economía cognitiva en curso de ese cuerpo viviente, porque, de todas las cosas del entorno sobre las cuales un cuerpo activo debe construir modelos mentales, ninguno es tan importante como el modelo que el agente tiene de sí mismo. (p. 437)

El autor señala que el hecho de que el yo sea una abstracción no implica que no existamos. Es verdad que en cierto sentido puede decirse que el yo es una ficción creada por nuestro cerebro, pero esa abstracción es algo que nuestro organismo ha creado y que es de gran importancia para él. Esta abstracción ha sido creada a partir de la continuidad existente en nuestros pensamientos y experiencias y consiste en suponer una entidad continua a la que todas esas propiedades mentales, así como las propiedades físicas del organismo, pertenezcan. Siendo esto así, nuestra persistencia “depende de la presencia de esa narración [...] que *en teoría* podría sobrevivir a una serie indefinida de cambios de *medio*, podría teletransportarse fácilmente [...] y almacenarse por tiempo indefinido en forma de mera información” (Dennett, 1995, p. 440). De este modo, nuestra continuidad a través del tiempo consiste en la continuidad de esa historia sobre nosotros mismos, y esta segunda continuidad depende de nuestras creencias, depende de que postulemos la existencia de un yo continuo y de que le adjudiquemos todas esas propiedades en las que vemos, a lo largo del tiempo, cierto patrón.

No es difícil ver que esta teoría es coherente con la idea de que la identidad personal es una propiedad subjetiva, relativa al hablante. Cuando decimos que una persona presente es la misma que cierta persona pasada estamos expresando que la identificamos con esa persona pasada y, por lo tanto, que postulamos un mismo yo continuo al que atribuimos las propiedades de ambas. Si conocemos todos los hechos, no hay modo de que podamos estar equivocados en uno de esos juicios de identidad. Esta teoría puede parecer cercana al punto de vista de Hume, que consideraba que imaginábamos la existencia del yo al observar ciertos sucesos mentales interrelacionados. Pero a diferencia de esta, la teoría de Dennett explica perfectamente porque no podemos evitar atribuir los sucesos mentales a un sujeto de experiencias. La suposición de este sujeto de experiencias forma parte de la manera en que funciona nuestra mente y nos es imprescindible. La creación de este yo forma parte, por así decir, de la manera en que estamos “programados”.

El argumento del animal pensante no supone ningún problema para la teoría de Dennett. Los organismos, debido a su funcionamiento, dan lugar, de algún modo, a ciertos sucesos mentales, pero no son los sujetos de dichas experiencias. En sentido estricto, no existe ningún sujeto de experiencias, pero hay cierto sentido en que sí que existe. El cerebro crea un yo psicológico al que convierte en el sujeto de experiencias. Es a dicho yo al que atribuimos todas nuestras propiedades mentales y al que consideramos el sujeto de los sucesos mentales que nos acaecen. Puede que haya quien considere que el hecho de que la actividad del organismo dé lugar a ciertos sucesos mentales basta para considerarlo un sujeto de experiencias. No hay problema en ello, siempre que se tenga en cuenta cuál es el sentido en que puede hablarse del animal como sujeto de experiencias. Johnston (2016) parece estar pensando en una solución similar al problema cuando afirma que la persona y el organismo son pensadores en distinto sentido. El autor afirma que en un organismo tienen lugar ciertos sucesos o procesos mentales y afirma que ese es el único sentido en que puede hablarse de este como siendo un pensador o un sujeto de experiencias. Cuando Johnston habla de la persona no está pensando en un yo psicológico como el que postula Dennett, pero está claro que la solución que puede ofrecerse desde la teoría de este último al problema del animal pensante es similar a la que Johnston tiene en mente. Existen ciertos organismos humanos en los cuales tienen lugar ciertos sucesos mentales. Estos organismos tienen condiciones de persistencia biológicas, incompatibles con cualquier versión del criterio psicológico de identidad personal o, más en general, con cualquier criterio en que la continuidad psicológica tenga o pueda tener un papel relevante. Como parte de su funcionamiento, estos organismos crean un yo psicológico. Este yo es una ficción, un agente al que poder convertir en el sujeto de los sucesos mentales que tienen lugar en el organismo. De este modo, el animal crea una narración a partir de la continuidad que observa en sus propiedades, una narración cuyo centro de gravedad es ese yo. Esta narración podría, aunque no

sucedan en los casos actuales, persistir más allá de la destrucción del organismo, cosa que provoca que el yo psicológico tenga condiciones de persistencia no biológicas. Nosotros, aunque estamos constituidos por un organismo, en el sentido de que vemos dicho organismo como parte de nosotros, somos algo distinto de ese organismo, algo que puede seguir existiendo cuando el animal que somos haya dejado de hacerlo. Somos, además, el único sujeto de experiencias, puesto que el animal, si lo es, lo es en un sentido muy distinto a aquel en que lo somos nosotros. De este modo queda resuelto el problema del animal pensante, y lo mismo sucede con el de las partes pensantes. Lo único que hay es ciertas partes del organismo involucradas de cierta forma en los procesos mentales, pero esas partes no son sujetos de experiencias en el sentido en que lo es el yo psicológico. El problema de los muchos que plantea Unger también queda resuelto. Ninguno de esos billones de candidatos igualmente buenos para ser el organismo es, en sentido estricto, un pensador. El único pensador que hay es el yo, que no es más que una abstracción. Si el problema planteado por Unger sigue siendo un problema por el hecho de que tenga como conclusión la existencia de billones de organismos donde sólo parece haber uno, entonces también lo es por postular la existencia de billones de mesas donde sólo parecía haber una. Si esta conclusión parece tan contraria a la intuición, o casi, como el nihilismo ontológico que propuso Unger según el cual nada existe, entonces tal vez exista una solución al problema. Exista o no tal solución, aceptar la existencia de billones de organismos no afecta de ningún modo a la teoría de Dennett.

Podría pensarse que puesto que, al fin y al cabo, el yo psicológico no es más que una abstracción, deberíamos considerar que nosotros somos el organismo, que es algo mucho más tangible y real. Sin embargo, en dicho caso, no seríamos más que un ser que teje un yo de manera tan ciega como la araña teje su tela. Seríamos un organismo en el cual tienen lugar ciertos sucesos mentales pero que no es realmente el sujeto de ningún tipo de experiencia. El yo psicológico creado por dicho organismo está mucho más cerca del concepto que tenemos de nosotros mismos. Puede que sea una abstracción, pero es también el sujeto de nuestras experiencias y el pensador de nuestros pensamientos. Una cascada está formada por la caída continua, por un mismo lugar, de muchas gotas de agua. No hay nada en la cascada que persista a lo largo del tiempo, esta es sólo una abstracción que creamos a partir de la continuidad que observamos en el caer del agua. Las gotas de agua que caen en un momento dado son algo mucho más real y tangible que la propia cascada y, sin embargo, esta última parece ser lo importante. Nos resulta mucho más útil saber que hay una cascada en cierto lugar que preocuparnos por las moléculas de agua que cayeron por ella en un momento dado. Del mismo modo, aunque el organismo sea algo mucho más real que el yo, es este último lo importante, y si pudiera pasar de un organismo a otro, haríamos bien en preocuparnos mucho más de lo que le

sucedan a nuestro yo que de lo que pase con un organismo particular. Al fin y al cabo, el yo es lo que somos, mientras que el organismo es algo sin lo cual podríamos seguir existiendo.

11. Conclusiones

Hemos llegado a una teoría de la identidad personal que difiere en muchos puntos de la de Parfit, pero que, a pesar de ello, tiene mucho en común con ella. El punto de vista que he defendido continúa siendo reduccionista. No hay nada más involucrado en la existencia continua de las personas que la existencia de un cuerpo, un cerebro y el darse de ciertos sucesos físicos y mentales interrelacionados. Es posible una descripción completa del mundo que sea impersonal, es decir, que no mencione a ninguna persona, salvo al describir los contenidos de ciertos pensamientos, ni atribuya los sucesos mentales a ningún sujeto de experiencias. Sin embargo, he defendido que no existe un criterio fijo de identidad personal, sino que la verdad de un juicio sobre la identidad numérica entre dos personas depende en parte del contexto psicológico de la persona que lo realiza. La identidad personal es, en este sentido, siempre indeterminada, no hay una respuesta a las preguntas sobre la identidad que sea objetivamente verdadera. El tipo de continuidad que suele tener más peso en nuestros juicios de identidad personal es la continuidad psicológica, aunque la continuidad física también tiene cierta importancia, especialmente cuando juzgamos la identidad de otras personas. La manera en que efectivamente realizamos los juicios de identidad tal vez se acerque al esquema del continuador más cercano propuesto por Robert Nozick (2003), aunque dotando de un mayor peso a la continuidad psicológica en el caso de la autoidentificación que el que tendría en el caso de los juicios acerca de la identidad personal de otros. Además, es de esperar que el peso dado a las distintas continuidades varíe dependiendo de quién realice el juicio de identidad en cuestión y de las circunstancias en que lo realice. En cualquier caso, esta es una cuestión empírica carente de implicaciones metafísicas. No existe, en sentido estricto, ningún sujeto de experiencias. Este es sólo, como defiende Dennett, una abstracción creada por nuestra mente a la que atribuimos propiedades mentales y a la que convertimos en el sujeto de nuestros pensamientos y experiencias. Es a dicha abstracción a lo que mejor se ajusta el concepto que tenemos de nosotros mismos, por lo que podemos concluir que somos dicha abstracción. Hemos llegado también a la conclusión de que Parfit tiene razón al defender que la identidad personal no es lo que importa. Lo que tiene la importancia que solemos atribuir a la identidad personal es, como sostiene Parfit, la relación R, es decir, continuidad y/o conexividad psicológica. Sin embargo, a diferencia de lo que creía Parfit, para que exista conexividad psicológica no es necesario que exista ningún suceso mental idéntico en ambas personas, sino que basta con que se den en ambas sucesos mentales suficientemente similares. Tampoco es necesario que exista, entre las dos personas R-relacionadas, ningún tipo de nexo causal. Lo que importa no es la causa sino el

resultado, de modo que la relación R es la relación que importa aunque se dé sin ninguna causa⁷². La verdad del reduccionismo podría tener como consecuencia que dejara de ser racional preocuparse por el propio futuro o, mejor dicho, por aquellas personas futuras que estén R-relacionadas con nosotros. Aunque, como le sucedió a Parfit, no he podido dar con ningún argumento que demuestre lo contrario, sí que creo haber demostrado que preocuparnos por dichas personas futuras es, por lo menos, racional de manera indirecta. Si no estuviéramos R-relacionados con ninguna persona futura, no podríamos pensar ni actuar y viviríamos como si nuestra existencia durara un único instante. Si no nos preocupáramos por las personas futuras que están R-relacionadas con nosotros, nos sucedería lo mismo. Por lo tanto, aunque puede que, en sentido estricto, sea irracional preocuparnos por el futuro de dichas personas como si fuera el nuestro, tenemos buenas razones para adoptar esa actitud irracional.

Parfit opina que de su teoría de la identidad personal se siguen ciertas consecuencias para la racionalidad y la moral. En particular, Parfit opina que su teoría conduce al abandono de la teoría clásica del propio interés y a la defensa del utilitarismo. Puesto que, en la práctica, la relación que importa es prácticamente la misma según mi punto de vista y el de Parfit, si realmente su teoría tiene estas implicaciones, parece de esperar que también las tenga el punto de vista que yo he defendido. Es cierto que mi argumento en favor de que la relación R importa sólo demuestra que importa en un sentido débil, pero Parfit tampoco ofrece un argumento que demuestre que dicha relación importe en un sentido más fuerte y la existencia de tal argumento no puede descartarse. Sería necesario examinar si de la teoría de Parfit se siguen realmente las conclusiones que él pretende extraer de ella⁷³. Sin embargo, dicha tarea nos pondría por delante un número de páginas similar al que ya hemos dejado atrás.

⁷² Debido a que esta relación se aleja bastante de la definida por Parfit, he propuesto llamarla similitud psicológica en lugar de relación R.

⁷³ Parfit (2004) argumenta que puesto que lo que importa es la relación R y esta puede darse en mayor o menor grado, la teoría del propio interés, según la cual debemos preocuparnos por igual de todos los momentos de nuestra vida, fracasa. El autor argumenta que, puesto que eso es así, deberíamos considerar moralmente a nuestros yo's futuros como si fueran otras personas, de modo que tendríamos responsabilidades morales para con ellos. En opinión de Parfit, este argumento demuestra que el paternalismo está justificado. Sin embargo, podría objetarse que, a efectos prácticos, el paternalismo es incorrecto. La mayor parte de las personas verían con malos ojos que, en el pasado, se hubiera interferido en su libertad aunque fuera para evitarles un mal que padecerían en el presente. Puesto que no podemos saber cuál será la opinión de la persona futura, lo más sensato es suponer que, como sucede en la mayor parte de los casos, esta será similar a la de la persona presente con que se encuentre R-relacionada, de modo que si actuáramos de manera paternalista estaríamos, seguramente, actuando contra los deseos no sólo de la persona presente sino también de la persona futura a la que pretendemos ayudar. Por lo tanto, el paternalismo sigue siendo incorrecto aunque la tesis reduccionista defendida por Parfit sea verdadera.

Referencias

- Baker, L. R. (2016). Animalism vs. constitutionalism. En S. Blatti y P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 50-63). Oxford: Oxford University Press.
- Blatti, S. (2016). Headhunters. En S. Blatti y P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 162-179). Oxford: Oxford University Press.
- Borges, J. L. (1983/2005). La memoria de Shakespeare. En *Obras completas II* (pp. 391-397). Barcelona: RBA.
- Braun-Latour, K. A., Latour, M. S., Pickrell, J. E., & Loftus, E. F. (2004). How And When Advertising Can Influence Memory For Consumer Experience. *Journal of Advertising*, 33(4), 7-25.
- Cotel, S. C., Gallo, D. A., & Seamon, J. G. (2008). Evidence that nonconscious processes are sufficient to produce false memories. *Consciousness and Cognition*, 17(1), 210-218.
- Curzer, H. (1991). An Ambiguity In Parfit's Theory Of Personal Identity. *Ratio*, 4(1), 16-24.
- Dennett, D. (1995). *La conciencia explicada* (S. Balari Ravera, Trad.). Barcelona: Paidós (Obra original publicada en 1991).
- Drivdahl, S. B., Zaragoza, M. S., & Learned, D. M. (2009). The role of emotional elaboration in the creation of false memories. *Applied Cognitive Psychology*, 23(1), 13-35.
- Ellis, R. (2000). Parfit and the buddha: Identity and identification in Reasons and persons. *Contemporary Buddhism*, 1(1), 91-106.
- Fara, D. G. (2000). Shifting sands: An interest relative theory of vagueness. *Philosophical Topics* 28(1), 45-81.
- Fields, L. (1987). Parfit on Personal Identity and Desert. *The Philosophical Quarterly*, 37(149), 432-441.
- Garrett, B. (1991). Personal Identity and Reductionism. *Philosophy and Phenomenological Research*, 51(2), 361-373.
- Gazzaniga, M. S., Ledoux, J. E., & Wilson, D. H. (1977). Language, praxis, and the right hemisphere: Clues to some mechanisms of consciousness. *Neurology*, 27(12), 1144-1147.
- Gendler, T. S. (2009). Personal identity and metaphysics. En B. P. McLaughlin, A. Beckermann, y S. Walter (Eds.), *The oxford handbook of philosophy of mind* (pp. 578-591). Oxford: Oxford University Press.
- Gillett, G. (1986). Brain Bisection and Personal Identity. *Mind*, 95(378), 224-229.

Hershenov, D. B., & Taylor, A. P. (2014). Split brains: No headache for the soul theorist. *Religious Studies*, 50(04), 487-503.

Hossack, K. (2006). Vagueness and personal identity. En F. MacBride (Ed.), *Identity and Modality* (pp. 221-241). Oxford: Oxford University Press.

Hume, D. (1984). *Tratado de la naturaleza humana* (F. Duque, Trad.). Barcelona: Orbis (Obra original publicada en 1738).

Johnston, M. (2016). Remnant persons: Animalism's undoing. En S. Blatti y P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 89-127). Oxford: Oxford University Press.

Lewis, D. (1983). Survival and Identity. *Philosophical Papers Volume I*, 55-72.

Lishman, W. A. (1971). Emotion, Consciousness and will after Brain Bisection in Man. *Cortex*, 7(2), 181-192.

Locke, J. (1956). *Ensayo sobre el entendimiento humano* (3ª reimpresión de la 2ª ed. en español; E. O'Gorman, Trad.). México: Fondo de Cultura Económica (Obra original publicada en 1690).

Mackie, D. (1999). Personal Identity and Dead People. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 95(3), 219-242.

Madden, R. (2016). Thinking parts. En S. Blatti y P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 180-207). Oxford: Oxford University Press.

Madell, G. (1985). Derek Parfit and Greta Garbo. *Analysis*, 45(2), 105-109.

Nagel, T. (1979). Brain bisection and the unity of consciousness. En *Mortal questions* (pp. 147-164). Cambridge: Cambridge University Press.

Noonan, H. W. (1989). *Personal identity* (2ª ed.). London: Routledge.

Nozick, R. (2003). Personal identity through time. En R. Martin y J. Barresi (Eds.), *Personal identity* (pp. 92-114). Malden, MA: Blackwell.

Olson, E. T. (1994). Is psychology relevant to personal identity? *Australasian Journal of Philosophy*, 72(2), 173-186.

Olson, E. T. (2002). What does Functionalism Tell Us about Personal Identity? *Nous*, 36(4), 682-698.

Olson, E. T. (2003). An argument for animalism. En R. Martin y J. Barresi (Eds.), *Personal identity* (pp. 318-334). Malden, MA: Blackwell.

Olson, E. T. (2004). Animalism and the Corpse Problem. *Australasian Journal of Philosophy*, 82(2), 265-274.

Olson, E. T. (2015). On Parfit's View That We Are Not Human Beings. *Royal Institute of Philosophy Supplement*, 76, 39-56.

Olson, E. T. (2016). The remnant-person problem. En S. Blatti y P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 145-161). Oxford: Oxford University Press.

Olson, Eric T (2017) "Personal identity", *Stanford Encyclopedia of Philosophy*. Disponible en: plato.stanford.edu/entries/identity-personal [Acedido: 11 de agosto de 2018].

Parfit, D. (1971a). On "The Importance of Self-Identity". *The Journal of Philosophy*, 68(20), 683-690.

Parfit, D. (1971b). Personal Identity. *The Philosophical Review*, 80(1), 3-27.

Parfit, D. (1982). Personal identity and rationality. *Synthese*, 53(2), 227-241.

Parfit, D. (1986). Comments. *Ethics*, 96(4), 832-872.

Parfit, D. (1993). The indeterminacy of identity: A reply to Brueckner. *Philosophical Studies*, 70(1), 23-33.

Parfit, D. (2003). The unimportance of identity. En J. Barresi, y R. Martin (Eds.). *Personal identity* (pp. 292-317). Malden, Mass.: Blackwell Publ.

Parfit, D. (2004). *Razones y personas* (Mariano Rodríguez, Trad.). Boadilla del Monte, Madrid: A. Machado Libros (Obra original publicada en 1984).

Parfit, D. (2012). We Are Not Human Beings. *Philosophy*, 87(01), 5-28.

Patihis, L., Frenda, S. J., Leport, A. K., Petersen, N., Nichols, R. M., Stark, C. E., ... Loftus, E. F. (2013). False memories in highly superior autobiographical memory individuals. *Proceedings of the National Academy of Sciences*, 110(52), 20947-20952.

Penrose, R. (2015). *La nueva mente del emperador* (J. García Sanz, Trad.). Barcelona: Debolsillo (Obra original publicada en 1989).

Persson, I. (2016). Parfit on Personal Identity: Its Analysis and (Un)importance. *Theoria*, 82(2), 148-165.

Putnam, H. (1983). Vagueness and Alternative Logic. *Erkenntnis*, 19(1/3), 297-314.

Raffman, D. (1994). Vagueness without paradox. *Philosophical Review*, 103(1), 41-74.

Read, R. (2015). The Tale Parfit Tells: Analytic Metaphysics of Personal Identity vs. Wittgensteinian Film and Literature. *Philosophy and Literature*, 39(1), 128-153.

- Schechter, E. (2015). The Subject in Neuropsychology: Individuating Minds in the Split-Brain Case. *Mind & Language*, 30(5), 501-525.
- Shoemaker, S. (1970). Persons and Their Pasts. *American Philosophical Quarterly*, 7(4), 269-285.
- Shoemaker, S. (1997). Parfit on identity. En J. Dancy (Ed.), *Reading Parfit* (pp. 135-148). Oxford: Blackwell.
- Shoemaker, S. (2016). Thinking animals without animalism. En S. Blatti y P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 128-141). Oxford: Oxford University Press.
- Sperry, R. W. (1966). Brain bisection and mechanisms of consciousness. En J. Eccles (Ed.), *Brain and conscious experience* (pp. 298-313). Berlín: Springer.
- Stanley, J. (2003). Context, interest relativity and the sorites. *Analysis*, 63(280), 269-281.
- Thomson, J. J. (1997). People and their bodies. En J. Dancy (Ed.), *Reading Parfit* (pp. 202-229). Oxford: Blackwell.
- Unger, P. (1979a). I Do Not Exist. En G. F. Macdonald (Ed.) *Perception and Identity* (pp. 235-251). London: Macmillan.
- Unger, P. (1979b). There are no ordinary things. *Synthese*, 41(2), 117-154.
- Unger, P. (1980). The Problem of the Many. *Midwest Studies In Philosophy*, 5(1), 411-468.
- Unger, P. (2007). *All the power in the world*. Oxford: Oxford University Press.
- Williams, B. (1986a). La continuidad corporal y la identidad personal. En *Problemas del yo* (pp. 35-42). Mexico: Universidad Nacional Autónoma de México (Obra original publicada en 1973).
- Williams, B. (1986b). La identidad personal y la individuación. En *Problemas del yo* (pp. 11-34). Mexico: Universidad Nacional Autónoma de México (Obra original publicada en 1973).
- Williams, B. (1986c). La imaginación y el yo. En *Problemas del yo* (pp. 43-66). Mexico: Universidad Nacional Autónoma de México (Obra original publicada en 1973).
- Williams, B. (1986d). ¿Son las personas cuerpos? En *Problemas del yo* (pp. 89-111). Mexico: Universidad Nacional Autónoma de México (Obra original publicada en 1973).
- Williams, B. (1986e). El yo y el futuro. En *Problemas del yo* (pp. 67-87). Mexico: Universidad Nacional Autónoma de México (Obra original publicada en 1973).

Williamson, T. (1997). Vagueness and ignorance. En R. Keefe y P. J. Smith (Eds.), *Vagueness: A reader* (pp. 266-281). Cambridge, Mass.: The MIT Press.

Wilson, D. H., Reeves, A., Gazzaniga, M., & Culver, C. (1977). Cerebral commissurotomy for control of intractable seizures. *Neurology*, 27(8), 708-715.

Wolf, S. (1986). Self-Interest and Interest in Selves. *Ethics*, 96(4), 704-720.