Racionalidad práctica

MARINA VELASCO

Razones internas *vs* razones externas. Reflexiones sobre una distinción*

Departamento de Filosofía.
Universidade Federal de Rio de Janeiro.

EN "INTERNAL AND External Reasons", Bernard Williams (1981) introdujo una distinción que ha dado muchísimo que hablar y ha producido muchísimos malentendidos. Con su defensa de la visión "internalista" (y su rechazo de la visión "externalista") de las razones para actuar, Williams se propone explícitamente defender una versión plausible de la concepción humeana de la motivación. Más concretamente, lo que pretende es negar que consideraciones de tipo moral o prudencial se apliquen *a priori* a la cuestión de "qué tipo de cosas un agente tiene razones para hacer". Aunque el tema no es, ciertamente, nuevo –ya que se trata de defender con nuevos argumentos la tesis humeana de que la razón es la esclava de las pasiones[1]–, hay que reconocer que el autor tiene la virtud de poner asuntos en la agenda. Parece que en este momento la polvareda comienza a asentarse, por lo que vale la pena preguntarse sobre los alcances de la polémica distinción: si se trata de una distinción útil, si verdaderamente vale la pena trazarla, o si, tal vez, se trata de una falsa dicotomía.

Después de discutir paso a paso el argumento de Williams, voy a concluir que si la distinción es interpretada en un sentido fuerte –que es el sentido que me parece más adecuado para los propósitos de Williams–, entonces plantea una falsa dicotomía, que no hay por qué aceptar.

* Parte de este trabajo (algunos párrafos textuales) ha sido publicada en un artículo mucho más extenso, donde comparo y critico las defensas de la teoría de la motivación humeana de Bernard Williams y de Michael Smith, en *Manuscrito – Revista Internacional de Filosofia*, Campinas, v. 26, n. 1, jan.-jun. 2003, pp. 135-182, con el título "Motivação Neo-Humeana: Por que acreditar nela?"

1. El *locus classicus* está en el *Treatise of Human Nature*: "La razón es, y sólo debe ser, esclava de las pasiones, y no puede pretender otro oficio que el de servirlas y obedecerlas" (Ed. Selby-Bigge, p. 415).

I

Williams introduce la distinción basándose en dos posibles interpretaciones de los enunciados que atribuyen razones para actuar a las personas. Así, los siguientes enunciados (en los que Φ representa un verbo de acción):

A tiene una razón para Φ

Existe una razón para que A Φ

serían susceptibles de dos interpretaciones[2]: en la interpretación *interna*, el enunciado (si es verdadero) implica que la persona tiene alguna motivación que se verá satisfecha o favorecida por el hecho de que haga A. La idea es que existe una condición relacionada con los objetivos [*aims*] del agente, y si ésta no se satisface, el enunciado resulta falso. Eso no sucede en la segunda interpretación, llamada *externa*. De acuerdo con esta segunda interpretación, el enunciado puede ser verdadero aun estando ausente la motivación adecuada en el agente.

Nótese que la distinción es introducida a partir de afirmaciones que hacemos *en tercera persona* sobre las razones para actuar de las personas. Como al pasar, Williams advierte que habla de "razones internas" y de "razones externas", como lo hace en el título del artículo, sólo por conveniencia[3], y, de hecho, nunca explica cómo este esquema para interpretar atribuciones de razones puede ser transformado en un esquema para clasificar a las razones mismas. Ésta es la fuente de muchas confusiones en la literatura posterior.

¿Por qué las razones internas implican la existencia de motivos y las externas no? Por definición, una razón es "interna" cuando es "relativa al conjunto motivacional subjetivo" [*subjetive motivational set*] de la persona. El contenido de este conjunto, *S*, en principio se deja abierto, aunque lo que no puede dejar de contener son deseos. Las razones internas son razones a las que se llega, por deliberación, a partir del conjunto motivacional subjetivo: ellas pueden motivarnos porque están conectadas con ese conjunto. Las razones externas, por el contrario, pretenderían ser verdaderas independientemente de su relación con los contenidos de ese conjunto, y por eso no queda claro cómo podrían motivar

2. No es necesario correlacionar las dos interpretaciones con las dos formas de enunciados mencionadas (es decir, pensar que el primer enunciado deba recibir la interpretación interna y el segundo la interpretación externa). De hecho, aclara Williams, a cada uno de los enunciados podría dársele tanto una interpretación interna como una externa.

3. "Es cuestión de investigar si existen dos tipos de razones para la acción, en oposición a dos tipos de afirmaciones sobre las razones de la gente para actuar" (Williams, 1981, p. 101).

a esa persona a actuar[4]. Establecida la distinción, Williams va a concluir que sólo existen razones internas, o, más precisamente, que todos los enunciados sobre razones externas son falsos[5]. Veamos el argumento.

Las razones internas son, de un lado, razones explicativas:

> Si hay razones para la acción, ha de ser un hecho que la gente actúa algunas veces por esas razones, y si lo hace, sus razones tienen que figurar en alguna explicación correcta de su acción (Williams, 1981, p. 102).

Pero, de otro lado, las razones internas no se limitan a la explicación, sino que se relacionan también con la racionalidad del agente. Ellas tienen que ser accesibles en primera persona. Sólo que la forma de la explicación no puede cambiar:

> Lo que podemos adscribirle en un enunciado sobre razones internas en tercera persona es también lo que él puede adscribirse a sí mismo como resultado de la deliberación (*Ibid.*, p. 103).

De modo que, para Williams, las razones internas tienen dos rasgos: ellas pueden explicar acciones, y ellas pueden justificarlas racionalmente. Si existe una ra-

4. Williams ilustra su argumento con el caso de Owen Wingrave, el personaje de un cuento de James. En esa historia, la familia de Owen Wingrave le insiste en la necesidad e importancia de que él se una al ejército, puesto que todos sus ancestros varones fueron soldados y el orgullo de la familia requiere que él también lo sea. Pero Owen Wingrave no tiene la más mínima motivación para unirse al ejército, y todos sus deseos lo llevan en una dirección distinta: odia la vida militar y todo lo que ella significa. Para su familia sería verdadero el enunciado: *existe una razón para que Owen se una al ejército*, esto es, no retirarían la afirmación aunque supieran que no hay nada en el *S* de Owen que pudiera llevarlo, por una deliberación racional, a creer en esa razón (*Ibid.*, p. 106). La razón "externa" es problemática porque lo que tiene que ser demostrado es cómo el agente podría llegar racionalmente a reconocer que tiene esa razón. Y sin demostrarse eso, aunque el agente realice la acción mencionada en la razón externa, no la estaría realizando por esa razón. (Owen Wingrave puede acabar uniéndose al ejército, después de todo, pero no por las razones que la familia cree que debería. Su razón sería otra. Tal vez el temor a la desaprobación de la familia. La razón "externa", entonces, ni explica esta acción de Owen ni la motiva).

5. En rigor, el argumento no es contra la posibilidad de las razones externas. Las razones externas son perfectamente reales: la gente suele usar ese tipo de enunciados. Lo que el argumento quiere mostrar es, más bien, que todos los enunciados sobre razones externas son falsos.

zón para hacer algo, esto es, una razón genuinamente práctica, (a) ella tiene que poder figurar en alguna explicación correcta de la acción, y además, (b) un agente que actúa por esa razón se comporta racionalmente, en el sentido de que podría llegar a reconocer que tenía esa razón como resultado de una deliberación.

Williams introduce mejoras y aditamentos al modelo que la tradición reconoce como humeano, al que considera demasiado simple y por eso inadecuado[6]. Según aquel modelo, como es sabido, las únicas consideraciones capaces de llevar a alguien a realizar una acción son aquellas que la representan como un medio para alcanzar aquello que el agente desea. Para esa concepción, que sería "el modelo más simple de razón interna", un agente sólo tiene una razón para Φ cuando tiene un deseo que satisface directamente haciendo Φ, o, indirectamente, cuando la realización de Φ se relaciona con ese deseo como un medio causal para un fin. Williams se propone ampliar y enriquecer ese modelo subhumeano "por ambos lados": por el lado de los contenidos que puede haber en el conjunto motivacional subjetivo del agente, y –también– por el lado de los procesos racionales que intervendrían en la deliberación.

De un lado, el conjunto motivacional subjetivo del agente no contiene sólo deseos:

> He discutido sobre *S* primero en términos de deseos, y esta expresión puede aplicarse, formalmente, para todos los elementos de *S*. Pero esta terminología podría conducirnos a olvidar que *S* puede incluir cosas tales como disposiciones de evaluación, patrones de reacción emotiva, lealtades personales, y varios proyectos, como podría llamárseles de forma abstracta, que incluyen compromisos del agente (*Ibid.*, p. 105).

De otro lado, el razonamiento medio-fin no es el único proceso racional por medio del cual obtenemos razones para actuar; se trata sólo de *un* caso. En rigor, el mero descubrimiento de que un curso de acción es un medio causal para un fin no sería, en sí mismo, una pieza de razonamiento práctico.

> Un claro ejemplo de razonamiento práctico es el que lleva a la conclusión de que uno tiene razones para Φ porque llevar a cabo Φ sería la manera más conveniente, económica, placentera, etc., de satisfacer algún elemento de *S*, y esto, por su-

6. En rigor, el modelo que la tradición reconoce como humeano no puede ser atribuido propiamente a Hume, por eso Williams lo llama "modelo subhumeano". La concepción de Hume depende enteramente de su teoría de las pasiones. Analizo el modelo de motivación de Hume en Velasco (2002).

> puesto, está controlado por otros elementos de *S*, [...] Pero hay posibilidades mucho más amplias de deliberación, por ejemplo: pensar cómo se puede combinar la satisfacción de los elementos de *S*, por organización temporal; o cuando existe algún conflicto irresoluble entre los elementos de *S*, considerar a cuál se le asigna más peso [...] o, de nuevo, encontrar soluciones constitutivas, como cuando se decide cómo se lograría pasar una tarde entretenida, suponiendo que uno quiere entretenerse (Williams, 1981, pp. 104-105).

Como resultado de estos procesos, el agente puede llegar a ver que tiene una razón para hacer algo para lo cual no veía que tenía una razón. De ese modo, el proceso deliberativo puede agregar nuevas acciones para las cuales hay razones internas, así como puede agregar nuevas razones internas para acciones dadas. La deliberación puede también sustraer elementos a *S* (el agente puede percibir que su creencia era falsa y entonces darse cuenta de que no tiene razones para hacer algo para lo cual creía que tenía razones para hacer); y –lo que es más importante– puede agregar elementos a *S* por medio del ejercicio de la *imaginación*: "... la imaginación puede crear nuevas posibilidades y nuevos deseos" (*ibid.*, p. 105). Podemos resumir, entonces, las actividades en las que consistiría, para Williams, la deliberación racional:

- hallar medios causales para los fines que uno está motivado a alcanzar;
- hallar realizaciones constitutivas para esos fines;
- armonizar los fines, viendo cómo pueden combinarse;
- jerarquizar los fines, cuando la armonización se prueba imposible, e
- imaginar completamente la realización de fines.

Las razones internas, entonces, son aquellas que surgen por medio de una deliberación correcta –una expresión muy usada por Williams es "*sound deliberative route*"– a partir de motivaciones que el agente ya tiene (Williams, 1994, p. 36). Esto significa que no todo elemento de *S* origina una razón interna. "A tiene una razón para Φ" significa más que "A está dispuesto a Φ". (Por eso las razones internas desempeñan un rol tan importante en los consejos de la forma "Si yo estuviera en tu lugar..."). Al decir que el agente tiene razones para hacer algo estamos autorizados a corregir sus creencias sobre hechos y su razonamiento. Como el agente es racional, pero puede tener creencias falsas, hay que admitir que el agente:

(a) puede creer que tiene una razón interna cuando en realidad no la tiene; y

(b) puede no saber un enunciado sobre una razón interna respecto de sí mismo.

Esto sería suficiente, según Williams, para que la noción de razón interna sea normativa. El ejemplo que Williams usa más de una vez es: un agente cree que el líquido que hay en una botella que tiene en frente de sí es ron, cuando en realidad es gasolina. Y quiere una cuba. ¿Tiene alguna razón para mezclar el líquido con refresco y bebérselo? Por una parte, parece natural decir que no tiene ninguna razón para beberse el líquido, aunque crea tenerla. Por otra parte, si se lo bebe, no sólo hay una razón para que lo haga sino que esto demuestra que, en relación con su falsa creencia, está actuando racionalmente, y por tanto, tenemos una explicación para que lo haya hecho.

El argumento concreto contra la posibilidad de las razones externas tiene dos partes. Dado que una razón práctica tiene dos rasgos: capacidad explicativa y conexión con la deliberación racional, Williams muestra, primero (a), que la presunta razón externa no podría explicar la acción; y segundo (b), que, aun suponiendo que pudiera explicarla, la presunta razón externa no sería una razón a la cual el agente podría llegar por medio de una deliberación racional.

(a) Como una razón externa es un enunciado que, por definición, podría ser verdadero independientemente de las motivaciones del agente; entonces una razón externa nunca podría explicar la acción de una persona.

> El punto principal de los enunciados sobre razones externas es que pueden ser verdaderos independientemente de las motivaciones del agente. Pero nada puede explicar las acciones (intencionales) de un agente excepto algo que lo motive a actuar así. De manera que se necesita algo más aparte de la verdad del enunciado sobre razones externas para explicar la acción, algún enlace psicológico; y tal enlace psicológico parece ser una creencia. El hecho de que A crea un enunciado sobre razones externas sobre sí mismo puede ayudar a explicar su acción (Williams, 1981, p. 107).

La premisa clave del argumento es la afirmación de que *nada puede explicar las acciones intencionales de una persona excepto algo que la motive a actuar de ese modo*. La verdad de un enunciado no puede explicar la acción de una persona. Para explicar la acción de una persona necesitamos citar "algo más", algún estado psicológico de esa persona que establezca el enlace con el enunciado: "Algo que la motive a actuar de ese modo". La acción intencional es imposible, o al menos inexplicable, en ausencia de algún elemento de *S* que la acción satisface. Presumiblemente, el estado psicológico tendría que ser una creencia: la creencia del agente en la verdad del enunciado. *Creer* que una consideración particular es una razón para actuar proporciona (en rigor, constituye) una razón para ac-

tuar. Pero si el agente la cree, entonces el agente es alguien respecto del cual se puede hacer un enunciado verdadero sobre razones internas: el agente tenía una motivación apropiada en su *S*[7].

(b) Como las razones prácticas tienen que revelar también la racionalidad del agente, la segunda parte del argumento quiere mostrar que, dada una razón externa, el agente no podría llegar por medio de una deliberación racional a tener esa razón. Para que pudiéramos decir que una razón externa es verdadera –o sea, que el agente tiene verdaderamente esa razón para actuar–, lo que habría que mostrar es cómo un agente *podría llegar a creer* el enunciado sobre razones externas acerca de sí mismo. Si el agente llegara a creerlo, ya sabemos, estaría motivado a actuar; la pregunta es cómo adquiriría esa nueva motivación que, por hipótesis, no surge a partir de ningún contenido de *S*. Ciertamente, Williams ha admitido que podemos llegar a nuevas motivaciones por un proceso de deliberación racional, pero sólo *a partir* de los contenidos que ya están en *S*. La deliberación puede agregar elementos a *S*, pero sólo a partir de elementos que ya están en él. (Esto implica que, para Williams, no todos los elementos de *S* pueden ser derivados de otros: algunos tienen que poder surgir "espontáneamente", sin un proceso racional previo). El caso de la razón externa es diferente, pues aquí –por hipótesis– se trata de una nueva motivación que no surge de ninguna motivación previa del agente, sino de un "proceso puramente racional"[8].

7. En este punto se podría criticar que, para Williams, las razones externas no existen porque colapsan en las internas. Tan pronto como una razón "externa" se vuelve capaz de motivar a un agente, el agente será alguien sobre el que se puede hacer un enunciado verdadero sobre razones internas, con lo cual la razón no era "externa" después de todo. *Cf.*, por ejemplo, Dancy, 1994-1995. No obstante, aunque en el momento en que el agente llegó a creer en la razón los dos tipos de razones son indistinguibles, antes de ese momento son diferentes. *Cf.* nota 8.
8. Cuando alguien tiene una razón interna, el único fundamento que hay para atribuirle esa razón es la existencia de un elemento relevante de *S*. Si es un elemento de *S* que no fue derivado de otros, sino que simplemente surgió –por ejemplo, un deseo–, entonces comenzó a existir en un momento, y antes de ese momento el agente no tenía una razón interna. Sin embargo, el caso de alguien que tiene una razón externa, y que más tarde llega a tener el elemento apropiado en su *S*, es diferente. Tiene que haber alguna razón para atribuir la razón a esa persona antes de que adquiera el elemento de *S*. La existencia del elemento en *S* no puede ser el fundamento para atribuirle la razón, porque se supone que la persona tenía la razón antes de que el elemento de *S* existiera. *Cf.* Cohon (1986).

Los que sostienen que puede haber razones externas entienden que si el agente deliberara racionalmente, sin importar las motivaciones que tuviera originalmente, llegaría a reconocer que tiene esa razón y, por tanto, estaría motivado a realizar la acción. El reproche que dirigen a quien no es sensible a la razón ofrecida es el de irracionalidad. Por eso, ante quien formula el enunciado sobre razones externas, no basta citar el "reconocimiento" de una razón por parte del agente, como un estado mental que pudiera haber surgido de cualquier modo en él: tiene que tratarse de un reconocimiento *racional.* No basta que el agente adquiera la motivación o llegue a creer en el enunciado de cualquier modo –tal vez porque fue persuadido por una retórica emotiva–. Quien formula el enunciado sobre razones externas supone que el agente adquiere la motivación *porque* llega a creer el enunciado sobre las razones, y que lo hace, además, porque está considerando el asunto correctamente. Por eso acusa de irracional a quien no cree en el enunciado:

> Por supuesto existen muchas cosas que podría decirle un hablante a alguien que no está dispuesto a Φ cuando el hablante piensa que debería estarlo, como que es desconsiderado, o cruel, o egoísta, o imprudente; o que las cosas, y él mismo, estarían mucho mejor si tuviera tal motivación. Cualquiera de estas cosas puede ser sensato decir. Pero quien da mucha importancia a plantear la crítica en forma de un enunciado sobre razones externas parece interesarse en expresar que lo que está mal particulamente con el agente es que es *irracional.* Es el teórico quien necesita particularmente precisar esta acusación: en especial, porque quiere que cualquier agente racional, como tal, reconozca que se require realizar la acción en cuestión (Williams, 1981, p. 110).

Quien formula un enunciado sobre razones externas supone, entonces, que la *única condición* para que la persona adquiera la nueva motivación es que delibere de la manera correcta. Dados los supuestos de Williams, obviamente, no se entiende cómo se podría satisfacer esa condición. Williams deja claro que la carga de la prueba queda para el "teórico de las razones externas". A él le cabría explicar esa "forma especial" de concebir la relación entre adquirir una motivación y llegar a creer un enunciado acerca de las razones. La palabra queda en el teórico de las razones externas. Mientras tanto, tendríamos que admitir que "la única racionalidad de la acción es la racionalidad de las razones internas" (*ibid.*, p. 111), o sea, aceptar que, de cualquier manera que caractericemos los estados del agente en que consiste su estar-motivado, las razones prácticas del agente serán siempre relativas a esos estados. Para Williams, los contenidos del conjunto motivacional subjetivo establecen las *condiciones* para

la presencia de razones para actuar, y eso significa admitir que todas las razones prácticas son hipotéticas.

II

¿Prueba Williams verdaderamente que ningún proceso racional puede hacer surgir un motivo para actuar, es decir, en sus propios términos, generar una razón "interna"? En rigor, no. Sólo lo prueba si aceptamos de antemano que las razones internas únicamente pueden surgir a partir de un estado "motivacional" que ya estaba en el agente, y sólo de acuerdo con alguno de los procedimientos que, para Williams, cuentan como racionales. La argumentación de Williams es circular. La afirmación de que las razones prácticas (o "internas") son relativas a los estados motivacionales del agente es el punto de partida de su argumento, de modo que no es sorprendente que sea también su conclusión. El argumento debe leerse, más bien, como un desafío al "teórico de las razones externas" para que explique esa "forma especial" de concebir la motivación que, bajo los presupuestos de Williams, aparece como inexplicable. Pero el argumento no prueba que sea imposible explicar la motivación de otra manera. Lo que Williams hace es plantear una dicotomía. Lo que debemos preguntarnos es si tenemos que aceptar esa dicotomía.

Williams supone, plausiblemente, que para explicar la acción intencional de una persona necesitamos apelar a "algo que motiva al agente", y que ese algo tiene que ser algún estado interno del agente –una causa mental, tal vez–. Dado que el término "deseo" se aplica formalmente a todos los elementos de *S*, lo que se supone es que los *deseos* son los estados mentales apropiados. La tesis genérica defendida por Williams podría formularse así: *no hay motivación sin deseos del agente que la acción satisface*. Sin embargo, la tesis humeana que quiere defender no se reduce a esta afirmación, que –depende de cómo se la interprete– puede ser incontrovertida. Después de todo, se puede admitir –como lo hace Nagel (1970)– que las explicaciones intencionales precisan mencionar deseos del agente y no admitir que los deseos funcionen como condiciones para la presencia de razones para actuar. O se puede admitir –desde una perspectiva kantiana– que toda deliberación tiene que tomar como punto de partida deseos presentes del agente sin que eso implique aceptar que esos deseos no admitan una revisión racional o un posicionamiento del agente frente a los mismos. El concepto de deseo necesario para la explicación intencional es puramente *formal*, y podemos entender que los deseos que precisamos atribuir al agente para hacer inteligible su acción se explican por las razones que el agente tiene, o sea, son deseos motivados por razones.

En rigor, Williams se compromete con una tesis más fuerte. Lo que supone es que la presencia de los deseos limita el alcance de las explicaciones racionales. Así, aunque admite que algunos deseos puedan ser explicados por razones[9], lo que supone es que *no todos* los deseos pueden serlo; o que los deseos sólo pueden ser explicados por razones que se explican a partir de otros deseos que, a su vez, no pueden ser explicados sólo por razones. Los deseos son entendidos como la *fuente* de la motivación en el sentido, más fuerte y *no cognitivo*, de que no pueden haber surgido de razones que no se basen, a su vez, en otros deseos. Lo que hace a la posición distintivamente humeana, entonces, sería la pretensión de que toda la deliberación práctica debe tomar deseos *presentes* como punto de partida, los cuales, a su vez, no puedan ser derivados racionalmente[10]. Como este proceso es, justamente, *la* deliberación racional, se entiende que los deseos primitivos no pueden haber sido producidos por deliberación y, por tanto, deben ser considerados como dados no-racionalmente. Éste es el corazón humeano del argumento, en la medida en que se apela a deseos *últimos* no susceptibles de justificación racional. El problema es que Williams no proporciona ningún argumento para defender esta tesis.

¿Qué decir, entonces, sobre la distinción entre razones internas y externas? Es central para la concepción de las razones internas que ellas puedan ser tanto razones explicativas como razones normativas (o justificativas). Con esto Williams establece una exigencia perfectamente plausible sobre las *razones para actuar*: la exigencia de que puedan motivar al agente. Las razones prácticas tienen que ser motivos y, por tanto, ser potencialmente explicativas de las acciones. Si las razones no son motivos, ellas no podrían desencadenar ni explicar acciones, y si las razo-

9. Obviamente, no todos los deseos pueden ser tomados como primitivos (porque eso significaría atribuir un conjunto infinito de disposiciones motivacionales al agente). El modelo tiene que proporcionar alguna explicación sobre cómo el agente adquiere nuevos deseos. La solución es afirmar que los nuevos deseos derivan sistemáticamente de los viejos. Los deseos se dividen, entonces, en dos clases: aquellos que surgen sin la intervención de ningún proceso racional, y los que derivan racionalmente de otros. En este punto puede haber varios mecanismos. En el modelo "subhumeano" los nuevos deseos se crean *vía* creencias. (Deseo azúcar, creo que el chocolate es dulce, entonces deseo chocolate). En el modelo "mejorado" de Williams, como vimos, es posible derivar un nuevo deseo variando imaginativamente o recombinando los viejos.

10. R. J. Wallace (1999) llama a esta tesis "*desire-in desire-out*" porque lo que se afirma es que cualquier proceso que produce un deseo como *output* tiene que tener otro deseo como *input*.

nes no son las que desencadenan ni explican acciones, no podemos decir que seamos racionales en sentido práctico. Pero decir todo esto es compatible con muchas otras posiciones que también se dicen "internalistas" y que no suscribirían la tesis humeana de Williams. De hecho, así es usada la distinción entre internalismo y externalismo en general, y la cuestión de la motivación moral en particular. En este contexto, se habla de una "exigencia internalista" que toda razón práctica debe satisfacer, en el sentido de ser capaz de motivar, de estar conectada con motivos[11]. El problema que veo en esta distinción es que ella acaba siendo banal, porque son muy pocos los que no serían internalistas. De hecho, hay una interpretación de la posición de Williams –que yo llamaría débil, por no tener otro nombre mejor– que entiende la tesis de las "razones internas" en este sentido débil. Según esta interpretación, las razones internas serían aquellas que pueden incorporar alguno de los motivos del agente de manera que lo llevan a realizar aquello para lo cual son razones. No es ésta, obviamente, la interpretación que yo presenté aquí.

En la interpretación que yo favorecí, las razones internas son aquellas cuyo *status* en tanto razones depende de su capacidad de incorporar los motivos del agente. La idea es que ellas no serían razones, si el agente no tuviera los motivos que tiene (*cf.* Velleman, 2000). Creo que esta es la tesis de Williams. Y creo que la distinción entre razones internas y externas, en este sentido "fuerte", plantea una falsa dicotomía. La tesis de Williams sólo representa un desafío para el "teórico de las razones externas" si éste comparte sus presupuestos humeanos, o sea, la idea de que las razones justificativas son "verdades inertes" y que las razones explicativas son "paquetes de deseos y creencias".

Para finalizar, un comentario sobre lo que creo que es el problema de fondo en la posición de Williams. A pesar de que Williams dice tener en cuenta la racionalidad del agente en su concepción de las razones internas, creo que no consigue dar cuenta plenamente del carácter normativo de la racionalidad. No se deja abierta la posibilidad de que el agente reconozca una razón como su razón, que la razón sea verdadera, y que, al realizar la acción, no la realice *por* esa razón. No se deja abierta la posibilidad de que la conducta sea irracional. Y eso tiene que ser admitido una vez introducida la categoría de la racionalidad. En este aspecto, comparto las críticas "racionalistas" de kantianos y realistas[12]. Creo

11. La aceptación de esta exigencia internalista suele ser el punto de consenso a partir del cual se generan las grandes divergencias; en ese sentido parece constituir un suelo común para la discusión actual entre humeanos, kantianos y aristotélicos. *Cf.* G. Cullity/B. Gaut (1997) (Introducción de los editores).

12. *Cf.* Korsgaard (1996); también McDowell (1995).

que toda tesis escéptica respecto de la motivación depende de la concepción que se tenga sobre la racionalidad, que toda teoría de la motivación depende de la visión que tenemos acerca de la racionalidad y que, por tanto, no puede ser usada como parte de un argumento independiente en favor de cualquier visión particular de la racionalidad. Que seamos o no humeanos respecto de la motivación depende de si somos o no humeanos respecto de la racionalidad.

Bibliografía

Cohon, Rachel (1986): "Are External Reasons Impossible?", en: *Ethics*, 96, pp. 545-556.

Cullity G., Gaut, B. (eds.) (1997): *Ethics and Practical Reason*, Oxford: Oxford University Press.

Dancy, Jonathan (1994-1995): "Why there is really no *such* thing as the theory of motivation?", en: *Proceeding of the Aristotelian Society*, New Series, Vol. 95, Papers, pp. 1-18.

Hume, David (1978): *Treatise of Human Nature* (Edición Selby-Bigge, Oxford: Clarendon Press), [Tratado de la naturaleza humana, F. Duque (trad.), Madrid: Ed. Nacional, 1977].

Korsgaard, Christine (1996): "Skepticism about Practical Reason", en: *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press, pp. 311-334.

Korsgaard, Christine (1997): "The Normativity of Instrumental Reason", en: Cullity, G., Gaut, B. (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press, pp. 213-254.

McDowell, John (1995): "Might There Be External Reasons?", en: Altham & Harrison, *World, Mind, and Ethics, Essays on the ethical philosophy of Bernard Williams*, Cambridge: Cambridge University Press.

McDowell, John (1978): "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society*, Supplementary Volume 52, pp. 13-29.

Nagel, Thomas (1970): *The Possibility of Altruism*, Princeton: Princeton University Press.

Velasco, Marina (2002): "Hume. As paixões e a Motivação", en: *Analytica*, Vol. 6, número 2, pp. 33-60.

Velleman, David (2000): "The Possibility of Practical Reason", en: *The Possibility of Practical Reason*, Oxford: Oxford University Press.

Wallace, R. J. (1999): "How to Argue about Practical Reason", en: *Mind*, 99 (395).

Williams, Bernard (1981): "Internal and External Reasons", en: *Moral Luck*, Cambridge: Cambridge University Press, pp. 101-113.

Williams, Bernard (1985): *Ethics and the Limits of Philosophy*, Harvard: Harvard University Press.

Williams, Bernard (1994): "Internal Reasons and the Obscurity of Blame", en: *Making Sense of Humanity,* Cambridge: Cambridge University Press.

LUIS EDUARDO HOYOS

Razones y motivos para actuar*

Profesor asociado, Departamento de Filosofía,
Universidad Nacional de Colombia.

1. BASTA CON ACEPTAR que no hay razones para actuar que sean independientes del deseo, para contar ya con la posibilidad de cumplir con un requisito de la explicación de la acción de acuerdo con un modelo no sustancialmente diferente al modelo causal, pues el deseo de algo, el de hacer algo, o también el deseo por alguien, se tendrían que referir siempre en tal caso a un sistema de motivaciones que opera en un contexto que se podría llamar sin reservas "natural". Pero entonces queda como un problema el modo como se han de incorporar las razones para actuar a un marco normativo, social y moral, al que no siempre se acomodan nuestros deseos. Bernard Williams ha sostenido que se puede defender la primera opinión sin por ello caer en el segundo problema. Según Williams, si se quieren explicar de un modo adecuado las razones para actuar, se debe aceptar que esas razones son "internas", es decir, razones que están ligadas a lo que él llama el "*conjunto motivacional subjetivo* del agente" ("the agent's *subjective motivational set*". Williams, 1981, p. 102). Para Williams, el enunciado "A tiene una razón para Φ" –en el que Φ representa una acción o un verbo de acción–, puede ser considerado como verdadero, si "A tiene alguna motivación que se verá satisfecha o favorecida por el hecho de que A haga Φ, y si no es así, el enunciado será falso". En otras palabras, no sería cierto que A tiene una razón para hacer Φ, si A no tuviera una motivación para hacerla. Ésta es la interpretación internalista de la racionalidad práctica. Según la interpretación externalista, el enunciado "A tiene una razón para hacer Φ" no se-

* Este trabajo es parte del proyecto de investigación "El concepto social de racionalidad y la idea de persona", inscrito en la DIB. Agradezco a Colciencias y al DAAD que me permitieron terminarlo gracias a una beca de investigación en Mainz y en Frankfurt durante el invierno 2002-2003. A Ignacio Ávila, Elke Brendel, Andreas Niederberger y Marina Velasco quisiera agradecer sus observaciones críticas y sus comentarios.

ría falso en el caso de que A no tuviera un motivo para hacer Φ, o en el caso en que hacer Φ no estuviera ligado al "conjunto motivacional subjetivo" de A. Aun cuando el "conjunto motivacional subjetivo" de un agente se halla directa o indirectamente en relación con los deseos, no consta él, para Williams, sólo de deseos, sino también de "disposiciones de evaluación, patrones de reacción emotiva, lealtades personales" y "compromisos del agente", etc. (*Ibid.*, p. 105). Me interesa en este ensayo someter a debate la distinción que hace Williams entre "razones internas y razones externas". Quisiera servirme de ese debate, y del contexto temático de la discusión generada por el muy comentado artículo de Williams en la filosofía contemporánea, para adelantar una breve exposición del modo como debe, en mi opinión, ser entendida la racionalidad práctica.

2. Williams elige la interpretación internalista de las razones para actuar por su mayor poder explicativo. Y esa es, sin duda, una motivación teórica bastante comprensible y aceptable. Pero, aunque pueda concederse que el internalismo –tal como Williams lo interpreta– está dotado de un importante potencial explicativo de la acción, o mejor, de las razones que nos llevan a actuar, no creo que él agote todo lo que se ha de comprender bajo el término "razón" (*reason, ratio*) práctica, o bajo el concepto de "racionalidad práctica". No todo lo que pueda decirse en términos explicativos de una acción y que está expresado por sus razones, comprendidas como motivos –que son, a su vez, un tipo de causas–, agota el carácter racional de la acción. Entre otras, porque si así fuera, me temo que todas las acciones podrían ser consideradas como racionales. Y sostener tal cosa es abiertamente no plausible. Muchos simples ejemplos de la vida diaria y de la historia muestran que las acciones humanas son con frecuencia irracionales. La idea básica del internalismo de elegir los motivos de la acción por su alto poder explicativo como indicadores de la racionalidad práctica no contribuye a esclarecer esta elemental intuición cotidiana. Junto a esta insinuación surge también la obvia pregunta de si es realmente aceptable que sólo haya razones para actuar que dependan directa o indirectamente de los deseos. Estas dos insinuaciones pueden ser tenidas como guía de la presente confrontación crítica con el internalismo.

3. Con frecuencia se ha insistido en que la interpretación internalista de la racionalidad práctica defendida por Williams representa un desafío para toda concepción racionalista sobre la razón práctica[1]. Se trata de un desafío al racionalismo práctico muy emparentado con el de Hume, aunque mejorado, ya que, si bien propone que lo que un agente reconoce como una razón para actuar tiene

1. *Cf.* Korsgaard (1986), McDowell (1998), Searle (2001) y Velasco (2004).

que surgir de algún elemento que ya se halla en su conjunto motivacional subjetivo y, por tanto, no tiene origen en un proceso racional, este conjunto motivacional subjetivo es algo más rico que un mero conjunto de deseos. En este orden de ideas, se piensa que la interpretación internalista de las razones para actuar no da cuenta del carácter normativo de la motivación y, por tanto, trae consigo implicaciones muy problemáticas a la hora de abordar el tema de la motivación moral. Con todo, esa objeción sólo me parece justificada en la medida en que se asuma que el carácter racional de una acción tiene algo que ver con su carácter moral. Y eso no es obvio sin más. Por cierto que Williams comparte esta presunción, pero no –me parece– del todo conscientemente. Esa circunstancia es relativamente fácil de observar cuando se tiene en cuenta que Williams –como he dicho– privilegia el internalismo por su potencial explicativo, pero no porque esté interesado en una consideración normativa. No obstante, sus preocupaciones normativas aparecen aquí y allá en sitios aislados y se vuelven explícitas y transparentes al final de su ensayo. Y entonces su concepción sucumbe a las mismas dificultades a las que está expuesta toda teoría moral de corte subjetivista: ¿Cómo se puede según ella garantizar que un juicio moral sea objetivo? Dicho de otro modo: ¿Cómo se puede mostrar mediante una teoría semejante que haya, por así decir, correspondencia entre motivación subjetiva y justificación moral? Sobre el punto volveré más adelante.

4. El poder explicativo del internalismo descansa en la idea de que la motivación de un agente para llevar a cabo una acción brinda la mejor posibilidad de comprender sus razones para actuar. Toda acción que pueda ser llamada racional debe poder ser entendida y explicada, y no hay mejor manera de comprender la acción que atendiendo a los motivos del agente. Aquí hay dos ideas implicadas, de algún valor para mi argumentación: la primera, que la comprensión y explicación de una acción son indicativos de racionalidad. Y esa explicación y comprensión no serían posibles, obviamente, sin referencia a razones por las cuales se actúa. La segunda, que las razones que llevan a actuar a un agente son motivos, esto es, razones internas, no externas. Esto quiere decir que no se puede considerar que un agente tenga razones para actuar si ellas no están ligadas en algún sentido a su conjunto motivacional subjetivo. No es necesario, al parecer, que surjan todas de él, pero sí tienen que ser de alguna manera remitidas a él. En otras palabras: un agente no puede afirmar con pretensiones de verdad que tiene una razón para hacer algo si no tiene un motivo para hacerlo.

Supongamos que alguien es presionado a hacer algo por obediencia a una costumbre o a una regla social, y que no quiere hacerlo. Por ejemplo, alguien tiene que casarse con una persona que no conoce ni quiere en virtud de un pac-

to dinástico. Esa persona no puede decir que tuvo una razón para actuar hasta tanto no incorpore, por así decir, la orden de casarse a su conjunto motivacional subjetivo, y entonces haga de un factor *externo* (la orden de casarse por razones de índole política y dinástica) un motivo *interno*.

"Nunca digas que no te casaste por amor" –le recomendó Isabel de Castilla ("la Católica") a su hija Juana (conocida como "la Loca") en vísperas de su unión obligatoria con Felipe el Hermoso. "Cásate con quien se te ordena para conveniencia del reino. El amor vendrá después" –insistió la madre. Y esa creencia, al parecer, convenció a la princesa[2].

No es que no haya razones externas, sino que éstas no tienen el poder explicativo de la racionalidad de la acción que tienen las razones internas, o sea, los motivos. Y si una razón externa puede ser presentada como una razón explicativa de una acción racional, debido a que, entre otras, activa un proceso de deliberación, es porque ella se ha hecho interna. Esto me parece –como dije– harto comprensible y aceptable, aunque no estoy seguro de que no sea trivial; de que no sea tan trivial como afirmar: los motivos son razones que explican las acciones y los motivos son internos. Luego las razones que explican las acciones (racionales) son internas.

5. Algo más difícil me parece entender qué es una razón externa y por qué ella no tiene el poder explicativo de la racionalidad de la acción que tiene la motivación, o qué poder tiene –si tiene alguno.

Al entrar a considerar lo que es una razón externa, no sólo se torna más difícil la explicación de las razones para actuar, sino que, además, lo que acabo de señalar como trivial (que la motivación explica la acción, las razones de la acción, y que no puede ser externa sino interna), parece ya no serlo tanto. Esta última es una presunción mía basada en la impresión de que a las características de *interno* para la motivación y *externo* para lo que Williams llama razones externas (pero que todavía no sabemos exactamente qué es) se añaden otros aspectos. Sospecho incluso que muchos de esos aspectos son más relevantes para el tema de la racionalidad práctica que los calificativos "interno" y "externo".

¿Qué es exactamente una razón externa? Éste es un asunto que, en mi opinión, dista de ser claro en el texto de Williams. Al principio, las razones exter-

2. Williams se sirve del ejemplo de Owen Wingrave, un personaje de una novela de Henry James, que es incitado por su familia para que –siguiendo la tradición familiar– se vuelva soldado a pesar de que él odia la vida militar (*Cf.* Williams, 1981, pp. 106 y ss.). Como se trata en todo esto de la racionalidad de la acción, me parece más próximo a nuestro tema el ejemplo de "Juana la Loca".

nas son caracterizadas negativamente, por mero contraste con las internas. Esto es, una razón externa es todo lo que no es una razón interna o una motivación; o también: todo lo que en principio no forma parte del conjunto motivacional subjetivo. Esta caracterización es relativamente clara, pero es subsidiaria de la de razón interna. Y por eso es que es relativamente clara, porque la idea de motivo como razón para actuar es comprensible.

Más adelante, Williams ofrece una caracterización algo más complicada, pero a mi modo de ver más interesante, de lo que es una razón externa. Williams imagina lo que él llama un "teórico de las razones externas" (Williams, 1981, p. 108). Este teórico defiende la existencia de razones explicativas del carácter racional de la acción diferentes a las internas, a las motivaciones del agente. Aun cuando se acepte que la motivación para actuar es la razón para actuar, para el teórico de las razones externas es necesario "que el agente adquiera la motivación porque llegó a creer el enunciado sobre las razones, y que lo haga, además, porque, de alguna manera, está considerando el asunto correctamente" (*ibid.*, pp. 108-109).

Piénsese otra vez en el ejemplo de Juana la Loca. Ella se casa con Felipe el Hermoso, teniendo una razón para ello, cuando considera como correcta la opinión de su madre acerca de la unión marital por conveniencia política. Según Williams, lo que para el teórico de las razones externas ocurrió aquí fue lo siguiente: las convicciones de la reina Isabel sobre el primado del poder sobre el amor, el relato de su propia experiencia, en fin, una serie de creencias en principio muy bien fundadas, activaron en Juana un proceso de deliberación racional. Si ese proceso de deliberación es genuinamente racional, entonces, "sin importar las motivaciones que tuviera originalmente" (*ibid.*, p. 109), Juana estaría motivada a casarse con Felipe en virtud de, o gracias a, esa deliberación. La motivación es resultado de la deliberación y no al contrario.

Lo que Williams sugiere es que situaciones como ésta pueden ser interpretadas de dos modos. Según el primer modo (la interpretación llamada internalista), la razón externa (los consejos y órdenes de la reina Isabel) sólo explica la acción, y las razones de la acción, cuando se vuelve interna, cuando se vuelve un motivo. Y una razón externa no se vuelve un motivo si no se relaciona con algún elemento del conjunto motivacional subjetivo del agente. De acuerdo con el segundo modo, en cambio (la interpretación externalista), se diría que el hecho de que la razón externa se vuelva un motivo debe ser entendido en el sentido de que ella *genera* un motivo, o que ella por sí misma motiva. La interpretación externalista sostiene que un proceso de deliberación racional puede llegar por sí mismo a engendrar una motivación sin que haya motivaciones detrás de ese

proceso, o sin que aparentemente él se encuentre en relación con el conjunto motivacional subjetivo.

A mí me parece que es muy difícil aceptar que un proceso de deliberación racional empiece por sí mismo a generar una acción sin que él esté soportado, por así decir, por una base disposicional para actuar. Por eso me parece más plausible la interpretación internalista, tal y como la presenta Williams. Ahí no veo mucho problema, siempre y cuando –claro está– esa interpretación internalista sea considerada como un punto de vista *explicativo teórico*, es más, quizás fundamentalmente *psicológico*, sobre las razones para actuar. Creo que los problemas empiezan a aparecer cuando Williams sugiere de un modo más o menos ambiguo el vínculo de la posición externalista con dos ideas sobre la racionalidad práctica de raigambre kantiana. La primera de esas ideas es la equivalencia kantiana entre el enunciado "existe una razón para que A..." y el enunciado "A debe..." (*ibid.*, p. 106). La segunda es la idea según la cual "la razón puede hacer surgir una motivación" (*ibid.*, p. 108).

Aunque Williams es cauteloso al emparentar el externalismo con el concepto kantiano del "debe" y con el imperativo categórico (de hecho no lo hace *expressis verbis*), no es muy difícil observar que, bajo sus supuestos, esa postura sería externalista, pues para Kant la acción "por deber" tiene que ser una acción concebida como independiente de nuestro "conjunto motivacional subjetivo", de nuestra base disposicional para actuar, subjetiva y empíricamente condicionada, y que podemos suponer conformada por preferencias, deseos, intereses, etc.[3]. La otra idea que Williams emparenta con el externalismo ("la razón puede hacer surgir una motivación") genera confusión. ¿Qué quiere decir ahí "la razón"? A mí me parece esa palabra muy difícil de manejar. Francamente cada día entiendo menos cómo han podido agitarla con tanta soltura tantos filósofos. Se entiende con relativa facilidad que haya disposiciones racionales, acciones racionales, opiniones racionales, deliberación racional, estrategias racionales, etc. Pero que haya algo así como "*la* razón" no me parece muy sencillo de pasar. No es que esto sea muy importante, pero toca en alguna medida el punto, porque si lo que dice Williams es que el externalismo se puede emparentar con la posición filosófica que cree que ese mito secular llamado La Razón (¿la pura o la impura?) puede hacer surgir una motiva-

3. No es para nada evidente, por supuesto, que la posición kantiana corresponda a una visión externalista. Christine Korsgaard ha insistido en el carácter internalista de la posición kantiana. Esta sugerencia fue aceptada más tarde por Williams (1994): si un deliberador racional acepta las constricciones de la moralidad, entonces puede decirse correctamente que una razón moral es una razón interna, ya que en tal caso ella sería parte de su conjunto motivacional subjetivo (Williams, 1994, pp. 37 y ss.).

ción, entonces es muy elemental desvirtuarlo. Bastaría para tal efecto la indicación de que semejante facultad, comprendida de un modo tan sustantivo, no existe[4]. Pero si lo que se desea decir es, más bien, que no se comprende cómo procesos de deliberación racional, de elección de estrategias racionales, etc., puedan estar liberados de una base motivacional del agente cuando él está buscando razones, entonces la crítica al externalismo adquiere más calidad. Lo cual no quiere decir que deba ser aceptada, por supuesto.

6. Desde un punto de vista consecuentemente internalista, la motivación no tiene, en estricto sentido, justificación racional. La motivación brinda una razón para actuar, pero ésta no surge, por así decir, del vacío, sino que se halla ligada a un conjunto motivacional subjetivo que opera como base disposicional. Ahora bien, a la base de esa base no hay ninguna otra base que cumpla con la función de ser la última y a la que –para usar el lenguaje lógico de las razones y las consecuencias, que es en buena medida el lenguaje de la justificación racional– corresponda una última razón o un último fundamento. El "conjunto motivacional subjetivo" de la acción, nuestra base disposicional para obrar, es como la famosa rosa de Angelus Silesius: ella "es sin un por qué", "florece porque florece".

Según el externalismo, en cambio, la motivación sí tiene una justificación racional, pues surge de un proceso de deliberación racional.

La mayor dificultad en la controversia entre estas dos perspectivas reside, en mi opinión, en que ella se suele asociar a la relación entre *moralidad* y *racionalidad*, y a las diferentes versiones que tenían Hume y Kant de esa relación.

Antes que nada se ha de llamar la atención sobre el hecho de que, para resolver el problema que surge de reconocer que hay una relación entre moralidad y racionalidad, no es necesario resolver el problema de si la motivación subjetiva tiene o no una justificación moral, o si ella se encuentra en alguna relación con la justificación moral. Ésta es una cuestión que desempeña un papel crucial para muchos autores esforzados en hallar alguna correlación entre motivación y justificación de cara al problema de una fundamentación de la moral[5]. Yo me temo, no obstante, que ese es un camino equivocado, pues ninguno de esos intentos logra impedir que se mezclen la psicología de la acción con la moralidad de la

4. Esta es la razón principal por la que la crítica de Ch. Korsgaard a Williams deja tanta insatisfacción: ella apela muy fuertemente a la presunta existencia de una facultad (la "razón pura"), entendida –en mi opinión– de un modo extremadamente sustantivo. *Cf.* Korsgaard (1968, pp. 21-23).

5. *Cf.*, por ejemplo, Nagel (1970), McDowell (1998) y Forst (1999).

acción de una manera que no se me hace muy prometedora. Y no muy prometedora es esa mezcla, en primer lugar, porque la perspectiva psicológica es ante todo una perspectiva explicativa (no la única, por lo demás) y no una normativa. Y, en segundo lugar, porque pienso que la diferencia entre una dimensión teórico-explicativa y una dimensión normativa es de tipo categorial, por así decir, esto es, se trata de una diferencia que tiene a la base dos distintas formas de hablar. En esa medida me parece que la búsqueda de una relación entre explicaciones psicológicas de la acción, basadas en motivos, y sus justificaciones morales y normativas no tiene futuro. No niego que haya motivos morales; niego que un motivo que es moral sea moral por ser un motivo.

Hume tiene razón al sostener que el asunto de la justificación racional de nuestros fines o razones para obrar nos lleva a un regreso. Tiene razón, ciertamente, siempre y cuando se entienda aquí que se está hablando de justificación en un sentido estrictamente teórico-racional. Ese regreso sólo se interrumpe, para Hume, cuando hallamos un elemento del cual no se puede dar razón porque no es un principio racional sino algo que anima, por así decir, nuestro deseo de vivir:

> Parece evidente que los fines últimos de los actos humanos no pueden en ningún caso explicarse por la *razón*, sino que se encomiendan enteramente a los sentimientos y afectos de la humanidad, sin dependencia alguna de las facultades intelectuales. Preguntad a un hombre *por qué hace ejercicio*, y os responderá que *porque desea conservar la salud*. Si le preguntáis entonces *por qué desea la salud*, inmediatamente os contestará que *porque la enfermedad es dolorosa*. Si lleváis vuestras inquisiciones más allá y deseáis que os dé una razón de *por qué odia el dolor*, es imposible que jamás pueda daros ninguna. Se trata de un fin último, y no puede ser referido a ningún otro objeto.
>
> Quizás a la pregunta segunda de *por qué desea la salud*, pueda también contestaros que *porque es necesaria para el ejercicio de su profesión*. Si vosotros le preguntáis *por qué está deseoso de hacer eso*, os responderá que *porque desea ganar dinero*. Y si le preguntáis: *¿Por qué?* Él os contestará: *Porque (el dinero) es el instrumento del placer*. Y más allá de esto, sería absurdo seguir pidiendo razones. Es imposible que aquí haya un proceso *in infinitum*, y que una cosa pueda ser siempre razón de por qué otra es deseada. *Tiene que haber algo que sea deseable por sí mismo, debido a su inmediata concordancia o acuerdo con el sentimiento y afecto humanos*[6].

6. D. Hume, *An Enquiry Concerning the Principles of Morals*. Ap. I. § 112. (La última cursiva es mía).

Por otra parte, y más esencial, creo que para juzgar acerca del valor moral de una acción es irrelevante decidir algo sobre su motivación. Salvo a alguien –pongamos por caso– de ser asesinado injustamente. Puede ser que lo salve por benevolencia, o por deber, o por temor a Dios, o por interés, debido a que esa persona tiene una deuda conmigo, o para ganar prestigio ante los demás, o por dinero, pues he decidido formar parte del millón de informantes del presidente Uribe[7], etc. Ninguno de esos motivos da valor moral a la acción. Lo que da valor moral a la acción es lo que estipulemos acerca de la justicia y la acción justa, acerca del bien y la acción buena. No quiero con ello decir –repito– que no hay motivaciones para obrar moralmente. Ciertamente las hay, y es mejor que las haya a que no las haya. También hay motivaciones para obrar inmoralmente, y es peor que las haya a que no las haya. Todo lo que sostengo es que la motivación de una acción no es lo que da su valor moral o, si se quiere, lo que le imprime un carácter normativo.

Kant fue más o menos consciente de esto. Cierto es que hay una insistencia excesiva en su filosofía moral por relacionar el valor moral de una acción con aquello que la *causa*. De ahí surgió ese engendro que él llamó acción "por deber". Pienso que esa insistencia de Kant en buscar un motivo de la acción que la dote de valor moral incondicional es sumamente problemática, incluso perniciosa. La teoría kantiana de la acción por deber como la única acción dotada de genuino valor moral, es una teoría demasiado idiosincrática.

Hay, en cambio, otras dos ideas de la filosofía moral kantiana de mucho mayor valor, en mi opinión, para comprender la relación entre moralidad y racionalidad práctica, y para contribuir a establecer una plausible caracterización de lo que es la racionalidad práctica.

---

7. Álvaro Uribe, el presidente de Colombia elegido en mayo de 2002 para el período constitucional 2002-2006, ha declarado que Colombia sólo puede ser salvada del violento caos en el que se halla actualmente sumida, si la *sociedad civil* participa activamente en el conflicto y colabora con las autoridades militares. En ese sentido, ha proclamado abiertamente a la opinión pública que el Estado requiere la ayuda de al menos un millón de informantes civiles para combatir tanto a la guerrilla de izquierda como a las milicias paramilitares. En eso consiste el programa del "millón de informantes" –muy publicitado al principio de su gobierno pero cada día que pasa, defendido más bien con prudencia. Lo más descabellado de la idea es que el gobierno ofrece recompensa a esos "colaboradores informales" por su cooperación. Se cae de su peso que una medida semejante podría llevar a una mayor desinstitucionalización del país y a una aún más grave devastación de su sistema jurídico y penal.

La primera de ellas se hace manifiesta en la preocupación fundamental de Kant por proponer un test *racional* para dictaminar el valor moral de una máxima, o sea, de aquello que, según él, orienta *subjetivamente* la acción. Ese test es el "imperativo categórico", o está posibilitado por él. La segunda, íntimamente ligada a la anterior, es la idea de que la racionalidad práctica no se agota en el esquema medio-fin, es decir, que la racionalidad práctica no es sólo racionalidad instrumental. Según Kant, el esquema medio-fin es un buen indicador de racionalidad práctica, de modo que puede decirse de alguien que actúa racionalmente si se conforma al criterio fundamental de la llamada racionalidad instrumental, a saber: quien pretende un determinado fin, debe querer los medios para alcanzar el fin. El niño que tiene dolor de pecho y desea aliviarse pero no se toma el jarabe, infortunadamente amargo, que conducirá a la obtención de ese fin, actúa por ello irracionalmente, ya que desea un fin pero no acepta los medios que –como se le ha explicado– lo llevarán a su consecución. No sólo es irracional, vale anotar, por no querer algo amargo, sino por no ser capaz de aceptar un rato de desagrado para recuperar la salud, es decir, por no ser capaz de sacrificarse temporalmente de acuerdo con lo que se suele llamar una "conducta estratégica"[8].

Pero el criterio de la racionalidad instrumental –así como la estructura general de lo que Kant llama "imperativos hipotéticos"– no estipula aún, ni puede por sí mismo estipular, que el fin pretendido sea racional. De ahí que sea forzoso introducir un criterio para decidir sobre la racionalidad de un fin. Y es aquí donde la relación entre racionalidad y moralidad se torna virulenta:

> Pues bien, todos los *imperativos* mandan o *hipotética* o *categóricamente*. Aquéllos representan la necesidad práctica de una acción posible como medio para llegar a otra cosa que se quiere (o es posible que se quiera). El imperativo categórico sería el que representase una acción como objetivamente necesaria por sí misma, *sin referencia a otro fin*. (La última cursiva es mía -LEH).
>
> [...] Ahora bien, si la acción fuese buena meramente como medio *para otra cosa*, el imperativo es *hipotético*; si es representada como buena en sí, y por tanto como necesaria en una voluntad conforme en sí a la razón, como principio de esa voluntad, entonces es *categórico*[9].

8. *Cf.* Jon Elster (1984, Cap. I).

9. I. Kant, *Grundlegung zur Metaphysik der Sitten*. Sección II, p. 414 [159]. Con respecto a aquello que constituye un criterio o principio para estatuir tanto la racionalidad como la moralidad de un fin es, por supuesto, también indispensable atender al concepto kantiano de "persona", pues es justamente este concepto el que hace comprensible el concepto de "fin en sí mismo". Eso significa que sólo aquel que se con-

Una de las principales contribuciones del pensamiento moral kantiano está en ahondar sobre la dificultad que supone decidir algo sobre la racionalidad práctica ya no solamente cuando se la considera como la adecuada coordinación de medios y fines, sino cuando se acepta que tiene que ver con la racionalidad de los fines. Es justo en este punto que se vuelve necesario contemplar un aspecto normativo para tener una idea más completa de la racionalidad práctica, pues si el fin pretendido de una acción, cuyas consecuencias para los otros son consideradas con significado para ellos, para su vida, para su salud, para su integridad, llegara a ser juzgado como incorrecto moralmente, no sería de esperar de la acción por medio de la cual se quiere alcanzar ese fin que sea moral, ni tampoco racional.

Si, por ejemplo, Carlos Castaño, el jefe supremo de las milicias paramilitares de Colombia, desea conservar el corredor de Urabá, en la frontera con Panamá, para fines estratégicos (pues por ahí se facilita el ingreso de armas a Colombia y la salida de cocaína, esencial para financiar la guerra), y un equipo de sus mejores asesores le indica que el mejor método para conservar esa zona consiste en acabar mediante el asesinato y el amedrentamiento con cualquier vestigio de apoyo a las FARC –la brutal guerrilla comunista colombiana–, pues no se puede menos que reconocer como racional que adopte esos medios si quiere aquel fin. El problema es si el fin mismo se puede considerar racional. O mejor, el problema es si un fin que no es moral se puede considerar racional, de modo que medios subordinados a él puedan ser llamados sin más racionales sólo porque cumplen con un criterio instrumental.

No estoy seguro de que la filosofía moral kantiana haya establecido el método definitivo para evaluar tanto la moralidad como la racionalidad de un fin. Pero decidir sobre eso no es absolutamente importante en este momento. Lo importante aquí es la sugerencia que se puede extraer del pensamiento moral kantiano en el sentido de que lo que llamamos racionalidad práctica no se agota en el concepto de la relación medio-fin, pues este punto de vista no da cuenta por sí mismo de un elemento esencial de nuestra racionalidad práctica: la normatividad. El internalista tiene este problema. No es que él no dé cuenta del

---

sidera como persona en un mundo de personas –las cuales, a su vez, son tratadas como tales– puede disponer de un principio del actuar que no es meramente instrumental ya que al concepto kantiano de persona le es propio que una persona "no" pueda ser usada "meramente como medio" para otra cosa, sino que deba ser considerada "siempre al mismo tiempo como fin" (*GMS*, p. 429 [188]). Un "*fin en sí mismo*" constituye, por su parte, aquello que Kant llama "un principio *objetivo* de la voluntad" (*op. cit.*, *loc. cit.*).

carácter normativo de la motivación, como sostiene Marina Velasco (2004), sino que su caracterización de la racionalidad práctica deja por fuera ese aspecto.

El aspecto normativo y moral de nuestra racionalidad práctica es esencial a ella porque está íntimamente ligado a la existencia de la sociedad y a la existencia del individuo dentro de la sociedad. Acción inmoral es acción irracional porque ella *puede* implicar que el individuo sea excluido de la sociedad y porque en muchos casos ella *puede* ocasionar la destrucción de la sociedad en cuanto contribuye en alguna medida a esa destrucción. Ahora bien, ningún ser racional que quiera permanecer vivo puede querer al mismo tiempo ser excluido de la sociedad o destruirla.

El internalista –tal como es presentado por Williams– tiene el gran problema de que no atiende suficientemente al hecho de que solamente se puede dar cuenta del aspecto normativo de una acción racional en cuanto el agente gana una perspectiva *objetiva*, esto quiere decir aquí, una perspectiva *intersubjetiva*. Gracias a esa perspectiva, él está en condiciones de comportarse de tal manera que su acción no rompa reglas constitutivas de la vida colectiva. Y esto va más allá de la motivación, pues ella se refiere ante todo a la perspectiva meramente subjetiva, meramente personal del agente. No es que el internalista no atienda al carácter normativo de la motivación, como suele decirse de Hume, sino que su caracterización de la racionalidad práctica y también de la moralidad de una acción no cuenta con ninguna garantía objetiva. No se entiende del todo, sin embargo, por qué esto tenga que ser así, pues juzgar moralmente y actuar moralmente tienen siempre que ver de algún modo con instancias objetivas[10].

No es del todo correcto, en mi opinión, afirmar que el "subjetivismo moral" de Hume no posee pretensiones *normativas*. También es injusta la interpretación de su filosofía moral que sostiene que ella es completamente *irracionalista*. Hume tiene, por supuesto, un interés de hallar un principio normativo de la acción: es buena aquella acción que me procura satisfacción y bienestar. Igualmente, es bueno aquello que deseo. También es uno de los intereses de Hume sacar a luz una estructura plausible de la racionalidad práctica. Él cree haber hallado esa estructura cuando hace explícito el único logro de la razón en la acción moral. La razón debe, primero, establecer si los objetos que causan determinados afectos (*passions*) realmente existen, es decir, debe establecer si las acciones se basan en suposiciones (*suppositions*) correctas. Y ella debe estar, en segundo lugar, en capacidad de ayudarnos en la evaluación de los medios para la consecución de determinados fines pretendidos; más exactamente, ella es esa

10. Sobre esta crítica a Hume, *véase* Foot (1978, pp. 74-80); también Nagel (1970).

capacidad de evaluar si los medios que emplazamos para la consecución de determinados fines son suficientes o no[11]. Por supuesto que hay para Hume racionalidad práctica, sólo que ella es básicamente instrumental. La razón es como una linterna que nos permite ver en la noche el camino indicado. No es ella, sin embargo, la que nos pone en movimiento por ese camino. El principio –la causa– del movimiento viene de otro lado.

El problema de la filosofía moral de Hume consiste, en primer término, en no atender suficientemente al *aspecto objetivo de la normatividad*, o mejor, en no atender suficientemente al hecho de que en toda evaluación –de carácter normativo– hay forzosamente implicada una referencia objetiva. No es suficiente decir que yo deseo y que el deseo es criterio de evaluación, esto es, que es bueno lo que deseo. Sino que, puesto que el *deseo* es siempre *deseo de algo*, ese algo ha de ejercer, como objeto público, alguna influencia sobre mi capacidad de desear. No puede excluirse que las opiniones –públicamente formadas y aprendidas a través de un medio público– ejerzan una influencia decisiva en la formación del deseo. Si lo que deseo es bueno porque lo deseo, o si, justamente, lo deseo porque es bueno, es algo sobre lo cual el subjetivista moral no puede decidir, salvo que crea que las opiniones, la formación cultural y otras instancias de índole pública, no están implicadas en el surgimiento del deseo. Ahora bien, si lo creyera, tendría una concepción harto incompleta del surgimiento del deseo. Acéptese, ciertamente, que el surgimiento del deseo de algo es más o menos oscuro, más o menos subjetivo. No obstante ello, una elemental reflexión cotidiana nos indica que no puede ser tan absolutamente oscuro y subjetivo que en su formación no intervengan prejuicios y opiniones sobre sus objetos. Y es improbable, a su vez, que estos prejucios y opiniones sean el resultado de consideraciones de personas aisladas, de personas no influidas por los sistemas públicos de valoración y de conformación de opinión.

Por otra parte, el aspecto objetivo implicado en toda evaluación debe ser atendido si se tienen pretensiones normativas –y Hume evidentemente las tiene–, pues si creo que es bueno sólo aquello que me hace feliz o sólo aquello que deseo en una determinada circunstancia, o en un estado particular, entonces no tengo forma de garantizar que eso podría ser dañino para otros, o para mí mismo, en otras circunstancias o en otro estado. En un momento de odio y resentimiento provocado por una acción injusta, por ejemplo, la venganza podría significar para mí reparación. No podría negarse que aquel resentimiento esté ligado a un sentimiento moral, a una capacidad de evaluación moral. Pero no creo que haya un solo sentimentalista,

11. *Cf.* D. Hume, *Treatise of Human Nature*, Book II. Part III. Sec. III, pp. 415-416.

o algún subjetivista moral, que de ahí extraiga la conclusión de que la venganza puede estar justificada moralmente para fines de reparación. La evaluación moral de la venganza va más allá de las necesidades subjetivas de reparación, y es a eso a lo que se ha de atender cuando se piensa en una instancia objetiva forzosamente implicada en toda consideración normativa o evaluativa.

Hay una segunda característica de la filosofía moral de Hume que, en relación con la racionalidad de la acción, la hace problemática. Se trata de que él sólo vea como pertinente racionalmente a lo que en nuestra acción se puede enmarcar en el modelo medios-fines. Sólo considerar ese aspecto de la racionalidad práctica es sencillamente problemático porque el esquema medios-fines aporta un criterio para aplicar el predicado "racional" a una acción en el caso exclusivo en el que previamente se ha elegido un fin, pero no aporta, en cambio, criterio alguno para aplicar el mismo predicado a la elección del fin. La racionalidad instrumental no sirve por sí misma para decidir cuándo un fin es racional, salvo que se considere que *todo* fin es, a su vez, medio, es decir, que sólo hay fines intermedios y no fines que se pretendan por sí mismos. No creo, sin embargo, posible desconocer que haya fines que se pretendan por sí mismos, pero tampoco considero necesario incurrir en una metafísica de los fines, en una mistificación de ciertos fines como fines absolutos, sólo por hacer ese reconocimiento. Ahora bien, si estamos en condiciones de decir algo sobre la racionalidad de los fines es porque creemos que se ha de tratar con sumo cuidado la relación entre ser racional y ser moral. Pues ni aun concediendo que no haya forma de establecer fines por sí mismos, es aceptable que no haya mejores fines que otros, y no tan sólo como fines intermedios para otros fines. Ahora bien, si hay unos fines mejores que otros, pero no en el sentido de ser fines intermedios para otros fines, es porque el criterio de evaluación no es instrumental sino moral.

6. Siguiendo esta dirección quisiera mostrar brevemente el modo como nos deja insatisfechos el internalismo de Williams cada vez que queremos dar cuenta del carácter normativo de las razones internas para actuar, es decir, de los motivos.

Si se atiende con cuidado a las respuestas que da Williams al final de su ensayo a tres preguntas que él plantea con la intención de poner en concordancia su posición internalista con una evaluación negativa –un reproche– de la actitud del así llamado *free rider* con respecto a los bienes públicos, se podrá ver por qué una teoría de la racionalidad práctica basada solamente en la motivación subjetiva no resuelve el problema de la normatividad de las razones para actuar.

Empiezo por la tercera pregunta:

"¿Podemos definir una noción de racionalidad donde la acción racional para A de ninguna manera se refiera a las motivaciones existentes de A?" (Williams,

1981, p. 112). La respuesta de Williams es, por supuesto: "No". No es de esperar algo distinto, pues ésta es, en esencia, la tesis del internalismo.

La primera y segunda preguntas son:

"1. ¿Podemos definir nociones de racionalidad que no sean puramente egoístas?" "2. ¿Podemos definir nociones de racionalidad que no sean puramente medios-fin?" (*op. cit., loc. cit.*). Y a esas dos preguntas Williams responde: "Sí". Es decir, desde el punto de vista que sostiene que la mejor explicación de la racionalidad de la acción es la internalista se pueden determinar nociones de la racionalidad que no sean puramente egoístas y que no se reduzcan meramente al esquema medio-fin. No niego que esto se pueda hacer y que el internalista lo pueda hacer. Creo simplemente que si el internalista lo puede hacer, no es porque sea un internalista, o sea, un filósofo de la acción que piensa que la mejor explicación de las razones para actuar es la que se refiere a los motivos del agente. No es por decir que las razones para actuar son los motivos para actuar que puedo decir que las acciones racionales no pueden ser puramente egoístas o que la racionalidad práctica no se agota en el esquema medio-fin. Hay que agregar a la motivación un criterio normativo que muy poco tiene que ver con ella como tal, es decir, que no es "interno".

Para poder llamar racional a una acción, hay que tener claridad sobre el hecho de que ésta debe desplegarse en un medio social institucionalizado y reglamentado. Supongo que ese hecho basta para traer consigo la forzosa aceptación de razones para actuar "independientes del deseo"[12]. Por eso la noción de racionalidad práctica excluye el egoísmo a ultranza, y por eso mismo no basta con decir que una acción es racional sencillamente por haberse conformado al esquema medio-fin, sin haber aún decidido sobre el carácter (social) racional de los fines.

Aunque saber acerca del motivo de la acción nos ofrece sin duda una buena explicación –ante todo psicológica, repito– sobre las razones del agente, y aunque, así mismo, no pueda negarse que la relación en la que se halla la motivación con los deseos del agente tiene que ser tenida en cuenta en toda comprensión sensata y realista de la acción[13], también es cierto que ni aquella explicación ni esta comprensión bastan para decidir sobre la moralidad de una acción y sobre la relación de esa moralidad con la racionalidad de la acción. Una acción motivada por la

12. Sobre el tema, *cf.* John Searle (2002).

13. *Cf.* Anscombe (1957) y, sobre todo, lo que en su explicación de las razones para actuar ella llama "desirability-characterisation" que es inherente a todo acto intencional (§§ 36-38).

benevolencia o por amor al prójimo puede ser ciertamente vista como buena y justa. Pero lo bueno y lo justo de esa acción no dependen del motivo. Una buena motivación no es buena por ser una motivación. La estipulación "bueno" o "justo" es normativa, no psicológica, vale decir, tiene lugar cuando se considera el ámbito social e institucionalizado de la acción, y sus repercusiones para los otros. Y esa es una cuestión eminentemente moral. La pregunta aquí es si este componente social-moral de la acción intencional tiene algo que ver con su racionalidad. Yo respondería que sí y destacaría, además, que la relación entre ambos componentes es esencial. La normatividad social es, desde un punto de vista especial, *constitutiva*, y no solamente *regulativa* (u opcional), de la racionalidad de una acción.

7. Para lograr una comprensión adecuada de la racionalidad práctica es necesario, desde mi punto de vista, tener en cuenta dos cosas: el motivo explicativo de la acción y la instancia normativa.

El motivo de una acción puede ser tenido como indicador de su racionalidad debido a su potencial explicativo y puede por ello ingresar en el "espacio de las justificaciones" (*space of reasons*) –para utilizar la conocida fórmula de Wilfrid Sellars, o también la otra complementaria de Robert Brandom: el "espacio social de las razones" o justificaciones–. Aunque las justificaciones, las razones, tengan un fin, por supuesto. El tema de la motivación como indicador de racionalidad de una acción nos lleva a tener en cuenta el papel que desempeña la explicación causal en toda explicación racional. Tanto las explicaciones *causales* como las *motivacionales* pueden ser tenidas como respuestas teórico-racionales a la pregunta: "¿Por qué?", es decir, a la pregunta que hace todo ser humano racional cuando quiere saber la explicación de un hecho o también de una acción. Sin embargo, la identificación de *motivo* y *causa* no puede ser aceptada sin reservas, cuando lo que está en juego es la explicación de las *razones* para actuar. Se puede ciertamente mostrar que una acción tenida por racional siempre tiene una causa tras de sí por la cual ella fue realizada, pues de lo contrario no podría ser considerada como un suceso en el mundo real. Y las acciones son, evidentemente, sucesos en el mundo real[14]. La explicación causal de una acción no es, sin embargo, la explicación que da cuenta de la acción *como acción*. Lo que es relevante en la explicación de una acción como tal es ante todo la razón o el motivo que el agente tiene para procurarse algo a través de ella. Y procurarse, o también proponerse, algo significa en cierto sentido desearlo.

Aun cuando los motivos, tomados como razones explicativas de la acción, representan indicadores de racionalidad más adecuados que las meras causas, es de

14. *Cf.* D. Davidson (1980), especialmente los *Essays* 1, 11 y 12.

observar que los motivos sólo cumplen con ser una significativa función en el contexto explicativo en cuanto son tratados retrospectivamente, esto es, en cuanto que ellos explican por qué esto o aquello *fue* hecho. Por eso la explicación motivacional de la acción no constituye todo lo que se ha de decir sobre su racionalidad. También es de relevancia lo que se diga de esa acción desde una perspectiva orientada hacia el futuro. Para esa determinación –orientada al futuro– de la racionalidad de la acción es indispensable considerar las razones para actuar en una dimensión normativa y social. En esa medida, una acción puede ser tenida por racional si ella resulta ser no solamente una acción explicable, sino si ella se puede al mismo tiempo introducir en un marco social e institucional, de tal modo que ella sea la manifestación de un comportamiento previsible que, a su vez, puede desencadenar otros comportamientos sociales previsibles.

El ejemplo de "conducir auto" explica bien mi punto de vista. Conducir auto no puede ser considerado únicamente como la acción de manejar bien el propio auto, esto es, de conocer todos los mecanismos para echar a andar el vehículo propio, sino que debe ser, además, comprendido como un procedimiento de correcta adaptación a un medio social e institucional. Esa adaptación no excluye la posibilidad de estrategias imaginativas y creativas para alcanzar una determinada meta, pero sí supone, esencial y constitutivamente, la obediencia mínima a ciertas normas inviolables. Cuando conducimos por una calle o por una autopista nos necesitamos los unos a los otros y necesitamos que cada uno esté atento y esté en condiciones de responder por lo que hace. Dos cosas son, entonces, indispensables cuando conducimos por una calle o por una autopista: la *conciencia de esa mutua necesidad*, y la conciencia de que sólo se puede conducir por una autopista o una calle en estado de *atención y de responsabilidad*. Quien no esté dispuesto a estar plenamente alerta y responsable cuando conduce por una calle o una autopista públicas, no debería conducir a través de un medio público. Pero, ¿qué significa conducir a través de un medio que no es público? Eso, en estricto sentido, contradice la noción de conducir. Es como si le dijéramos a alguien que no está dispuesto a respetar las normas mínimas del lenguaje que hable solo. Hablar solo es, en cierto y muy fundamental sentido, un contrasentido. Visto así, el ejemplo de "conducir un auto" debe dar suficiente ilustración del elemento *normativo* y *constitutivo* de la racionalidad práctica de un agente.

Si se quisiera hacer valer una perspectiva según la cual no cuentan solamente las explicaciones motivacionales y causales como indicios de racionalidad, entonces se debe exigir algo distinto y algo más de lo que se exige cuando se piden explicaciones retrospectivas de la acción. Junto a la condición de explicatividad –que es, ciertamente, condición necesaria, pero no suficiente– se debe exigir de esa acción que se

encuentre inmersa en un marco institucionalizado y reglado. Ahora bien, un marco institucionalizado y reglado, pero también justo, para el juego social contribuye esencialmente a que la sociedad como tal se mantenga. Esa condición adicional nos permite comprender por qué acciones que indican un comportamiento previsible y transparente también son acciones que *pueden* contribuir a la conservación y el mejoramiento de la vida en común. De acuerdo con eso se juzgará si la acción es buena o mala, correcta o incorrecta. Semejante visión *prospectiva* u orientada al futuro nos permite adscribir responsabilidad al agente. Es más, esta visión prospectiva exige que el agente sea responsable por cuanto en ella se presupone que el agente sea una persona con la cual se pueda contar, o al menos en cierta medida se pueda confiar. Ahora bien, para poder decir de una acción que ella es el resultado de una persona responsable, imputable y digna de confianza es irrelevante la referencia a los motivos subjetivos de esa acción –aunque siempre los haya–. Por eso puede decirse que el internalista, ciertamente, cumple con una importante condición para la comprensión de la racionalidad práctica, a saber: la condición de que una acción racional tenga que poder ser explicada retrospectivamente. No obstante, su propuesta no alcanza para comprender por completo lo que significa actuar racionalmente.

Bibliografía

Anscombe, G. E. M. (1957): *Intention*, Oxford: Basil Blackwell.

Davidson, Donald (1980): *Essays on Actions and Events*, Oxford: Clarendon Press.

Elster, Jon (1984): *Ulyses and the Sirens. Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.

Foot, Philipa (1978): "Hume on Moral Judgement", en: Foot, Philipa: *Virtues and Vices and Other Essays in Moral Philosophy*, Oxford: Basil Blackwell, pp. 74-80.

Forst, Rainer (1999): "Praktische Vernunft und rechtfertigende Gründe. Zur Begründung der Moral", en: Stefan Gosepath (ed.), *Motive, Gründe, Zwecke. Theorien praktischer Rationalität*, Frankfurt a. Main: Fischer, pp. 168-205.

Hume, David (1989): *Treatise of Human Nature*, L. A. Selby-Bigge (ed.), Oxford: Clarendon Press.

Hume, David (1957): *An Enquiry Concerning the Principles of Morals*, L. A. Selby-Bigge (ed.), *Enquiries Concerning the Human Understanding and Concerning the Principles of Moral by David Hume*. Oxford: Clarendon Press.

Kant, Immanuel (1989): *Grundlegung zur Metaphysik der Sitten*, en: Wilhelm Weischedel (ed.), *Werksausgabe*, Band VII, Frankfurt: Suhrkamp.

Korsgaard, Christine M. (1986): "Skepticism About Practical Reason", en: *The Journal of Philosophy*, Vol. 83, No. 1 (5-25).

McDowell, John (1998): "Might There Be External Reasons?", en: McDowell, John, *Mind, Value and Reality*, Cambridge (Mass.)-London: Harvard University Press, pp. 95-111.

Nagel, Thomas (1970): *The Possibility of Altruism*, Princeton-New Jersey: Princeton University Press.

Searle, John (2002): *Rationality in Action*, Cambridge, Mass.- London: The MIT Press.

Velasco, Marina (2004): "Razones internas *vs.* razones externas. Reflexiones sobre una distinción", en este volumen, pp. 181-193.

Williams, Bernard (1981): "Internal and External Reasons", en: Williams, Bernard, *Moral Luck*, Cambridge: Cambridge University Press, pp.101-113.

Williams, Bernard (1994): "Internal Reasons and the Obscurity of Blame", en: *Making Sense of Humanity*, Cambridge: Cambridge University Press.

ROBERTO HORÁCIO
DE SÁ PEREIRA

La paradoja de la irracionalidad según Donald Davidson*

Universidade Federal de Rio de Janeiro/CNPq.

1. Introducción

Desde el período clásico, la tradición filosófica rechaza por contradictoria la descripción de acciones, convicciones, inferencias y deseos simultáneamente intencionales (esto es, basados en razones), y, sin embargo, contrarios al juicio de valor del propio agente. En principio, solamente podríamos describir estados mentales de forma al mismo tiempo intencional e irracional en el sentido prudencial, o sea, según juicios de valor o evaluaciones que nosotros –terceros– realizamos, pero jamás a la luz de lo que el propio agente juzga mejor hacer, juzga mejor creer e inferir, juzga mejor desear, etc.

La discusión filosófica en torno a tal paradoja de la irracionalidad es conocida en la historia de la filosofía bajo el rótulo del problema de la acrasia (incontinencia) o de la debilidad de la voluntad. De este modo, Sócrates, en el *Protágoras* de Platón, rechaza como absurda la visión del sentido común de que alguien podría actuar de forma contraria a su conocimiento del bien, cediendo a tentaciones o placeres inmediatos, o coartado por apetitos e inclinaciones. Nadie podría actuar "dominado" por deseos irreflexivos (como apetitos e inclinaciones) y al mismo tiempo de forma intencional, o sea, basado en su conocimiento del bien. Aunque Aristóteles y Santo Tomás busquen mostrar, contra Sócrates, la posibilidad de acciones incontinentes, sus explicaciones terminan por privar a la incontinencia de su carácter intencional: los incontinentes se dejarían llevar por deseos irreflexivos (como causas puramente físicas) por no practicar, en el momento de la acción, su conocimiento del bien.

* Traducción del portugués de Tomás Andrés Barrero Guzmán, becario CNPq, Brasilia/Brasil (maestría).

A través de una relativización de los juicios de valor que, como veremos, atribuye a los mismos una forma semántica similar a la forma semántica de los juicios de probabilidad, Davidson cree poder mostrar en la filosofía contemporánea de la acción –de cierto modo, en contra de toda la tradición– que no hay paradoja alguna o contradicción cuando describimos acciones, convicciones y deseos como siendo al mismo tiempo intencionales e irracionales (naturalmente, según lo que el propio agente juzga mejor hacer, creer y desear).

En este trabajo me propongo mostrar que tal intento fracasa. No me parece posible afirmar que el agente, actuando en contra de las razones que él mismo reconoce en su juicio como las mejores, pueda, aun así, estar actuando intencionalmente, esto es, con base en el *reconocimiento* de las razones involucradas. El error de Davidson proviene de su concepción insuficiente de lo que sería una "razón primaria". Para caracterizar *G* como una acción primaria de por qué alguien llevó a cabo una acción *A*, no basta que afirmemos que *G* se compone (i) de una "actitud favorable", por parte del agente ante la acción *A*, con una determinada propiedad y (ii) de la convicción, por parte del mismo agente, de que *A* posee efectivamente tal propiedad. Es necesario agregar la suposición (iii) de que el agente llevó a cabo la acción *A* por *el reconocimiento* de la actitud favorable y de la convicción expresadas por *G*. Según mi diagnóstico, la condición expresada por (iii) es normalmente pasada por alto porque, en ausencia de razones en conflicto, la satisfacción de las condiciones (i) y (ii) nos lleva a suponer tácitamente que también la condición (iii) habría sido satisfecha. Incluso así, (iii) me parece constituir condición indispensable para que podamos hablar de un actuar, creer o desear intencionales.

2. Descripción pre-analítica del fenómeno

La palabra "acrasia" en el griego clásico significa no-ejercicio voluntario de la capacidad de autocontrol frente a deseos irreflexivos. Se la traduce por la expresión latina "*incontinentia*" y, de forma menos precisa, por la expresión compuesta "debilidad de la voluntad". Por incontinencia entendemos pre-analíticamente formas de conducta que contrarían el juicio del propio agente sobre la mejor (o más correcta) alternativa de acción, en la medida en que aquel deja de ejercer su capacidad de autocontrol frente a deseos contrarios a sus juicios sobre lo mejor. En esta acepción pre-analítica, la incontinencia constituiría una forma híbrida de conducta, por decirlo así, entre lo que entendemos usualmente por comportamiento imprudente y lo que entendemos por comportamiento compulsivo. Como el imprudente, pero a diferencia del compulsivo, el incontinente actuaría de forma libre e intencional. Tendría razones para lo que hace y se creería

libre para actuar o no de forma diferente. Pero, del mismo modo que el compulsivo y a diferencia del imprudente, actuaría de forma contraria a su propio juicio sobre la mejor (o más correcta) alternativa de acción, dejando de ejercer voluntariamente su autocontrol o resistencia frente a sus deseos.

Supongamos, para dar un ejemplo, que lo mejor (o más correcto) para alguien, que ya ha bebido lo suficiente, sea dejar de beber en razón de las obligaciones y compromisos del día siguiente. Sin embargo –sigue el ejemplo– tal individuo continúa bebiendo. Pre-analíticamente podemos describir su conducta (de continuar bebiendo) de tres formas diferentes. (1) En primer lugar, podemos imaginar que tal persona sigue bebiendo (aunque lo mejor sería parar) debido a auto-indulgencia o imprudencia. Tiene plena conciencia de lo que está haciendo y escoge libre e intencionalmente continuar bebiendo de forma imprudente o auto-indulgente. Es importante resaltar que de acuerdo con esta posible descripción de la conducta, *somos nosotros* los que, al suponer que la mejor opción en su estado sería dejar de beber, censuramos al agente, por irracional, su conducta de continuar bebiendo. El imprudente o auto-indulgente continúa bebiendo porque él mismo juzga que lo mejor que puede hacer en su situación es continuar bebiendo.

(2) De acuerdo con una segunda descripción posible de la conducta en cuestión, podemos imaginar que tal persona continúa bebiendo, aunque juzgue que lo mejor que podría hacer en su situación sería parar inmediatamente. En este caso, aunque el agente pueda tener plena conciencia de lo que está haciendo, no tendría alternativa real de acción, ya que no podría dejar de beber aunque quisiera: su deseo de continuar bebiendo sobrepasaría su juicio de que lo mejor que podría hacer sería dejar de beber. En los términos de esta descripción, la capacidad de resistencia o de auto-control, si existe en alguna medida en el agente, sería a todas luces insuficiente para vencer el deseo irreflexivo de continuar bebiendo. Los conceptos psicológicos tradicionales de fobia, manía y adicción parecen presuponer la noción de compulsión, por lo menos en sus formas límites o extremas. Probablemente este concepto tradicional de compulsión surge como una extensión analógica de la idea de una coerción externa. Actuando de forma compulsiva estaríamos siendo coartados por nuestros deseos o apetitos irreflexivos a actuar en contra de nuestra propia voluntad reflexiva, del mismo modo que podemos ser coartados por terceros a hacer lo que no deseamos.

(3) La acción de continuar bebiendo puede ser descrita, por último, como una forma intermediaria de conducta entre la imprudencia y la compulsión. En esta situación posible, la persona no sería compelida a continuar bebiendo. Permanecería bebiendo de forma libre, teniendo la alternativa de dejar de beber, ya que tiene, a diferencia del compulsivo, capacidad de auto-control o resistencia.

Además, su conducta sería intencional, en el sentido de que el agente tendría razones (aunque no las mejores) para continuar bebiendo o para no ejercer su capacidad de auto-control. Sin embargo, del mismo modo que el compulsivo, el incontinente estaría contrariando con su conducta aquello que considera como lo mejor (o más correcto) que puede hacer en la situación en cuestión, a saber, dejar de beber. La intencionalidad y libertad de conducta nos permitirían, entonces, afirmar que el agente continuaría bebiendo, no porque su deseo de hacerlo fuera irresistible, como en el caso del compulsivo, sino simplemente porque no ejercería su capacidad de auto-control sobre su deseo de beber. Haría una concesión al deseo contraria a su propia valoración, permitiéndose, así, continuar entregado a la bebida.

En estos términos pre-analíticos, la incontinencia parece constituir un fenómeno tan frecuente que nos preguntamos cómo la gran mayoría de filósofos puede cuestionar o negar su existencia. La creencia usual en la existencia de actos incontinentes parece apoyarse en dos órdenes de razones. La primera razón es oriunda de un empleo usual del predicado "irracional" que tiene el objetivo de señalar una incoherencia en el propio sistema de creencias, intenciones y acciones de un agente. En este sentido, caracterizamos justamente como irracionales aquellas acciones, creencias y deseos contrarios al propio juicio del agente sobre lo que sería mejor hacer, creer y desear. A la luz del empleo de tal predicado, la conducta incontinente parecer ser irreducible tanto a la compulsión como a la imprudencia. No podría ser reducida a la mera compulsión simplemente porque no tiene sentido que caractericemos como irracional una forma de conducta en la cual el agente no tiene alternativa de elección. Pero tampoco podría ser reducida a la imprudencia porque, en este caso, empleamos el predicado "irracional" para señalar nuestra discordancia con respecto al juicio y la intención del propio agente sobre el bien o lo mejor (por ejemplo, la intención de escalar el Everest) y no una incoherencia o inconsistencia interna al sistema de creencias, intenciones y acciones del propio agente.

El segundo orden de consideraciones que parece apoyar la creencia en la existencia de actos incontinentes tiene que ver con los elementos involucrados en cada forma específica de conducta. Mientras que la compulsión parece caracterizarse por un conflicto de orden exclusivamente emocional, y la imprudencia por una ausencia total de conflicto, la incontinencia parece caracterizarse fundamentalmente por un conflicto consciente, al mismo tiempo decisorio (sobre qué hacer, a partir de dos órdenes de razones opuestas) y emocional. Mientras que el compulsivo es coartado involuntariamente por un deseo irresistible y el imprudente actúa según su propio juicio (en el cual el agente escoge fomentar

su propio vicio), el incontinente se vería enfrentado siempre con dos raciocinios excluyentes. Por un lado, juzgaría que, a la luz de todas las evidencias disponibles, lo mejor que podría hacer en la situación en la que se halla es, por ejemplo, dejar de beber; sin embargo, por otro lado, juzgaría que es bueno o agradable continuar bebiendo de forma ininterrumpida. Sin ejercer su capacidad de resistencia y dejándose llevar por lo bueno en detrimento de lo mejor, se avergonzaría por su propia conducta y sería objeto del desdén ajeno.

3. Formulación general del problema de la incontinencia

Parece, entonces, que la existencia de formas de conducta que serían al mismo tiempo, intencionales, libres, pero contrarias al propio juicio del agente sobre lo mejor que puede hacer, parece cuestionar principios fundamentales de la teoría de la acción. Según estos principios, juicios o evaluaciones se reflejan en deseos o intenciones de acción, y deseos e intenciones se reflejan en acciones intencionales. Por consiguiente, formas de conducta híbridas entre la imprudencia y la compulsión no parecen conceptualmente posibles. O bien una conducta se deja caracterizar como libre e intencional, reflejando las intenciones del agente –y éstas, a su vez, sus juicios– (y si lo censuramos no nos queda más que calificarlo como irracional en la acepción de imprudente), o bien tal comportamiento, contrariando juicios e intenciones del propio agente, está desprovisto de intencionalidad y no podemos censurarlo (no tendría sentido), sino que tendríamos que calificarlo como compulsivo.

Creo que la mejor manera de señalar las dificultades conceptuales relacionadas con la suposición de la existencia de actos incontinentes consiste en la formulación de un problema filosófico. En un sentido estricto (que se remonta a los *Tópicos* 104b de Aristóteles), un *problema* de naturaleza conceptual reside en una contradicción (real o aparente) entre dos o más proposiciones que consideramos igualmente verdaderas. En esta acepción técnica, podemos afirmar que el concepto de incontinencia expresa un *problema,* ya que buscamos comprender una determinada acción, atribuyendo simultáneamente a su agente (i) un juicio contrario a la realización de aquélla, (ii) libertad para actuar o no de la forma en cuestión, acompañada (iii) de la intención real de realizar una forma distinta de acción en vez de la acción en cuestión. Esquemáticamente, podemos expresar tal problema conceptual en la forma del conflicto entre la siguiente secuencia de proposiciones:

(1) El agente es libre de hacer x o y.

(2) A la luz de todas las evidencias disponibles, el agente juzga que la acción x es mejor (o más correcta) que la acción y.

(3) El juicio del agente con respecto a que la acción x es mejor (o más correcta) que la acción y, a la luz de todas las evidencias disponibles, se refleja en una intención real del agente de no realizar y en lugar de x.
(4) No obstante, el agente lleva a cabo intencionalmente (con base en razones) la acción y en lugar de la acción x.

La contradicción (real o aparente) resultante de la suposición de estas cuatro proposiciones es usualmente denominada en la literatura contemporánea de "*last ditch*" (*cf.*, Pears, 1982). Un agente que sucumbe ante tal forma de incontinencia estaría realizando una acción y (en términos de la proposición 4) de forma libre o voluntaria (en términos de la proposición 1), de forma ponderada e intencional (en términos de la proposición 3), de forma contraria, sin embargo, a su propio juicio (valorativo o normativo), según el cual, actuar de forma x en vez de y sería mejor (o más correcto) (en términos de la proposición 2). Antes de continuar, vale la pena hilar algunas consideraciones preliminares sobre cada una de las proposiciones en conflicto.

La primera proposición le asigna al agente la capacidad de llevar a cabo alternativas de acción excluyentes (x o y). En estos términos, el concepto de libertad presupuesto por la formulación general del problema sería aquel según el cual ser libre significa (o implica) poder escoger una entre varias alternativas disponibles de acción. Para aquellos filósofos que piensan que libertad no significa, ni supone, alternativas disponibles de acción, el problema de la acrasia ni siquiera se dejaría formular. En la formulación general del problema, el enunciado de tal proposición pretende apenas distinguir las posibles acciones incontinentes de las formas involuntarias usuales de comportamiento, en las cuales el agente se ve compelido a actuar por causas ajenas a su voluntad, sean éstas externas o internas, buscando hacer justicia, de este modo, a la descripción pre-analítica original en términos de la cual el incontinente no ejerce libremente su capacidad de autocontrol.

La segunda proposición le asigna al agente un juicio de valor (x es mejor que y) o un juicio normativo (x es más correcto que y). Naturalmente, la forma lógica de tales juicios es discutible y, como veremos, parte de la solución formulada por autores contemporáneos consiste justamente en hacer explícita la forma correcta de tales juicios. Sin embargo, la formulación del problema es independiente de consideraciones sobre el contenido de tales juicios, así como de una teoría sustantiva acerca de normas y valores. En la formulación general del problema, el enunciado de esta segunda proposición pretende apenas distinguir las posibles acciones incontinentes de las conocidas formas de conducta imprudente, resaltando que las incontinentes estarían contrariando una valoración, no de terceros, sino del propio agente.

La tercera proposición exige que la valoración o juicio (x es mejor o más correcto que y a la luz de todas las evidencias disponibles) que se atribuye al agente se refleje en una intención real correspondiente (desear hacer x en lugar de hacer y). Tal proposición no figura en las caracterizaciones más usuales del problema de la incontinencia, simplemente porque se parte del *internalismo motivacional* como un principio general evidente de la teoría de la acción intencional. Según tal principio, las evaluaciones siempre se reflejan en motivos reales de acción, y éstos en acciones concretas. Dado que podemos imaginar, sin cuestionar el internalismo, excepciones a la regla en las cuales las evaluaciones del agente, por una razón u otra, no estarían reflejándose en motivos reales, el enunciado de esta tercera proposición pretende apenas incorporar a la formulación general del problema la restricción mencionada. Así, las posibles acciones incontinentes serían aquellas en las cuales el agente contraría un juicio que expresa sus intenciones reales de acción.

La cuarta y última proposición le asigna a la acción un carácter intencional. En términos generales, afirmar que una acción es intencional significa decir que se basa en razones. Hay, naturalmente, gran controversia sobre la forma que deben asumir tales razones. En el modelo clásico aristotélico, tales razones asumirán la forma de premisas mayores y menores de un raciocinio práctico según el patrón deductivo de la silogística. En la filosofía contemporánea de la acción, en contraste, numerosos autores consideran tales razones bajo la forma de premisas de raciocinios que seguirían un patrón inductivo. Con todo, la formulación general del problema de la incontinencia es independiente de tales consideraciones teóricas.

Por último, el enunciado de la cuarta proposición pretende apenas distinguir las posibles acciones incontinentes de aquellas formas irreflexivas de comportamiento desprovistas de intencionalidad, caracterizando la incontinencia como forma de conducta resultante del convencimiento, por parte del agente, de determinadas razones.

Para hacer visible nuestro problema basta con observar que la suposición de cualquier conjunto de tres de las cuatro proposiciones, cuando éstas son comprendidas correctamente, implica necesariamente la falsedad de la cuarta proposición sobrante. Así, si suponemos (1) que el agente es libre para actuar de forma x o y, (2) que considera que el determinado curso de acción x es mejor (o más correcto) que y, a la luz de todas las evidencias disponibles, además, (3) que tiene la intención real de llevar a cabo no llevar a cabo y en vez de x, entonces (4) tiene que ser falsa, es decir, no podemos admitir que el agente actúe intencionalmente de forma y. Si suponemos, en contraste, (4) que el agente actúa intencionalmente

de forma y, (3) que tiene la intención real de realizar x en lugar de y, pero, además, (2) que considera que la concreción de x sería mejor (o más correcta) que la realización de y, a la luz de todas las evidencias disponibles, entonces (1) tiene que ser falsa, es decir, tenemos que suponer que el agente no es libre de escoger y llevar a cabo x o y. Por último, suponiendo, (1) que el agente es libre de actuar de forma x o y, (2) que el agente considera x como mejor (o más correcto) que y a la luz de todas las evidencias disponibles, (4) aun así, llevando a cabo, de forma intencional, la acción y en vez de la acción x, la conclusión que se impone es que (3) es falsa, es decir, tenemos que suponer que la valoración del agente no se reflejó en intenciones reales del agente de llevar a cabo la acción x en lugar de la acción y.

Formulado en los términos de este conflicto, el problema de la debilidad de la voluntad deberá, entonces, ser, o bien *resuelto*, o bien *disuelto*. *Resolución* significa aquí encontrar alguna formulación para los principales conceptos involucrados (libertad, objetivo, juicio de valor y raciocinio práctico), de modo que las cuatro proposiciones en conflicto puedan ser compatibles. Una resolución del problema implica el reconocimiento tácito de la existencia de acciones incontinentes. *Disolución* significa, en contraste, el rechazo de una o más proposiciones en conflicto, de modo que las tres proposiciones sobrantes sean compatibles. Una disolución del problema implica, por tanto, una actitud escéptica frente a la posibilidad de la existencia de acciones incontinentes. En vez de actos incontinentes existirían apenas, o bien a) acciones compulsivas involuntarias, o bien b) acciones irreflexivas (sin intencionalidad), o bien c) acciones imprudentes (en plena conformidad con el juicio del propio agente), o bien, por último, d) acciones contrarias al juicio o a la valoración del agente, pero sin contrariar sus intenciones de acción.

4. La relativización del juicio de valor propuesta por Davidson

Como observamos, el incontinente parece caracterizarse por un conflicto en la decisión de qué hacer a partir de dos órdenes de razones contrarias. Si, actuando de forma intencional, el agente sería movido por razones, es decir, por deseos y juicios de valor que expresan una actitud favorable a la acción que realiza, al actuar de forma incontinente estaría contrariando su propio juicio sobre lo mejor. En la tradición filosófica, este conflicto de razones asume invariablemente la forma del conflicto entre deseos sensibles inmediatos, por un lado, y deseos morales o prudenciales, por otro. Incontinente sería aquel que, al no ejercer su capacidad de autocontrol, estaría cediendo a sus deseos sensibles inmediatos en detrimento de los deseos superiores de la moral o de la prudencia, dictados por la razón. Así,

en toda acción incontinente, apetitos o deseos sensibles de orden inferior estarían triunfando sobre deseos reflexivos de orden superior, o a la inversa, nuestras intenciones morales más nobles estarían siendo vencidas por deseos egoístas.

Contra esta visión tradicional del problema, Davidson nos ofrece algunos ejemplos interesantes de posibles casos de incontinencia donde ni el apetito o deseo sensible inmediato sería la fuerza victoriosa, ni tampoco la moral o prudencia sería la fuerza vencida. Supongamos que me encuentro relajado en la cama después de un día de trabajo extenuante, cuando se me ocurre que no me cepillé los dientes. La preocupación por la salud de mis dientes exige que me levante y cepille mis dientes, mientras que la complacencia me sugiere olvidarme del cepillado por hoy. Pondero las alternativas a la luz de las razones: por un lado mis dientes son fuertes y a mi edad el deterioro todavía es lento. No sería tan fundamental cepillarlos hoy. Por otro lado, si me levanto, arruino mi tranquilidad, lo que puede acarrear una noche en vela. Ponderando entonces todos los aspectos relevantes, juzgo que lo mejor sería permanecer en la cama. Sin embargo, el imperativo de cepillarme los dientes se impone a mi voluntad. Fatigado, me levanto de la cama y me cepillo los dientes. En este y en muchos otros casos, tendríamos una posible acción incontinente en la cual el deseo sensible no triunfa sobre el deber, siendo, por el contrario, vencido por un imperativo.

Según Davidson, el problema de la debilidad de la voluntad surge cuando buscamos hacer compatibles dos principios fundamentales de la teoría de la acción con el reconocimiento de la existencia de acciones incontinentes. El primer principio expresaría una conexión natural entre intención y acción, entre querer algo y actuar en conformidad: preferencias expresadas por intenciones de acción (prefiero llevar a cabo x a llevar a cabo y) estarían siempre reflejadas en la elección concreta de acciones (realizo x en lugar de y). Numerosos filósofos acogieron este principio en sus diferentes teorías sobre la acción intencional. Para Anscombe, por ejemplo, el signo primitivo de querer algo sería justamente buscar alcanzarlo. De forma similar, Hare afirma que hay una *relación lógica* entre querer algo y actuar de modo en que se logre aquello que se quiere, y Hampshire plantea que "A desea actuar de forma x" sería equivalente a afirmar que " si todas las circunstancias permanecen inalteradas, A haría x, si A pudiera". Interpretando la cláusula *ceteris paribus* de Hampshire ("si todas las circunstancias permanecen inalteradas") de modo que signifique "desde que no haya nada que el agente desee más", Davidson presenta la siguiente formulación para el primer principio fundamental:

> P1. Si un agente desea más realizar x que y, y si se cree libre para hacer x o y, entonces actuará intencionalmente de forma x, si actúa de forma x o y (Davidson, 1980, p. 23).

El segundo principio fundamental viene a expresar una relación esencial entre valoración y motivación, enunciando una versión del llamado internalismo motivacional: las evaluaciones siempre se reflejan en motivos, deseos y voliciones. Davidson lo formula en los siguientes términos:

> P2. Si un agente juzga que la realización de x sería mejor (o más correcta) que la realización de y, entonces prefiere realizar x a realizar y (*Ibid.*).

Según Davidson, ambos principios tendrían, según sus propias palabras, "un aire de auto-evidencia". Con todo, como intenté señalar en la formulación general del problema, aunque Davidson tenga razón con respecto a la existencia de una conexión conceptual entre valoración y motivación, siempre es posible imaginarnos casos de evaluaciones que no se reflejan en motivaciones. De una forma general, evaluaciones y juicios tienden a no reflejarse en deseos y voliciones cuando valoramos desde un punto de vista impersonal o a la distancia experiencias que para el agente representan una imposición de orden social o moral no plenamente interiorizada o avalada. Así, si incorporamos esta restricción a la formulación original de Davidson, bajo la forma de una cláusula adicional, en lugar de P2, tendremos entonces el siguiente principio general P2':

> P2'. Si un agente juzga que sería mejor hacer x que hacer y, *y si su juicio es auténtico, expresando un valor plenamente interiorizado por el agente*, entonces éste prefiere hacer x a hacer y (*Ibid.*).

Así, se configura una vez más nuestro problema de la debilidad de la voluntad. Si suponemos, de acuerdo con el segundo principio P2', que un agente juzga que la alternativa de acción x sería mejor o superior a la alternativa y, *y, además, si tal juicio es auténtico*, entonces el agente preferirá la alternativa x en vez de la y, en la medida en que preferencias expresadas por juicios de valor siempre se reflejarían en preferencias expresadas en intención de acción. Si ahora, de acuerdo con el primer principio P1, el agente prefiere la alternativa x en vez de la y, y si cree que es libre de realizar x, entonces optará intencionalmente por x, en la medida en que sus preferencias volitivas se verían siempre reflejadas en la elección concreta de formas de acción. En estos términos, los dos principios P2' y P1, tomados conjuntamente, excluyen, justamente, la posibilidad de que una acción pueda ser descrita al mismo tiempo como intencional e incontinente. Esta

última posibilidad es contemplada por Davidson bajo la forma de un tercer y último principio general:

P3. Existen acciones intencionales e incontinentes.

Davidson también toma distancia respecto de la tradición en la medida en que cree en la verdad de este tercer principio que enuncia la existencia de acciones incontinentes. Así, en vez de *disolver* el problema, rechazando alguno de los elementos de la tríada en conflicto, el autor pretende, más bien, compatibilizar los dos principios fundamentales con el principio que reconoce la existencia de acciones incontinentes. Su tesis general puede ser resumida en los siguientes términos: la alegada contradicción entre los tres principios referidos sería apenas aparente y sólo tendría lugar debido a una incomprensión con respecto a la forma general de los juicios de valor y, por consiguiente, con respecto a la estructura de los raciocinios prácticos en los cuales tales juicios figuran como premisas.

El punto de partida del autor es la explicación ofrecida por Aristóteles, y retomada por Santo Tomás de Aquino, quien, en razón de la supuesta intencionalidad de la acción acrática, buscaba reconstituir el raciocinio práctico que estaría en la base del comportamiento y del estado mental del incontinente. Aristóteles buscaba explicar la supuesta acción incontinente de comer un dulce en los siguientes términos. En primer lugar, se atribuye al incontinente la opinión universal (normativa) de que los dulces deben ser evitados, así como la opinión contraria (valorativa), pero igualmente universal, de que todo lo que es dulce es agradable. En segundo lugar, se atribuye al incontinente el juicio singular (juicio de percepción) de que eso (que él tiene enfrente) es dulce. Este juicio serviría como premisa menor para ambos principios universales (para realzar la contradicción entre las respectivas conclusiones de los posibles raciocinios, podríamos interpretar las premisas mayores y las propias conclusiones bajo la forma de juicios de valor –o normativos– comparativos, suponiendo que la relación "mejor que" o "más correcto que" sea una relación asimétrica. A la luz del primer raciocinio, deberíamos concluir que es mejor –o más correcto– evitar este dulce que comerlo, mientras que a la luz del segundo, que es mejor –o más correcto– comer este dulce que evitarlo). Como el incontinente actuaría de forma contraria al primer juicio universal, pero en conformidad con el segundo, Aristóteles sugiere, por último, que un deseo irresistible de comer el dulce ofrecido dominaría al agente, imposibilitando así momentáneamente que él ejerciera su conocimiento expresado en el primer juicio universal, a través de la subsunción de la premisa menor a la regla expresada por tal juicio universal.

La dificultad de la solución aristotélica radica en la supresión del conflicto consciente vivido por el incontinente al decidir entre dos posibles acciones, basado en dos órdenes contrarias de razones. Buscando hacer justicia a tal dilema o conflicto de naturaleza moral, el primer paso de Davidson consiste en el examen de la forma de un posible raciocinio práctico, que emerge como fusión de los dos raciocinios prácticos paralelos atribuidos por Aristóteles al incontinente. En este posible raciocinio práctico, los juicios de valor en conflicto se fundirían en una misma premisa mayor, y los juicios de percepción en una misma premisa menor. Así, la premisa mayor de este raciocinio (M3) enunciaría simultáneamente (p) "es siempre mejor (o siempre más correcto) que evitemos los dulces en vez de comérnoslos", (p') "es siempre bueno que comamos dulces" (los dulces son siempre agradables). La premisa menor (m3) diría, en contrapartida, (m) "es mejor que evitemos comer este dulce que nos lo comamos" y (m') "es bueno que comamos este dulce" (este dulce es agradable). Como el incontinente actúa contrariamente a su juicio sobre la mejor alternativa de acción, parece razonable suponer que la conclusión de tal raciocinio enunciaría (C3) "es mejor que evitemos comer este dulce que nos lo comamos". Sin embargo, como los juicios de valor que figuran en la premisa mayor del supuesto raciocinio no pueden ser verdaderos al mismo tiempo de la misma acción, (C3) no puede seguirse lógicamente de las premisas mayor (M3) y menor (m3), lo que significa decir que, en estos términos, para Davidson no tendríamos ni siquiera un argumento.

Una primera sugerencia, en el sentido de eliminar la contradicción entre los juicios de valor en conflicto, sería introducir la expresión modal *prima facie* como un operador proposicional para los juicios en cuestión: SI algo es un dulce, entonces *prima facie* debe ser evitado, SI algo es un dulce, entonces *prima facie* es bueno o agradable. Eliminaríamos la contradicción en la medida en que podríamos decir de una acción que ella *prima facie* debe ser evitada, siendo, sin embargo, *prima facie* buena. Continuando con la sugerencia, buscaríamos ahora extraer el juicio de valor *sans phrase* de la conclusión C3 (es mejor *sans phrase* que evitemos comer este dulce) de los juicios *prima facie* expresados por las conclusiones parciales (es bueno *prima facie* comer este dulce) y (no es bueno *prima facie* comer este dulce). Aun así, la dificultad inicial de comprensión de la naturaleza del raciocinio práctico perdura, en la medida en que no tenemos simplemente cómo extraer conclusiones acerca de lo bueno (o debido) *sans phrase* a partir de consideraciones sobre lo bueno (o debido) *prima facie*.

Pretender extraer conclusiones sobre el bien (o el deber) *sans phrase* a partir de juicios de valor *prima facie* sería un error que, según Davidson, tendría su origen en la suposición tácita (presente en toda la tradición filosófica) de que

los juicios morales (valorativos o normativos) tendrían siempre la forma de condicionales universalizados: (x) (Mx → Ix): cualquier acción x, si es una mentira es incorrecta; (x) (Px → Rx): cualquier acción x, si es placentera, debe ser realizada. Bajo tal suposición, no hay nada que podamos hacer con el operador modal *prima facie,* en el sentido de hacer comprensible la posibilidad de dilemas morales o conflictos en la decisión, eliminando una contradicción entre juicios de valor en conflicto. Según Davidson, este error decisivo en la comprensión de la naturaleza del raciocinio práctico sería análogo a aquel observado por Hempel en la inferencia probabilística, al pretender extraer conclusiones modales de premisas caracterizadas por operadores modales. Con gran claridad, este autor habría mostrado que no podemos inferir (C) "casi seguramente lloverá" de las premisas que enuncian (M) "si el barómetro disminuye, casi con seguridad llueve" y (m) "El barómetro está disminuyendo". Pues a cada inferencia modal que formuláramos en estos términos, podríamos siempre contraponerle simultáneamente otra inferencia paralela, con la misma forma lógica, cuya conclusión, sin embargo, sería contraria: (M') "cuando el cielo está rojizo en la noche, casi con seguridad no llueve". (m') "El cielo está rojizo esta noche". (C') "Casi con seguridad no lloverá".

El error decisivo en las inferencias probabilísticas residiría en una comprensión equivocada de la expresión "casi con seguridad". En lugar de expresar un operador modal de proposiciones singulares, cuya función sería modificar el consecuente del condicional de la premisa M: "Si el barómetro disminuye, *casi con seguridad* llueve", la expresión "casi con seguridad" tendría como función modificar la conexión lógica entre las frases: "el barómetro disminuye" y "lloverá". De este modo, la aserción original ("Si el barómetro disminuye, casi con seguridad llueve") debe ser comprendida como enunciando simplemente: "(el hecho de) que el barómetro disminuya, hace *casi seguro* que llueva" o, haciendo uso de la expresión "probable", "que el barómetro disminuya, *hace probable* que llueva", en símbolos: *pr* (Llx, Cx), donde los valores de la variable "x" serían localizaciones espacio-temporales para ser caracterizadas tanto por la disminución en la medida del barómetro (C) , como por la presencia de lluvia (Ll).

Esa analogía entre inferencias probabilísticas e inferencias prácticas le permite a Davidson afirmar que, tal como la expresión "casi con seguridad", la expresión *prima facie* no constituye un operador de predicados o proposiciones aisladas, sino, más bien, un conectivo proposicional. En vez de exhibir la forma de condicionales universalizados, juicios de valor, tales como "mentir es *prima facie* errado", deben ser entendidos en los siguientes términos: "(la suposición) que una acción sea una mentira la convierte *prima facie* en errada", o que "la

caracterización de una acción como una mentira es una razón para considerarla como *prima facie* mala", en símbolos: *pf*(Ex, Mx), donde los valores de la variable "x" serían acciones. Lo mismo sería válido para juicios comparativos: al afirmar que la acción de evitar comer dulces es mejor que la acción de comerlos, decimos simplemente que "el hecho de que una acción se caracterice como la acción de evitar comer dulces y otra como la acción de comer dulces, hace a la primera mejor que a la segunda", en símbolos: *pf*(x es mejor que y, x es el acto de evitar comer el dulce y y es el acto de comer el dulce), donde, una vez más, x y y son variables para acciones. En estos términos, consideraciones morales (valorativas o normativas) poseerían un estatuto *condicional* o *relativo*. No decimos de una determinada acción que es buena o mala, mejor o peor que otra, y *punto*. Afirmamos, más bien, que determinadas características de una acción constituyen razones para considerarla *prima facie* buena o mala, o que determinadas características de una acción x y determinadas características de una acción diferente y nos permiten considerar x como *prima facie* mejor que y.

Con base en esta reflexión previa sobre la forma lógica de los juicios de valor, podemos ahora formalizar los raciocinios prácticos que caracterizarían el dilema acrático. En el ejemplo aristotélico, el incontinente se vería enfrentado a dos juicios de valor ("ES bueno evitar comer dulces" y "ES bueno [placentero] comer dulces"), que asumirían la forma de las dos premisas mayores de los dos silogismos prácticos potenciales. Si consideramos el prefijo "pf" como símbolo del conectivo *prima facie* y las letras *a* y *b* como nombres de las acciones de evitar comer dulces y comer dulces, respectivamente, tendríamos como premisa mayor del primer raciocinio la siguiente: (M) *pf*(x es mejor que y, x es el acto de evitar comer el dulce y y es el acto de comer el dulce). En contrapartida, tendríamos como premisa mayor del raciocinio opuesto la siguiente: (M') *pf*(y es mejor que x, y es el acto de comer el dulce y x es el acto de evitar comer el dulce). La premisa menor común a ambos raciocinios asumiría la forma de los siguientes juicios de observación: (m) *a* es el acto de evitar comer, mientras *b* es el acto de comer. Como conclusión del primer raciocinio tendríamos: (C): *pf*[*a* es mejor que *b*, (M) y (m)], y del segundo raciocinio: (C'): *pf*[*b* es mejor que *a*, (M') y (m)]. En buen romance, la conclusión (C) diría que la suposición (M) (evitar comer dulces es mejor que comerlos), junto con la suposición (m) de que las acciones *a* y *b* se caracterizan, respectivamente, como la acción de evitar comer este dulce y la acción de comer este dulce, hacen a la acción *a prima facie* mejor que a la acción *b*. La conclusión (C') diría, por el contrario, que la suposición (M') de que comer dulces es mejor que evitar comerlos, y la suposición (m) de que las acciones *a* y *b* se caracterizan, respectivamente, como la acción

de evitar comer este dulce y la acción de comer este dulce, hacen a la acción *b prima facie* mejor que a la acción *a*.

Algunas observaciones se hacen relevantes aquí. En primer lugar, si Davidson está en lo cierto con respecto al estatuto condicional de las evaluaciones, ni las premisas mayores (M) y (M'), ni las conclusiones parciales (C) y (C') se contradicen o son lógicamente incompatibles. Nada impide que las consideraciones (M) y (m) hagan a *a prima facie* mejor que *b*, mientras que las consideraciones (M') y (m) hagan a *b prima facie* mejor que *a*. Considerando, sin embargo, que la acción realizada *b* (de comer el presente dulce) se caracteriza como incontinente, contrariando (por definición) el juicio del agente sobre la mejor alternativa de acción, debemos suponer que la primera conclusión (C) habría suplantado a la segunda (C'). Eso nos permitiría decir que el agente, ponderando las dos conclusiones parciales (C) y (C'), habría llegado a la siguiente conclusión definitiva (Cd): *pf* (*a* es mejor que *b*, *e*), donde *e* debe ser entendido como el conjunto de todas las razones conocidas, en la situación (M), (M') y (m). En suma, la consideración de todas las razones conocidas [que son: (M) los dulces son nocivos para la salud, (M') los dulces son agradables y (m) aquí hay un dulce] hacen a la acción *a* (evitar comer este dulce) *prima facie* mejor que a la acción *b* (comerlo).

Eso le permite a Davidson afirmar, en segundo lugar y en oposición a lo que suponía Aristóteles y toda la tradición, que el raciocinio práctico no sigue el patrón de la inferencia deductiva (propio de la silogística), sino, más bien, el patrón de las inferencias inductivas, siempre sujetas a invalidación. De esta forma, cuando lo que importa es saber cuál sería la mejor alternativa de acción disponible, a la luz de todas las razones conocidas, la razón práctica no se ve en mejor situación que los raciocinios inductivos sobre la previsión del tiempo. En uno y otro caso, no tenemos a disposición una fórmula general para calcular hasta dónde una afirmación sobre una conjunción de razones o enunciados de indicios puede ser inferida de la conjunción de estas razones o indicios tomados aisladamente. En el caso específico de la inferencia práctica, no hay fórmula general alguna que nos permita afirmar cuándo una determinada conjunción de enunciados de indicios *e* (en nuestro ejemplo, M y M' y m) hace a una determinada acción *a prima facie* mejor que a otra *b* y no, digamos, a *b prima facie* mejor que a *a*, a partir de lo que cada miembro de esta conjunción enunciado afirma aisladamente: M hace a *a prima facie* mejor que a *b*, pero M' hace a *b prima facie* mejor que a *a*, etc.

Cabe resaltar, por último, que lo máximo que podemos inferir de un raciocinio práctico es un juicio condicional *prima facie*. En estos términos, los racio-

cinios serían *prácticos* apenas en la medida en que sus premisas y conclusiones expresen evaluaciones o normas sobre acciones, mas no –como una vez más suponía Aristóteles– porque sean capaces de engendrar las acciones como sus conclusiones. Así, las consideraciones aludidas en M y M' y m (o sea, *e*) nos permiten inferir, en la mejor de las hipótesis, que aquellas hacen a *a* (la acción de evitar comer este dulce) *prima facie* mejor que a *b* (la acción de comerlo). Si ninguna otra intención interfiere, tal juicio condicional se reflejará en la intención de realizar la acción *a* en lugar de *b* y esta intención en la realización de *a* en lugar de *b*.

Con todo, las consideraciones axiológicas o normativas expresadas en las premisas y conclusiones del raciocinio jamás nos permiten inferir que el agente realizará la acción *a* en vez de la acción *b*.

Según Davidson, tanto la intención como la realización concreta de la acción *a* en lugar de la acción *b* expresan, en términos prácticos, otro juicio de valor por parte del agente que, contrariamente a aquellos que figuran como conclusiones de las inferencias prácticas, tendría ahora una forma incondicional: *a* es mejor que *b sans phrase*, o *a* es mejor que *b y punto*. Pues, a partir del momento en que un agente tiene la intención de realizar *a* en lugar de *b* o realiza *a* en detrimento de *b*, tenemos que suponer aquellas consideraciones que en el raciocinio práctico hacían a *a prima facie* mejor que *b* y que pasaron a constituir para él una razón *suficiente* para querer actuar y para actuar de forma *a* en lugar de forma *b* en detrimento de las demás consideraciones involucradas. De esta forma, si después de haber inferido que la preocupación por la salud hace a la acción de evitar este dulce *prima facie* mejor que a la acción de comerlo, evito comer, es porque consideré la preocupación por la salud una *razón suficiente* para actuar en detrimento de las demás razones involucradas (como el placer de comer tal dulce). Mi intención de evitar este dulce y mi acción de evitarlo expresan en términos prácticos el juicio incondicional de que evitar comer este dulce es mejor que comerlo *y punto*.

Estas reflexiones sobre la naturaleza del raciocinio práctico le permiten entonces a Davidson formular su solución compatibilista para el problema de la acrasia. En la medida en que expresa en términos prácticos el juicio del agente de que *b* es mejor *sans phrase* que *a*, la realización de la acción *b* en lugar de *a* sería plenamente intencional a la luz de los principios fundamentales P1 y P2 (como observamos, P1 enuncia que, si el agente prefiere realizar $\underline{x}$ a realizar $\underline{y}$, y si se cree libre de hacer $\underline{x}$ o $\underline{y}$, actuará intencionalmente de forma $\underline{x}$, si actúa de forma $\underline{x}$ o $\underline{y}$. Y P2 afirma que, si el agente juzga que la realización de $\underline{x}$ sería mejor –o más correcta– que la realización de $\underline{y}$, y prefiere realizar $\underline{x}$ a realizar $\underline{y}$). Sin

embargo, en la medida en que la realización de la acción *b* en lugar de la acción *a* expresa en términos prácticos el juicio incondicional de que *b* es mejor *sans phrase* que *a*, aquella puede ser contraria al juicio condicional del propio agente de que *pf* (*a* es mejor que *b*, *e*) sin, al mismo tiempo, introducir ninguna contradicción o incoherencia en su sistema de creencias. Así, la realización de una determinada acción *b* en lugar de una acción *a* puede caracterizarse como una acción incontinente (P3) y, aun así, estar en conformidad con los principios P1 y P2.

Sin existir contradicción en juzgar que una determinada acción *b* es mejor *sans phrase* que *a*, y, al mismo tiempo, que la totalidad de las razones conocidas *e* hacen a *a prima facie* mejor que a *b*, pasamos a atribuir irracionalidad al agente cuando podemos atribuirle el siguiente principio de segundo orden: es mejor o debemos siempre actuar en conformidad con lo que consideramos mejor. Como observamos, al comer el dulce que se le presenta, el incontinente no deja de actuar de forma intencional, teniendo *razones* para su acción. Juzga que las consideraciones (M') y (m) hacen a la acción de comer este dulce *prima facie* mejor que a la acción de evitarlo. Considerando, sin embargo, que tal evaluación parcial habría sido suplantada en la ponderación final, según la cual la totalidad de las razones conocidas hacen a la acción de evitar comer el dulce *prima facie* mejor que a la acción de comerlo, al comer el dulce el incontinente pasaría a *ignorar* su propio principio de que debe actuar de acuerdo con aquello que él mismo juzga mejor. Apenas en este momento la irracionalidad y la inconsistencia entrarían en escena. Siendo así, Davidson debe poder explicar cómo alguien puede actuar contra su propio principio de que se debe actuar en conformidad con lo que se juzga mejor, o sea, cómo alguien puede actuar con base en una razón que habría sido suplantada por la ponderación global de todas las razones conocidas.

La respuesta freudiana del autor consiste en la suposición de que la mente del incontinente se divide en este momento en departamentos semi-autónomos. En cada departamento tendríamos la misma estructura de razones y causas bajo la forma de juicios de valor y convicciones que producen las acciones intencionales. El deseo de comer el presente dulce entraría en la decisión en dos momentos. En primer lugar, aparecería bajo la forma de una *razón* a favor de la acción de comer el presente dulce ("la suposición de que los dulces son agradables hace a la acción de comer este dulce *prima facie* mejor que a la acción de evitarlo"), razón ésta que a los ojos del propio agente habría sido suplantada por las razones en favor de la acción de evitar comerlo ("la suposición de que los dulces, aunque agradables, son nocivos para la salud, hace a la acción de evitar comer este dulce mejor *prima facie* que a la acción de comerlo"). Con todo, en la medida en que las mejores razones (contrarias a la acción de comer el dulce y

favorables a la acción de evitar comerlo) se encontrarían en un departamento mental distinto de aquel en el cual se encuentra el deseo en cuestión (de comer el dulce), aquéllas se volverían impotentes frente al deseo mencionado, posibilitando así que tal deseo se manifestara una segunda vez, esta vez ya no más como razón, sino como la *causa* que lleva al agente a ignorar, no solamente sus mejores razones sino también su propio principio de segundo orden, según el cual se debe actuar en conformidad con lo que se juzga mejor. De esta forma, aun suplantado desde el punto de vista racional por las *mejores razones* (en ponderación del mismo agente), el deseo de comer el dulce no sería superado en términos causales.

Sin embargo, al disociar las causas de las razones, reconociendo que el deseo incontinente se impone al agente porque las mejores razones (contrarias) se encuentran en un departamento mental distinto, Davidson (como Aristóteles) parece quitar a la acción incontinente su carácter intencional. A partir del momento en que suponemos que el agente pondera y reconoce que las ponderaciones iniciales, favorables a un determinado curso de acción, fueron *suplantadas* en el cálculo general de las razones involucradas por consideraciones contrarias a tal curso de acción, es difícil imaginar que pueda adoptar tal curso de acción con base en el *reconocimiento* de las ponderaciones iniciales suplantadas.

Autores afines a la posición de Davidson creen poder evadir esa dificultad distinguiendo dos órdenes de razones involucradas en la acción incontinente: por un lado, tendríamos las razones denominadas *directivas* (motivos); por otro, las llamadas *evaluativas* (argumentos). Esa distinción parece expresar las intenciones de Davidson, en la medida en que éste afirma que, aunque el deseo incontinente no constituya una *razón contraria* al principio de orden superior (según el cual debemos actuar en conformidad con nuestras mejores razones), constituiría al menos una *razón para ignorar* tal principio. En estos términos, en la explicación propuesta por Davidson, el incontinente estaría actuando en contra de sus mejores razones de orden evaluativo (juicios condicionales *prima facie*), pero en conformidad plena con razones directivas que harían del suyo un comportamiento intencional.

Esa distinción entre razones *directivas* y razones *evaluativas* se remonta a la disociación empirista tradicional de Hume entre motivos (pasiones) y razones (argumentos). Aun sin suponer, como Hume, que sólo los deseos (pasiones) serían capaces de motivar la acción humana, se supone en todo caso que los deseos serían capaces de constituir motivos para la acción, aunque no constituyan *razones convincentes* a los ojos del agente. Sin embargo, ejemplos del mismo Davidson ilustran de modo convincente que no basta atribuir a un agente una actitud favorable (deseo) hacia una determinada forma de acción, bajo una des-

cripción y una convicción de que tal acción posee propiedades capaces de satisfacer tal deseo, para que podamos afirmar que la realización de tal acción por parte del agente es efectivamente intencional. Así, supongamos que alguien desee recibir la herencia de una tía y se convenza de que sólo causándole la muerte recibirá tal herencia. Con todo –prosigue el ejemplo– imaginemos que la conciencia de tal deseo inescrupuloso deje al agente tan perturbado emocionalmente que conduzca de forma imprudente y acabe atropellando a la propia tía accidentalmente. En este caso, aunque el agente desee, de hecho, la herencia de la tía y esté convencido de que es necesario matarla para poder obtener tal herencia, no diríamos que su acción es intencional, aun en el caso en que tal acción haya sido causada por tal deseo y por la mencionada convicción (bajo la forma de lo que Davidson denomina *razón primaria*). La razón es simple: atropellando a la tía accidentalmente nuestro agente no está actuando a causa del *convencimiento* o del *reconocimiento* de tal deseo como una razón.

De este modo, la propia concepción davidsoniana de lo que cuente como una razón (primaria) me parece insatisfactoria. Para caracterizar *G* como una razón primaria de por qué alguien realizó una acción descrita como *A* (por ejemplo, como la acción de atropellar a la pobre tía), no basta que afirmemos que *G* se compone: (i) de una actitud favorable por parte del agente con respecto a la realización de la acción descrita como *A* con una determinada propiedad y (ii) de la convicción por parte del mismo agente de que la acción descrita como *A* tiene efectivamente la propiedad capaz de satisfacer tal actitud favorable. Es indispensable que supongamos (iii) que la realización de una determinada acción descrita como *A* sea el producto del *reconocimiento* por parte del agente de que su actitud favorable a la acción descrita como *A* constituye una buena razón o una razón de algún modo relevante para su realización. En estos términos, la disociación empirista tradicional entre motivos y razones o, contemporáneamente, entre razones directivas y razones evaluativas, se muestra insostenible.

Davidson no parece resolver, entonces, la paradoja de la irracionalidad. Aunque el juicio incondicional (de que comer este dulce es mejor que no comerlo y punto) no contradiga el juicio condicional (de que tanto la preocupación por el placer como la preocupación por la salud hacen a la acción de evitar comer *prima facie* mejor que a la de comer), nuestra tercera condición nos permite señalar una contradicción en el sistema de nociones del agente, pues, después del cálculo general de todas las razones involucradas, el agente *reconoce*, por un lado, que su preocupación inicial con el placer resultante del acto de comer el dulce no constituye más una buena razón o una razón relevante frente a la preocupación por la salud. Sin embargo, comiéndose el dulce de forma intencional, el

agente estaría, por otro lado, actuando a causa del *reconocimiento* de la preocupación por el placer como una razón relevante.

Solamente hay dos formas de resolver tal contradicción. Según una primera alternativa, podemos suponer que el juicio condicional del agente no es sincero ni auténtico. En este caso, el agente estaría o engañando a terceros o auto-engañándose. Pero podemos suponer también que, comiendo el dulce, el sujeto no esté actuando más de forma intencional, aunque su acción pueda expresar un juicio incondicional, en la medida en que no estaría actuando con base en el *reconocimiento* de la preocupación con el placer como una razón relevante. En esta situación, el sujeto perdería su condición de agente, volviéndose pasivo frente a su propio deseo. Haciendo justicia a la división de la mente sugerida por Freud, no sería más el "yo" o el *ego* quien juzga y desea ("*Ich* will"); tendríamos antes un deseo impersonal imponiéndose de forma ciega y compulsiva: "Se desea" ("*es* wird gewollt").

Bibliografía

Aristóteles (1993): *Ética Nicomáquea*, Julio Pallí Bonet (trad.), Madrid: Gredos.

Aristóteles (1982): "Tópicos", en: *Tratados de Lógica*, Miguel Candel Sanmartín (trad.), Madrid: Gredos.

Davidson, Donald (1980): *Essays on Actions and Events*, Oxford: Clarendon Press.

Pears, D. (1982): "How Easy is Akrasia?", en: *Philosophia* 11, 33-55.

ULYSSES PINHEIRO

Identidad personal e imaginación práctica*

Universidade Federal de Rio de Janeiro/CNPq.

PRESENTO A CONTINUACIÓN una crítica a las teorías reduccionistas acerca de la naturaleza de las personas. Me propongo mostrar que el modelo reduccionista no es capaz de explicar adecuadamente ciertas propiedades esenciales del concepto de persona. Para lograrlo, distinguiré, en primer lugar, algunos de los muchos fenómenos para los cuales tal modelo ofrece una explicación satisfactoria, para que podamos, en seguida, comprender exactamente por qué fracasa en la explicación de otros fenómenos. El examen de la deliberación y de lo que llamo aquí "imaginación práctica" será el caso privilegiado para exponer el mencionado fracaso. Pero el error de la teoría reduccionista en este caso es apenas el indicio o síntoma de una dificultad más general, que será señalada al final del texto.

Para aquellos que defienden una tesis reduccionista en relación con la identidad personal, como Hume y Parfit, una persona está *constituida* por un conjunto de eventos (físicos y/o mentales). A pesar de no ser una entidad separada de esos eventos (como una sustancia cartesiana), tampoco es *idéntica* a esos eventos (así como una estatua constituida de mármol no es idéntica al mármol, pues hay propiedades que atribuimos a la estatua que no pueden ser atribuidas al mármol; se trata de dos tipos de cosas –estatua, mármol–, que tienen criterios de individuación diferentes, y que, por tanto, también tienen criterios de identidad diferentes).

Lo que un reduccionista no podría aceptar es que

(1) Pensamientos, experiencias y acciones son estados o modificaciones de estados pertenecientes a una entidad distinta de esos estados, entidad ésta que permanece idéntica a través de sus modificaciones.

* Traducción del portugués de Tomás Andrés Barrero Guzmán, becario CNPq, Brasilia/Brasil (maestría).

O sea, lo que el reduccionista rechaza cuando rechaza (1) es que las acciones y experiencias dependan de una sustancia[1].

La teoría reduccionista tiene una cierta plausibilidad, ya que suponer la existencia de sustancias parece significar, o bien un compromiso con la existencia de una sustancia cartesiana o bien un compromiso con la tesis de que somos una sustancia extensa [nuestro cuerpo, o una parte de nuestro cuerpo (el cerebro)]. O sea, la tesis anti-reduccionista parece implicar una disyunción compuesta de dos proposiciones falsas, a saber: nuestra identidad personal estaría dada por la existencia continua de una sustancia mental o por la existencia continua de una sustancia corporal, lo que significaría que lo que somos de hecho es, o bien una sustancia cartesiana, o bien nuestro cuerpo.

A pesar de no ser contradictorio suponer la existencia de una sustancia cartesiana, esa suposición parece ser falsa[2], o por lo menos innecesaria como criterio de identidad personal. No parece ser el caso que la identidad personal dependa de una sustancia puramente mental distinta de la conexión de nuestras experiencias; al contrario, parece ser verdad que la continuidad psíquica[3] es una condición necesaria y suficiente de la identidad personal (llamemos a esa la "intuición lockeana" de la Tesis I). Aunque existan sustancias cartesianas, los ejemplos lockeanos parecen mostrar que no es contradictorio pensar que la continuidad psíquica sea "sustentada" por una sucesión de incontables "átomos espirituales", los cuales, por tanto, no harían parte necesaria del criterio de identidad personal –dicho de otra manera, el concepto de persona no parece incluir, entre sus

1. *Cf.* la definición de sustancia de Aristóteles, *Categorías*: el carácter más propio de la sustancia, que explica todas sus otras características, es que, "permaneciendo idéntica y numéricamente una, es apta para recibir los contrarios" (4 a 10). Veremos más adelante, sin embargo, que el anti-reduccionista no tiene que comprometerse con la existencia de sustancias.
2. En esta frase y de aquí en adelante, usaré expresiones como "parece ser falsa", para evitar argumentos circulares, pues lo que debe ser investigado aquí es la plausibilidad de la tesis enunciada en (1).
3. Usando las definiciones de Parfit (1984, pp. 206-207), podemos decir que (1)- "conexión psicológica" es una relación directa de estados mentales entre sí o con ciertas acciones (dejemos vago por ahora lo que significa "relación directa"); (2)- "conexión psicológica fuerte" denota la existencia de conexiones entre dos estadios temporales de persona en un número suficientemente grande (digamos, a cada día, por lo menos la mitad de los estados psicológicos del estadio temporal de una persona está conectada con estados psicológicos del estadio temporal de una persona al día siguiente); (3)- finalmente, "continuidad psicológica" denota la existencia de cadenas superpuestas de conexiones psicológicas fuertes.

notas características, el concepto designado aquí por la expresión "sustancia cartesiana"–. Ya en el caso del cuerpo, parece ser no sólo innecesario, sino imposible (conceptualmente) suponer que el sujeto de inherencia de nuestras experiencias sea simplemente un órgano vivo. Dicho de otra forma: una persona no es un cerebro, pues no es posible atribuir a personas y a cerebros las mismas propiedades[4] –incluso si el fisicalismo es verdadero, o sea, aunque aceptemos que toda propiedad y todo evento mental se reduce a o procede de eventos físicos del cerebro–. Ciertamente, la verdad de esa teoría no se compadece con la tesis intuitiva según la cual las experiencias dependen de sujetos de esas experiencias, pues el cerebro no puede ser tomado como el portador de mis pensamientos o como el sujeto de mis experiencias. Además, la continuidad de átomos corporales tampoco puede ser responsable de la identidad personal, debido a la misma intuición lockeana enunciada anteriormente.

Luego, dada la aparente falsedad de los dos miembros de esa disyunción, la tesis reduccionista tiene una gran plausibilidad inicial.

Pero, por otro lado, la teoría reduccionista carece también *prima facie* de una cierta plausibilidad, pues es preciso rescatar la tesis generalmente aceptada y probablemente verdadera (tal vez hasta conceptualmente verdadera) según la cual experiencias y acciones presuponen los sujetos de esas experiencias y acciones (llamemos a esa última la Tesis II). Ahora bien, la manera más intuitiva de explicar esa dependencia es suponer que esos sujetos son sustancias, y que las experiencias y acciones son cualidades, atributos o modificaciones de esas sustancias.

Lo que un reduccionista tendría que explicar es cómo puede aceptar que experiencias y acciones presuponen sujetos sin que esa tesis sea una instancia de una especie de principio metafísico anti-reduccionista, a saber:

(2) Eventos presuponen sustancias.

La tesis de que experiencias y acciones (en lo que sigue, designaré esa conjunción con EA) dependen de sujetos parece ser necesaria para explicar dos características conceptuales aparentemente no eliminables, conectadas con nuestro concepto de persona, a saber: EA son eventos que existen o *en* sujetos (EA no existen de forma autónoma, por sí mismos, así como el rojo no existe por sí mismo, independientemente de cosas que sean rojas); EA singulares, atribuidas a un cierto sujeto, no pueden ser atribuidas a otros sujetos diferentes.

4. Aun en el caso en que existan sustancias cartesianas, tendrían muchas de las propiedades que usualmente atribuimos a las personas, particularmente propiedades descritas en un vocabulario intencional.

Las tesis de la dependencia ontológica (establecida por la relación de inherencia) y de la predicación de singulares no deben ser entendidas como peticiones de principio contra el reduccionista (podría pensarse que éstas son tesis explícitamente anti-humeanas acerca de la naturaleza de las percepciones, y que, por tanto, no sería sorprendente que un humeano no pudiera ofrecer una interpretación de ellas). Deben ser vistas, más bien, como tesis esenciales a cualquier concepto de persona e, incluso, cuando remplazamos EA por una variable para cualidades en general, como tesis esenciales a cualquier atribución en general, de tal modo que el reduccionista tendría que darles un sentido (aunque diferente del sentido usual, sustancialista), so pena de no conseguir elaborar un concepto consistente de persona.

Pero el reduccionista afirma poder dar un sentido preciso para esas tesis, sin apelar a (2). Si puede lograr eso, su teoría habrá incorporado las Tesis I y II, o sea, será totalmente plausible y consistente, a pesar de contrariar el sentido común.

Parfit afirma que las personas son un tipo de ente similar a las naciones; Perry compara a las personas con juegos[5]; en ambos casos, la propiedad común que justifica la analogía es una cierta falta de definición objetiva en cuanto a la identidad de esos entes. En los dos casos, una cierta base física y mental (ciudadanos y territorio; jugadores y campo de juego) se reúne bajo un solo concepto y dispone de un criterio de identidad en el tiempo, en la medida en que los elementos de esa base interactúan según ciertas relaciones regulares (leyes políticas y jurídicas; reglas del juego). Según el reduccionista, tanto en lo que se refiere a naciones y juegos como en lo que se refiere a personas, las preguntas por la identidad permanecen en algunos casos indeterminadas (no por ausencia de conocimiento suficiente, sino porque el propio fenómeno sería en sí mismo indeterminado). Usemos como ejemplo otro tipo de entidad, clubes; suponiendo que un cierto club, después de estar clausurado por muchos años, fue reabierto por algunos de sus antiguos miembros con el mismo nombre y las mismas reglas que tenía antes, y que no había en las reglas del antiguo club ninguna sobre "refundación", la pregunta "¿Es este club el mismo club, o es otro club, muy parecido al primero?" puede no tener una respuesta determinada, o sea, puede ser intrínsecamente indefinida[6]. Esa indeterminación sería explicada justamente por el hecho de que una persona, una nación y un juego no son entidades independientes de las relaciones establecidas entre los elementos de su base, relaciones éstas

5. Perry (1978, pp. 23-25); Parfit (1984, p. 213).
6. Para este ejemplo, *cf.* Parfit (1984, p. 213).

que muchas veces se modifican gradualmente, preservando la continuidad, pero no la identidad (tanto la relación de continuidad entre etapas temporales totales de un objeto compuesto exclusivamente por tales relaciones como la relación de identidad son relaciones transitivas[7], pero la primera, y no la segunda, admite grados).

La indeterminación esencial de esos fenómenos no afecta, según el reduccionista, las tesis sobre la predicación (la inherencia de los accidentes al sujeto y la singularidad de las cualidades), pues la relación de constitución sería suficiente para introducir el tipo de restricción expresado por esas tesis. Tomemos como paradigma el modelo del juego, por ejemplo, el fútbol. Los partidos de fútbol son constituidos por 22 jugadores, por un campo dividido por la mitad, cada mitad con dos señales en los extremos y por una bola, además de un árbitro y dos auxiliares, etc. Esos elementos están conectados por ciertas reglas (relativas a los objetivos del juego, a las faltas, a su duración, etc.). Ahora bien, una jugada de fútbol depende esencialmente del juego para existir, o sea, sólo existe en un juego (lo que era exigido por la tesis de la inherencia). Las jugadas no existen aisladamente. Además, jugadas singulares, pertenecientes a un cierto juego, no pueden ser atribuidas a otros juegos diferentes, por la simple aplicación de la Ley de Leibniz (lo que era exigido por la tesis de la singularidad). O sea, esos eventos dependen de la serie que ellos mismos constituyen, y son atribuidos a esas series, y solamente a esas series.

Lo mismo, dice el reduccionista, vale para el concepto de persona. Una persona no es, según aquél, nada más que una serie de eventos tales como pensamientos, experiencias y actos, conectados entre sí por relaciones que preservan la continuidad psicológica, desde que sea dado el tipo adecuado de causa (a saber, la causalidad no-ramificada de un cuerpo humano vivo, o por lo menos de un cerebro). Las relaciones psicológicas que preservan la continuidad son variadas, incluyendo la memoria de experiencias y actos pasados; la intención, que conecta deseos con acciones futuras; el carácter, que mantiene unidas las acciones y disposiciones (en lo que se refiere al cuerpo, la continuidad implica al menos un cambio gradual de las partes que lo componen –digamos, no más que 50% de los elementos constituyentes pueden variar en un corto intervalo de tiempo–). También en ese caso, estaría asegurada la verdad de las tesis sobre la inherencia y la singularidad.

El hecho de que la metáfora del juego no presenta un análogo para todas esas características del concepto de persona (por ejemplo, no podemos decir que un

7. Como se sabe, una relación R es transitiva si de xRy e yRz, se sigue xRz.

juego "actúa" o "decide" algo) no se debe a un problema con la noción de constitución, sino simplemente a las diferencias entre los conceptos de persona y de juego[8]. Lo mismo vale para la construcción de metáforas similares.

Si la explicación elaborada por la teoría reduccionista sobre la naturaleza de las personas es correcta, eso la compromete con la tesis de que una redescripción impersonal de las EA que describimos como pertenecientes a una cierta persona es necesariamente posible, pues una persona no es una entidad aparte de la serie de eventos que la constituye, de tal modo que podemos redescribir sus acciones y experiencias sin mencionar aquello que ellas constituyen. En el caso eventual en que no pudiéramos hacer eso, no podríamos explicar la relación de constitución; análogamente, debemos poder describir un pedazo de mármol independientemente de la referencia al hecho de que el mármol constituye una estatua, si es que podemos afirmar que constituye la estatua, pues si eso fuera imposible, no podríamos explicar de forma no-circular en qué consiste esa relación de constitución. El reduccionista podría hacer una precisión en este punto, observando que está buscando un reduccionismo ontológico, y no un reduccionismo semántico; o sea, nos recordaría que su teoría no pretende eliminar el concepto de persona, ni afirmar que el concepto de persona tiene el mismo contenido que el concepto de sus elementos constituyentes, así como la existencia y el concepto de la estatua no pueden ser reducidos al concepto de mármol. En otras palabras, las personas existen, y el concepto de persona no es reducible al concepto de sus elementos constituyentes, pero las personas no existen independientemente de esos elementos, ni pueden causar eventos de otra forma que no sea a través de la relación causal entre sus elementos constituyentes y sus efectos[9]. Pero esa precisión no elimina el hecho de que, para el reduccionismo ontológico, debe ser posible describir los elementos constituyentes de una entidad sin mencionar aquello que ellos constituyen y que, una vez hecha esa descripción, no habría

8. Sería posible construir nuevas analogías para ilustrar metafóricamente esa tesis; por ejemplo, equipos de fútbol, constituidos por jugadores, "actúan" y, en cierto sentido, "deciden". Si esa analogía no es adecuada, podemos formular otras –o entonces admitir que hay especificidades en el concepto de persona que imponen un límite a esas comparaciones metafóricas.

9. Veremos más adelante que el anti-reduccionista puede concordar con la primera de esas tesis (personas no existen independientemente de sus partes constituyentes), pero no con la segunda (todas las relaciones causales entre personas y sus acciones pueden ser redescritas en términos de sus partes constituyentes). El reduccionista, por razones que examinaremos a continuación, debe aceptar ambas tesis.

ninguna entidad que hubiera sido omitida; la descripción habría sido, para todos los propósitos relacionados con la ontología, *completa*.

Sin embargo, es posible formular una objeción a la teoría reduccionista al considerar que el concepto de persona impone una restricción más fuerte a las tesis sobre la inherencia y la singularidad de las que se derivan de los conceptos de juego, de equipo o de nación (y otros similares). La metáfora, en este caso (como en muchos otros[10]), es engañosa. Esa restricción, como veremos, nos llevará a interpretar las tesis sobre la inherencia y la singularidad en términos de (1), o sea, como una aplicación del principio enunciado en (2) –o mejor, debido a razones adicionales, expuestas más adelante, como una reformulación de (2).

Además de las tesis sobre la inherencia y la singularidad (comunes a toda predicación), el concepto de persona también incluye entre sus notas características la capacidad de deliberación práctica[11]. La deliberación práctica, a su vez, involucra la facultad de la imaginación práctica[12], que presenta al sujeto posibilidades de futuros alternativos. Tomando *x*, *y* y *z* por variables para personas (o, más precisamente, para estados temporales de personas), parece que podemos afirmar que, *en los términos de la teoría reduccionista*, la efectuación de esa facultad es equivalente a (define) la identidad entre *x* y *y* dadas las siguientes condiciones[13]:

10. Lo que demuestra la necesidad de control de los "experimentos de pensamiento" y de la elaboración de ejemplos y contra-ejemplos "intuitivos". Ese control sólo puede ser dado por el análisis conceptual, que delimita la validez y la extensión de las comparaciones y los ejemplos.
11. Aunque fuera posible defender la tesis poco plausible de que todas nuestras deliberaciones prácticas son racionalizaciones *a posteriori* que encubren los "verdaderos" motivos de nuestras acciones (instintos, deseos inconscientes, etc.), aun así sería preciso afirmar que hace parte del concepto de persona la realización de pseudo-deliberaciones (pues las personas necesariamente creen que están deliberando). Veremos más adelante, sin embargo, en qué sentido esa tesis es no solamente poco plausible, sino falsa.
12. Al menos en el caso de los seres humanos, o, de modo todavía más restringido, al menos para algunas de las deliberaciones de los seres humanos (*cf.* nota 18 *infra*). Esa facultad no pertenece al concepto de persona tomado en general –por ejemplo, no pertenece al concepto de Dios.
13. Además, es claro, de otros requisitos generalmente aceptados por el reduccionismo como necesarios para la preservación de la identidad, tales como: no-ramificación, presencia del tipo adecuado de causa, etc., que serán tomados aquí como evidentes y satisfechos por los casos en examen.

($x = y$) si y solamente si: (i) x en t_1 anticipa imaginativamente EA de z en t_2; (ii) x piensa (cree) que $x = z$; (iii) y realiza en t_2 algunas de las EA anticipadas por x en t_1; (iv) y contiene en su memoria un recuerdo aparente de haber anticipado esas EA en t_1; (v) el hecho de que x en t_1 haya anticipado EA de z en t_2 está causalmente conectado tanto con esa memoria aparente de y como con la realización, en t_2 y por y, de las EA anticipadas por x en t_1.

Antes de mostrar por qué el análisis completo de ese fenómeno afecta la teoría reduccionista, es preciso hacer algunas observaciones preliminares: en primer lugar, debemos notar que la afirmación de que y realiza algunas de las EA anticipadas por x no significa, obviamente, que se trata de las mismas EA singulares en uno y otro caso, pero sí que x anticipa EA de un cierto tipo al imaginarlas siendo realizadas por z y que y realiza EA singulares que pertenecen al mismo tipo imaginado por x (podríamos designar el tipo con la expresión EA* y su instancia con la expresión EA). En segundo lugar, y para llamar la atención sobre una característica similar a la enunciada en el primer punto, debemos notar que z, estrictamente, no designa una persona ni al menos una etapa temporal de una persona, sino una ficción forjada por x (otra manera de describir la variable z sería decir que z designa un tipo de persona y que x y y concretizan singularmente ese tipo[14]; en ese caso, en vez de $x = z$, sería mejor decir que x es una instancia de z – o, si quisiéramos distinguir las variables para personas de las variables para tipos de personas, podríamos incluir un asterisco, como: z^*, y decir que x y y ejemplifican z^*–)[15]; en tercer lugar, es necesario notar que la realización efectiva de algunas de las EA por y es requisito para distinguir la facultad de la imaginación práctica de un mero devaneo de la imaginación; finalmente, debemos observar que la actividad de la imaginación práctica volcada para la conjunción EA se aplica más a acciones que a experiencias, pues nuestras deliberaciones normalmente resultan en la realización de acciones (aun cuando deliberamos acerca de la posibilidad de tener tal o

14. *Cf.* la noción de un "Adán en general" propuesta por Leibniz en su correspondencia con Arnauld. *Carta X*, 4/14 de julio de 1686.

15. Aplicando esas dos observaciones a lo que fue dicho antes, podríamos reformular la definición de identidad entre x e y de la siguiente forma:
($x = y$) si y solamente si: (i) x en t_1 imagina EA* realizadas por z^* en t_2; (ii) x piensa (cree) que x es una instancia de z^*; (iii) y realiza en t_2 algunas de las EA anticipadas por x en t_1; (iv) y contiene en su memoria un recuerdo aparente de haber anticipado esas EA en t_1; (v) el hecho de que x en t_2 haya imaginado EA* realizadas por z^* en t_2 está causalmente conectado tanto con esa memoria aparente de y como con la realización, en t_2 y por y, de las EA anticipadas por x en t_1.

cual experiencia, eso generalmente significa que deliberamos acerca de las acciones que nos permitirán tener esas experiencias[16]). Aun así, como veremos, las consecuencias relativas a las acciones, que intentaré establecer a continuación, deberán extenderse también a las experiencias.

Entre las cinco condiciones formuladas anteriormente en términos de la teoría reduccionista para que el ejercicio de la imaginación práctica garantice la identidad personal entre dos etapas de persona, las condiciones (ii), (iv) y (v) son indicios de que la tesis reduccionista es falsa. Recordemos esas tres condiciones (no señalaré por ahora las diferencias entre acciones singulares y tipos de acciones ni entre un individuo singular y un tipo de individuo imaginado): (ii) que *x* piense (crea) que *x* = *z*; (iv) que *y* contenga en su memoria un recuerdo aparente de haber anticipado esas acciones en t_1, y (v) que el hecho de que *x* haya anticipado acciones de *z* en t_2 esté causalmente conectado tanto con esa memoria aparente de *y* como con la realización, en t_2 y por *y*, de las acciones anticipadas por *x* en t_1. Veremos que, en el contexto de la explicación de la deliberación, la condición (iv) no puede ser interpretada en términos de "memoria *aparente*", pero sí de "memoria *real*", lo que significa que no es posible sostener una tesis reduccionista sobre la identidad personal[17]. Las condiciones (ii) y (v), por su parte, no pueden, en ese mismo contexto, ser interpretadas en términos impersonales –lo que, una vez más, tiene como consecuencia una refutación del reduccionismo–. Comentaré, pues, esas tres condiciones conjuntamente, para mostrar en qué sentido falsifican el reduccionismo.

Vimos anteriormente que, considerando que *z* es un ser imaginario, la expresión *x* = *z* tal vez debería ser interpretada más propiamente como: *x* es una instancia de *z** (de modo simétrico, distinguimos antes EA* de EA). Pero ¿qué significa precisamente afirmar que *z** es un "ser imaginario"? Ciertamente, nos imaginamos a nosotros mismos en situaciones futuras alternativas, y esa anticipación imaginativa es un elemento importante en nuestras deliberaciones, en la medida en que ayuda a determinar[18] el valor de las motivaciones en competen-

16. "Generalmente", pero no siempre; podemos, por ejemplo, deliberar acerca de la posibilidad alternativa de o bien entregarnos a ciertos sentimientos de melancolía y desespero, o bien mantener la concentración y la sobriedad.

17. Perry (1978, pp. 30-31) introduce la noción de "memoria aparente" conectada con relaciones causales para evitar la objeción de Butler a Locke, según la cual la explicación de la identidad personal en términos de memoria sería circular. *Cf.* también Perry (1975, p. 147).

18. ¿Eso significa que la imaginación práctica es una condición necesaria para la determinación del valor de nuestras motivaciones? No creo. Algunas veces, podemos

cia por nuestro asentimiento. Pero, al analizar esa anticipación, es importante notar que nosotros nos imaginamos *a nosotros mismos*; en ese sentido, la relación entre *x* y *z* no debe ser tomada como la relación entre una instancia concreta y su tipo general, sino como una relación simple de identidad[19]. Evidentemente, no se trata de afirmar una identidad a lo largo del tiempo entre *x* y *z*, pues *z* no es una entidad real, y sí un ser imaginario; ahora bien, sin entidad, no hay identidad. Tal vez, como una "*façon de parler*", podríamos decir como máximo que se trata de la identidad de *x* consigo mismo, por la simple tautología según la cual toda entidad es idéntica a sí misma. Nosotros nos consideramos a nosotros mismos en el futuro por la imaginación, lo que significa apenas que *x* se considera a sí mismo como si estuviera en el futuro, reaccionando, sin embargo, con las razones y motivaciones presentes para deliberar sobre los escenarios alternativos que se le presentan a su imaginación (*cf.* Perry, 1976, pp. 75-78). En ese caso, *x* y *z* no se relacionan como "estadios temporales de persona" diferentes, sino pura y simplemente como la misma persona, sin calificativos adicionales. Viendo las cosas de esa manera, no se trata exactamente de suponer que *x crea* que es el mismo que *z*; simplemente *sabe* que es *z*, pues está refiriéndose a sí mismo, imaginándose en el futuro.

Pretendo mostrar que esa identidad entre *x* y *z*, en la medida en que es condición de una deliberación *real*, es decir, en la medida en que involucra la atribución de un *poder de elección* efectivo a *x*, no puede ser expresada en términos impersonales, sino que involucra una referencia a sí como sujeto de la deliberación. En seguida, mostraré que esa referencia a sí es esencial para explicar la relación causal entre los elementos contenidos en *x* y *y*, de lo que se debe concluir que la persona que está constituida por los elementos presentes en *x* y *y* no es

---

deliberar considerando de modo puramente racional las alternativas, usando para la decisión reglas que también son ellas mismas puramente racionales (es eso lo que ocurre, por ejemplo, en ciertos problemas morales abstractos y complejos). Pero aun en el caso en haya motivaciones puramente racionales no ligadas a deseos considerados por los agentes en situaciones imaginarias, es preciso que estos últimos se refieran a sí mismos en el futuro por algo análogo a aquello que es designado aquí por *z*. Además, incluso en el caso de motivaciones puramente racionales, puede haber apelación a procesos imaginativos en la deliberación.

19. En algunos otros contextos teóricos será importante analizar esa relación como incidiendo sobre tipos de ente y sus instancias –por ejemplo, en el examen del compatibilismo entre libertad y determinismo, que es el caso que interesaba a Leibniz cuando propuso que Arnauld considerara la diferencia entre la noción completa y la noción general de Adán.

reducible a la serie que los contiene. O sea, del análisis del significado de la deliberación que una persona hace entre las posibilidades alternativas que se representa a sí misma, debemos concluir que el agente que delibera no puede reducirse a una serie de eventos (en el caso que investigamos, a intenciones, compuestas por creencias y deseos, y a acciones), sino que, por el contrario, la identidad entre *x* y *y* es una *condición* de la deliberación, y no el *resultado* de la conexión entre intenciones y acciones.

La demostración de esta tesis consiste, por tanto, en mostrar que la teoría reduccionista no puede explicar la deliberación, o sea, que esta última presupone un modelo anti-reduccionista. Veamos por qué. Deliberar involucra considerar acciones alternativas (digamos, A o B, donde la disyunción es entendida como exclusiva, o sea, donde hacer B implica no hacer A) atribuidas *al mismo* sujeto. O sea, es preciso que *x* se represente a sí mismo como permaneciendo igual, sea que haga A o B. Y todavía más: es preciso suponer que esa representación sea no solamente una *creencia* necesaria conectada con nuestra perspectiva pragmática, sino, más bien, que la *verdad* de los contrafácticos generados a partir de ella esté garantizada, si es que de hecho hay deliberaciones. Digamos que *x* está en el presente deliberando acerca de la posibilidad de hacer A o B; digamos, además, que la proposición que afirma que *y* hará A sea verdadera. Si la deliberación es un fenómeno real (y no una mera racionalización que oculta motivos inconscientes), entonces es preciso que sea verdadera la proposición contrafáctica que afirma que, si *y* no hubiera escogido A, *y* podría haber escogido que B sea el caso. Si la proposición contrafáctica que afirma que B habría sido el caso es verdadera, esta situación incluye el hecho trivial de que el sujeto designado por la misma variable *y* en el antecedente y en el consecuente del condicional contrafáctico sea el mismo ente[20] (ya porque hay identidad transmundana, ya porque hay contrapartes que dan las condiciones de verdad de la referencia contrafáctica a entes *de este* mundo actual), pues sin eso no habría de hecho deliberación de un sujeto ante dos posibilidades alternativas. Esto, a su vez, significa que *z* debe representar indiferentemente *y* en el mundo actual y en los

20. La tesis enunciada aquí no se resume a la comprobación de esa obviedad (a saber, que el empleo de *y* en los dos lados del contrafáctico designa el mismo individuo perteneciente a un cierto conjunto de individuos, que da el valor de aquella variable), sino que afirma que tal contrafáctico hace parte de toda y cualquier deliberación. Esta tesis es ella misma ciertamente obvia desde el punto de vista de nuestro concepto común de deliberación, pero puede plantear algunos problemas metafísicos difíciles y, por tanto, merece ser enunciada claramente.

mundos posibles imaginados por *x*, es decir (dado que en el mundo actual futuro $x = y$), debe representar *x*, sea que *x* esté conectado con *y* realizando A o realizando B.

Estoy asumiendo que una "descripción fenomenológica" cuidadosa de la deliberación incluye por lo menos los rasgos recién descritos. Lo que significa que, si la teoría reduccionista es verdadera, debe poder dar cuenta de todos los hechos mencionados en el último párrafo. Y, de hecho, creo que puede dar cuenta de *casi* todo lo que está contenido en esa descripción. Distinguiré, entonces, en primer lugar, los rasgos que pueden ser plenamente explicados por el modelo reduccionista, para que podamos, en seguida, comprender exactamente por qué éste fracasa al explicar los rasgos restantes.

Es *compatible* con la tesis reduccionista de que una persona es reducible a una serie de eventos psicofísicos:

1- *Que* x *en* t_1 *se imagine realizando las acciones alternativas A o B.* O sea, no es incompatible con el modelo reduccionista afirmar que la "idea de sí" requerida por la deliberación es necesariamente vaga, bastando que el agente imagine contrapartes más o menos parecidas consigo realizando acciones mutuamente excluyentes. La representación imaginativa involucra necesariamente una cierta vaguedad en relación con el pasado y con el futuro (por ejemplo, en relación con los escenarios alternativos de posibilidades, que son imaginados a grandes rasgos, sin detalles excesivos, poco importantes para la deliberación en curso, o en relación con el pasado, cuyas determinaciones causales complejas no pueden ni aproximadamente ser abarcadas por nuestra mente, o incluso en relación con el futuro, pues no sabemos ni al menos lo que decidiremos efectivamente de aquí a cinco minutos[21]). Esta vaguedad explica por qué, aunque la serie futura que contiene A no sea estrictamente idéntica a la serie contrafáctica que contiene B, el reduccionista puede afirmar, en los términos de su propia teoría, que *x* se imagina a sí mismo como haciendo A o B. En ese sentido, no sería preciso figurarse la relación entre *x* y *z* como la de un singular con su tipo, sino simplemente como la relación de un singular consigo mismo –esa imaginación consistiría en la representación (más o menos distinta) de la serie que contiene

21. Este último punto es importante para algunas teorías sobre responsabilidad moral, por ejemplo, para la teoría compatibilista de Moore, pues es, según este autor, la ignorancia de las determinaciones pasadas y de las decisiones futuras lo que garantiza la corrección de la atribución de responsabilidad moral en un mundo determinista.

x hasta t_1 como siendo igualmente compatible con la elección de A y con la elección de B (o, en el caso de *y*, con las acciones A y B).

II- *Que sean atribuidas propiedades disyuntivas a* x *en* t_1 *y a* y *en* t_2. De hecho, aun para un humeano, que pretende reducir personas a series de eventos, es decir, que pretende dar cuenta de la identidad de las personas por medios puramente cualitativos, excluyendo de la comprensión de ese concepto toda unidad no cualitativa (unidad ésta que consistiría en ser una entidad cualquiera separada de la serie de eventos, o sea, una entidad que permanecería la misma a través de modificaciones[22]), no es preciso comprometerse con la tesis leibniziana de que posibilidades alternativas no pueden ser atribuidas al mismo individuo. Basta que el individuo en cuestión sea definido por conceptos disyuntivos[23]. Así, un reduccionista puede expresar, en los términos de su propia teoría, que *x* es aquel que se representa haciendo A o B y que *y* es aquel que escoge y hace A o B[24].

III- *Que haya referencia contrafáctica a* y *como aquel que hace A pero podría haber hecho B.* Incluso para el reduccionista que comprende las personas como series de eventos, es decir, que comprende la identidad personal de modo puramente cualitativo, a través de la noción de una serie continua, es posible

22. Aquí es necesaria una distinción entre tipos de propiedades: la hecceidade (o ipseidad) y la cualidad; la primera, que es la propiedad de "ser un individuo", es distinta de las propiedades que expresan puras cualidades, que son generales y no-relacionales. O sea, la primera expresa justamente la propiedad de "no ser una cualidad". Esas definiciones son interdependientes e incluso circulares ("ser un individuo" es "no ser una cualidad", y viceversa), pues tal vez expresen conceptos primitivos, cuyo contenido puede ser elucidado, pero no definido propiamente.

23. *Cf.* sobre este punto Adams (1979, pp. 9-10). El autor llama la atención sobre el hecho de que, en la teoría de Leibniz, "el concepto de un individuo, el cual, por decirlo así, expresa la propiedad de ser aquel individuo, difiere de los conceptos más generales por ser *completo*", lo que significa que "ningún nuevo contenido puede serle adicionado consistentemente". Esto, por su parte, implica que no hay identidad transmundana, o sea, que no es posible construir conceptos completos a partir de predicados indexicados a mundos posibles.

24. Tal vez en este punto yo esté concediendo al reduccionista más de lo que debería. De hecho, ¿no sería problemático afirmar que una *serie* puede tener propiedades disyuntivas? ¿Eso no afectaría la individuación de esa serie? Cuando consideramos que esas disyunciones generan otras disyunciones, en ramificaciones crecientes de posibilidades alternativas, ¿hay una figuración precisa que pueda ser referida a la serie en cuestión? Quizás la inclusión de grados de complejidad no afecte la teoría reduccionista sobre propiedades disyuntivas, pero no estoy seguro con respecto a eso. De cualquier forma, no es sobre ese punto que mi crítica incidirá.

enunciar contrafácticos sobre personas[25]. La teoría reduccionista sobre este punto sería más o menos la siguiente[26]: algo existe si y solamente si la totalidad de ese algo existe; verbigracia, si y solamente si existe un mundo en el cual la totalidad de ese algo existe, v.g., si y solamente si, cuantificando sobre partes de ese mundo, la totalidad de ese algo existe, v.g., si y solamente si *es* esencialmente parte de algún mundo y, por tanto, si y solamente si *no es*, por esencia, un individuo transmundano. O sea, partes de mundos son individuos posibles, e individuos transmundanos son individuos imposibles (Lewis, 1995, p. 211). Aun así, el reduccionista puede afirmar que la referencia contrafáctica a eventos alternativos a aquellos que efectivamente constituyen la serie que es una persona puede ser pensada en los propios términos de su teoría. Individuos nunca existirían en más de un mundo posible, sino que existirían representados en otros mundos por sus contrapartes, es decir, por individuos semejantes a ellos, pero con algunas propiedades diferentes (por ejemplo, divergentes en cuanto al acto de hacer B en vez de A –lo que, presumiblemente, incluye divergencias subsecuentes, aunque antes de t_2 las series fueran idénticas)[27]. Así,

25. También en este otro punto tal vez yo esté concediendo al reduccionista más de lo que debería. De hecho, ¿será que la atribución de posibilidades alternativas no depende de alguna forma de la existencia de identidad transmundana? No estoy seguro con respecto a la respuesta, o sea, no estoy seguro acerca de cuáles son las condiciones formuladas en términos de mundos posibles para la atribución de posibilidades alternativas a un individuo. Para una crítica al reduccionismo de cualidades, *cf.* Adams (1979). Pero este tampoco es el punto sobre el cual incidirá mi crítica.
26. Adapto los puntos siguientes de la teoría sobre juicios contrafácticos de Lewis (1995), especialmente pp. 210 y ss.
27. Ellas serían idénticas incluso en relación con *x*, que tendría la propiedad disyuntiva de "devenir A o B". La teoría de Lewis sobre este punto es más compleja que la expuesta hasta aquí, pues, al admitir el principio de la composición irrestricta, este autor puede *definir* propiedades de un individuo transmundano. El punto de partida de esa definición sería dado por la consideración de que los estadios de un individuo transmundano serían sus partes posibles maximales, o sea, serían las intersecciones de ese individuo transmundano con los mundos a los cuales él se sobrepone, es decir, serían los individuos posibles. A partir de ahí, es posible *definir* un individuo transmundano como la suma de sus estadios. De esa forma, sin desechar una concepción puramente cualitativa de los individuos posibles, esto es, sin desechar la tesis de que un individuo posible existe en por lo menos un y no más que un mundo posible, su teoría puede, de forma ingeniosa, referirse a individuos transmundanos. Evidentemente, estos últimos no tendrían una unidad integrada de tal modo que se manifestara como auto-interés, deliberación práctica, planes, etc. En ese sentido, un individuo de ese tipo no podría ser llamado, propiamente, una *persona*.

es posible afirmar que *y* podría haber hecho B no porque él mismo hizo B en otro mundo posible diferente de este mundo actual –*él* no lo hizo–, sino porque tiene contrapartes que hacen B en otros mundos posibles. La relación de "ser una contraparte de" es simétrica, de modo que la contraparte de *y* tiene a *y* como su contraparte (o sea, *y* es la contraparte de su contraparte); de ese modo, es posible afirmar de *y*, y no de otro individuo cualquiera, que *él* podría haber hecho B –lo que, sin embargo, sólo puede ser representado por su contraparte–[28].

Es posible reunir en una sola descripción esos tres rasgos presentes en la deliberación práctica (imaginación práctica, propiedades disyuntivas y atribución de contrafácticos), de una manera compatible con la teoría reduccionista. Como vimos, según la tesis reduccionista, una persona se mantiene igual si hay continuidad psicológica entre sus estados mentales, garantizada por el tipo adecuado de causa y sin bifurcación, sin que sea preciso (ni posible) suponer la continuidad de una entidad separada de los elementos de la serie. Es compatible con esta tesis afirmar que tanto la serie A como la serie B son continuas con la serie que contiene *x* (en la medida en que la deliberación en t_1 causa la acción correspondiente en t_2) y que la serie que contiene *x* representa imaginativamente esa distinción en t_1, de modo que, según el reduccionista, el juicio contrafáctico que afirma que *aquella misma* persona que hizo A podría haber hecho B ganaría un sentido preciso, a saber: la serie B sería la contraparte contrafáctica que corresponde a la serie A, que es real.

O sea, aunque admitamos que una persona no es nada más que una colección de cualidades[29], es posible dar cuenta de esos aspectos de la deliberación. No hay paradojas a ese nivel, incluso cuando consideramos que la persona constituida por A es la misma persona designada por *x*, así como la persona constituida por B es la misma persona designada por *x*, pero la persona constituida por A no es la misma que la persona constituida por B. El reduccionista podría afirmar que, después de la división contrafáctica en t_2, la secuencia de la serie A es distinta de la secuencia de la

28. Lewis (1995, p. 217) expresa ese último punto afirmando que muchos mundos posibles representan *de re*, acerca de un individuo, que él existe, a través de las contrapartes.

29. Veremos, al final de este texto, que la alternativa al reduccionismo no sería la tesis según la cual una persona es algo separado (una sustancia) y distinto del conjunto de las cualidades o partes que componen un individuo complejo, sino, más bien, la tesis según la cual una persona tiene propiedades que solamente pueden ser atribuidas al todo, y cuya condición de atribución es que la propia persona se perciba como un todo.

serie B; luego, la afirmación de que la serie que contiene A es la "misma" persona que la serie que contiene B se debe apenas al hecho de que, al concentrarnos sólo en la elección restringida a A y B, podemos pensar (erróneamente) que lo que sobra de la serie continuaría "de la misma" forma en A o B. Que eso no es así, sin embargo, queda claro cuando interpretamos A y B como designando elecciones cruciales (morales o existenciales), que modificarán radicalmente el tipo de persona que hasta entonces existía; o cuando consideramos que la situación descrita por la elección entre A y B, incluso cuando se trata de elecciones banales, en verdad es repetida indefinidamente en nuestro día a día, generando una ramificación contrafáctica extremamente complicada, que distingue personas radicalmente, aunque imperceptiblemente, en el dominio de ese "jardín de senderos que se bifurcan".

Una última característica presente en la deliberación, pero que tal vez no valga la pena mencionar, por obvia, es la identidad entre *x* y *y*. Esa identidad fue justamente definida anteriormente en términos de la deliberación, o sea, listando las condiciones (i) a (v) que, descritas en términos reduccionistas, presumiblemente explicarían la deliberación práctica. En este caso, el reduccionista podría simplemente afirmar que la identidad entre *x* y *y* consiste en la relación de eventos mentales y físicos usando los conceptos ya mencionados, que serían, como acabamos de ver, *compatibles* con las tres características esenciales enumeradas hasta aquí para que haya deliberación. Independientemente de mi crítica subsecuente a la descripción reduccionista, es forzoso admitir que la teoría de la identidad personal formulada en términos reduccionistas no es, en cuanto tal, contradictoria. O sea, considerando el proyecto reduccionista en general, presenta un modelo consistente para dar cuenta del concepto de persona y de los criterios de identidad personal, y una eventual crítica a ese modelo no consiste en revelar que es "de hecho" contradictorio, sino apenas que no consigue describir no-circularmente la deliberación.

La compatibilidad de la teoría reduccionista con los puntos hasta ahora analizados se hace más evidente cuando consideramos que, incluso para los anti-reduccionistas que defienden la existencia de hecceidades, no está garantizada la identidad personal transtemporal, sino apenas (tal vez) la identidad transmundana[30]. Que no está garantizada la identidad personal transtemporal por el simple hecho de admitir hecceidades queda claro cuando consideramos la intuición lockeana. Según Locke, sería posible suponer la existencia de varias "sustancias cartesianas" sucediéndose bajo la continuidad psicológica dada por relaciones de memoria, sin que ninguna de esas incontables sustancias desempeñara ningún papel en la atribución de iden-

30. Esa última implicación estaría garantizada una vez se refutara el Principio de la identidad de los indiscernibles.

tidad personal. En ese caso, la propiedad de "ser Pedro" (su hecceidad) en cualquier momento del tiempo no sería equivalente a ninguna propiedad cualitativa, y no tendría nada que ver con su identidad numérica en el tiempo. Sería la conexión causal de propiedades cualitativas de Pedro en t_1 con propiedades cualitativas de Pedro en t_2 lo que explicaría su identidad personal transtemporal[31]. De esa forma, aunque el estadio temporal de Pedro en t_1 fuera diferente del estadio temporal de Judas en t_1 debido a que ambos tienen hecceidades diferentes, y aunque la propiedad de tener una hecceidad fuera transmitida a lo largo de los estadios temporales de cada uno de ellos, no sería necesario que se trate de la misma hecceidade a lo largo de la serie de estadios temporales de cada uno de ellos. Pedro podría distinguirse de Judas por tener una hecceidad diferente de la de Judas, en cada momento del tiempo, y por sus propiedades cualitativas diversas, tanto en un dado momento del tiempo como trans-temporalmente, y aun así su identidad personal podría ser explicada en términos de propiedades puramente cualitativas[32]. Luego, el reduccionista no estaría en desventaja en este punto con respecto al anti-reducionista.

Pero hay por lo menos una característica enumerada anteriormente en la "fenomenología de la deliberación" que no es captada por la teoría reduccionista. Se trata de la atribución de un poder de elección a la persona que se imagina haciendo A o B y a la persona (¿la misma?) que opta entre A y B y escoge A porque quiere. Examinemos, pues, en qué sentido la teoría reduccionista fracasa por ser *incompatible* con esta característica.

De hecho, en la descripción de deliberación propuesta anteriormente, uno de los rasgos esenciales era que, considerando que la deliberación es un fenómeno real (y no una mera racionalización), es preciso que sea verdadera la proposición contrafáctica que afirma que, *si y* no hubiera escogido A, *y* podría haber escogido que B fuera el caso. Es claro que podemos formular un contrafáctico semejante

31. Sobre este punto y la secuencia de este párrafo, *cf.* Adams (1979, p. 20).
32. Nótese que la admisión de hecceidades *puede* desechar la apelación a contrapartes, en la medida en que tal admisión implica la identidad transmundana, pero que, por otro lado, la admisión de hecceidades *no implica* el rechazo de que hay, en los mundos posibles, contrapartes de los individuos actuales. O sea, la referencia a contrapartes podría estar presente en el caso de teorías que afirmaran que una persona es irreducible a la serie de eventos que le son atribuidos. Digamos que una persona sea una sustancia cartesiana; si esa sustancia optara por B en vez de A (habiendo sido A su opción en el mundo real), entonces sería una contraparte de la sustancia real, pero no (por la Ley de Leibniz) *la misma* sustancia que de hecho hizo A. Esa es la dificultad general que Arnauld le señaló a Leibniz en su correspondencia sobre el Parágrafo 13 del *Discurso de metafísica*.

con respecto a *x* y a sus acciones futuras[33] –A o B–, teniendo en cuenta que, en el mundo actual, $x = y$; pero formular las cosas de ese modo indica que estamos atribuyendo un poder de elección no solamente a *y*, sino también a *x*, o sea, al antecedente causal de *y* en t_1. Al imaginar posibilidades alternativas, *x* ya tenía un poder de elección con respecto a ellas, si es que su imaginación es un elemento en una deliberación efectiva.

Lo que no puede ser representado por la teoría reduccionista es justamente ese poder de elección atribuido a *x* y a *y*. De hecho, tal poder no es representado ni por las propiedades disyuntivas, ni por la imaginación vaga de alternativas, ni por los juicios contrafácticos relativos a *y* (y a *x*). Afirmar que una persona tiene poder de escoger entre alternativas no se reduce a (aunque involucre) afirmar que le es atribuida la acción de "hacer A o B", ni que esa persona puede imaginarse haciendo A y haciendo B, ni que, escogiendo A, hay una representación contrafáctica en la cual ella es figurada escogiendo B. No hay ninguna cualidad, en la serie a la que se reduciría una persona, que corresponda a ese poder de elección. Esa ausencia se explica por el hecho de que el poder de elección involucra la totalidad de lo que la persona es. Dicho de otra forma, una elección no es algo que *le ocurre* a una persona, sino que la persona *es* sus elecciones[34]. Veremos más adelante que esto implica que una persona escoge apenas si está totalmente ejemplificada en todos los momentos del tiempo en que ella existe. Pero antes de deducir estas consecuencias generales, es necesario mostrar por qué el reduccionismo no puede representar, en los términos de su propia teoría, el poder de elección.

Afirmé anteriormente que lo que el reduccionismo puede representar es que el agente (o una tercera persona describiéndolo) imagine su futuro como un compuesto de posibilidades alternativas; que pueda ser atribuida una propiedad disyuntiva al agente, y que el agente (o una tercera persona describiéndolo) sepa que su presente es de tal forma, aunque pudiera haber sido de tal otra. Ser una

33. Esas acciones pueden ser descritas como "hacer A en t_3", "hacer B en t_3", "hacer A o B en t_3" o "ser aquel que hace A en t_3, pero podría haber hecho B en t_3". En gran medida, la elección entre esas formas depende de una demostración de la verdad o de la falsedad de la tesis determinista.

34. Sobre este punto, *cf.* Strawson (1992, pp. 134-135): al presentar una "fenomenología de la experiencia de la libertad", independientemente de decidir todavía si la libertad es una realidad metafísica, afirma que es innegable que, en lo atinente a la experiencia de la deliberación, "no somos meros espectadores de una escena en la cual –dejando de lado el elemento de cómputo, de cálculo– deseos en competición luchan por la victoria, teniéndonos por premio".

serie es compatible con todas esas descripciones. Lo que el reduccionismo no puede representar es la atribución de un poder de elección *a la serie*.

Atribuir un *poder* o capacidad a un individuo es afirmar que algo está en potencia en ese individuo; eso es compatible con ser una serie. Pero atribuir un *poder de elección* es afirmar que un individuo puede hacer verdadero o falso cierto estado de cosas. Eso es ir más lejos que decir que hay en él una "potencia para recibir contrarios": es decir que él, en la medida en que es aquel individuo, y no otro, puede hacer tal estado de cosas verdadero o falso. Pero hacer un estado de cosas verdadero o falso requiere que el individuo se piense a sí mismo como aquel que puede hacer A o B. O sea, la eficacia causal del agente depende de que éste se piense como el mismo haciendo A o B. No basta afirmar que ocurre un evento mental en la serie de la cual x es un estadio, evento ese que tendría por contenido: "x hace A o x hace B", pues eso capta el sentido en que algo *se hace* en la serie, pero no el sentido en que la serie *hace* algo. Es porque x *sabe* que es aquel que piensa en hacer A o B que tiene sentido afirmar que esta persona tiene el poder de hacer A o de hacer B. (Vimos anteriormente que imaginarse vagamente a sí mismo como haciendo A o B es *compatible* con el reduccionismo, pues esas imaginaciones son algunos de los eventos mentales –simultáneos o sucesivos– relacionados por la continuidad psicológica que, en ese modelo teórico, constituye una persona; vemos ahora que lo que es *incompatible* con el reduccionismo es una de las condiciones necesarias de esa imaginación de situaciones alternativas, a saber: la representación de sí mismo como el mismo, en muchas variaciones imaginativas que sean consideradas en el contexto de una deliberación práctica). Pero eso significa afirmar que x debe referirse a sí mismo como pudiendo hacer A o B, si es que puede hacer A o B. Ahora bien, esa referencia a sí es lo que *significa* el pronombre personal "yo". Pero el reduccionismo implica necesariamente, como vimos, que todo aquello que se dice en términos personales puede ser redescrito de forma impersonal. De otra forma, la tesis de la constitución ontológica de personas por eventos psicofísicos no tendría sentido. Si hay un caso en que tal redescripción se muestra imposible, entonces se sigue que el reduccionismo es falso.

Hay otra manera de presentar esa misma conclusión. Vimos antes que, al decir que y escoge B y al tener en cuenta que, en el mundo actual, $x = y$, estamos diciendo que atribuimos un poder de elección no solamente a y, sino también a x, o sea, al antecedente causal de y en t_1; o, incluso, es lo mismo que decir que x ya tenía un poder de elección sobre A y B, si es que le atribuimos una deliberación real. Pero ahora podemos precisar más esas afirmaciones: si x debe representarse a sí mismo para que su imaginación práctica sea parte de una deliberación efectiva, v.g., causalmente eficaz, entonces que $x = y$ no puede ser el resultado de

una relación entre elementos de los estadios temporales representados por *x* y *y*, sino una presuposición de la deliberación. No sería concebible que el individuo que se piensa a sí mismo como pudiendo hacer A o B en t_1 no fuera el mismo individuo que de hecho escoge A en vez de B en t_2, o sea, que ejerce poder de elección ya presente en *x* en t_1. (Esto vale no solamente para deliberaciones a "corto plazo", sino también para la identidad a "largo plazo", con proyectos de vida). La identidad, en vez de ser el resultado de una relación causal entre estadios de personas, está presupuesta en la deliberación si es que esta última es real. Si, por una parte, la conciencia de la identidad desempeña un papel causal, por otra es un mismo ser idéntico (o, más precisamente, la misma hecceidade) lo que debe permanecer en el tiempo para desempeñar esa función[35].

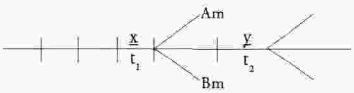
En este punto el reduccionista podría replicar: podemos redescribir cada pensamiento reflexivo expresado por el pronombre "yo" en términos impersonales. Como vimos, esa posibilidad de redescripción no es sólo una opción notacional, sino más bien una consecuencia necesaria de las tesis ontológicas reduccionistas. Incluso el "yo pienso, luego existo" cartesiano podría (y tal vez debería) ser redescrito como: "es pensado: el pensar está sucediendo". (*Cf.* Parfit, 1984, pp. 224 y ss.). Ese pensamiento sería al mismo tiempo subjetivo e impersonal (*ibid.*, p. 252); subjetivo porque sería inmediatamente auto-referente e inmune a errores, e impersonal porque no contendría ninguna referencia del sujeto a sí mismo como sujeto de ese pensamiento, a través del pronombre "yo". De modo general, diría el reduccionista, todos los indexicales como "aquí" y "ahora", así como el pronombre personal "yo", podrían ser explicados por el pronombre demostrativo "esto" y similares ("éste", "éstas", etc.); por ejemplo: con el uso de "ésta" y "en esta" que, en esta oración que el lector está leyendo ahora, se hace referencia a esta oración (*ibid.*, p. 252)[36]. De la misma forma, continuaría el reduccionista, sería posible describir decisiones de un cierto sistema de forma impersonal, y es por eso que podemos hablar propiamente de decisiones tomadas por organismos muy simples, que ciertamente no tienen ninguna "idea de sí mismos". Así, podríamos describir las tomas de decisión de un organismo más complejo en una forma también impersonal; aquel tendría conciencia de que existe un peligro enfrente y podría referirse a sí mismo de esta forma: "Si esto (refiriéndose a sí) doblara a la derecha, esto escapará del peligro; esto (refiriéndose a sí) está corriendo peligro y quiere salvar-

35. *En ese caso*, la *hecceidade* está necesariamente conectada con la identidad personal trans-temporal.

36. No sería, por tanto, necesario expresar el pensamiento expresado por la oración a la que esta nota se refiere como: "Yo, aquí y ahora, pronuncio la oración: .....".

se; entonces, esto (refiriéndose a sí) debe doblar a la derecha". He ahí un modelo reduccionista de deliberación, diría el humeano.

El reduccionista podría hacer más preciso ese modelo, mostrando cómo la atribución del poder de elección al agente en cuestión, en la imaginación práctica y requerida por la deliberación, podría ser explicada enteramente por su teoría, sin suponer que la identidad personal dependa de un hecho adicional separado de la serie que constituye una persona. Su argumento podría consistir en un comentario a la siguiente representación en diagrama de una persona:



El reduccionista podría argumentar que todo lo que él tendría que suponer para explicar este diagrama es que la serie que contiene *x* es compatible, considerándola hasta t_1, con la elección de A o de B. O sea, que, en t_1, *x* representa A y B como motivos posibles de una acción futura (lo que se señala en el diagrama con el índice 'm', que indica que se trata de la representación de A y de B como motivos que son candidatos a tornarse "causalmente eficaces" a través de la representación por *x* de A como mejor que B). El reduccionista también diría que la serie que contiene *y* es compatible con su secuencia como A o como B, o sea, es compatible con la atribución al sujeto de las acciones A y B.

En suma, lo que el reduccionista tendría que suponer sería apenas la verdad, *o bien* del indeterminismo, *o bien* del compatibilismo entre poder de elección y determinismo. Ahora, para suponer (y eventualmente probar) la verdad de una de esas tesis no es preciso, aparentemente, aceptar que una persona es una "entidad separada" o irreducible a sus elementos constituyentes. El reduccionista no tendría, así, que admitir que la serie es *la misma,* sea que contenga A o B (lo que es prohibido por la Ley de Leibniz), sino apenas que posibilidades alternativas son compatibles con la serie considerada hasta un cierto momento t (en el caso, hasta t_1 o hasta t_2). En otras palabras, las dificultades de suponer posibilidades alternativas no serían derivadas de la aceptación de la tesis reduccionista, sino que serían más bien dificultades inherentes a las tesis indeterminista y compatibilista –dificultades éstas también presentes en el caso que adoptemos un concepto anti-reduccionista de persona.

Volveré más adelante a la cuestión sobre organismos simples y sobre el sentido en que podemos decir que ellos "deciden" y "escogen" (ciertamente, hay un sentido propio para esas atribuciones). Por ahora, lo importante es observar que el reduccionista no recuperó, con su redescripción impersonal de la deliberación

(aunque admitiéramos que hay en ella algo de reflexivo y de subjetivo), la auto-referencia necesaria para la atribución del poder de elección. *x* no se limita a representar A y B como motivos posibles para *una* acción futura; se representa *a sí mismo* como haciendo A o B, o sea, representa A y B como *sus* acciones, como efectos consecuentes de *su* poder de elección, y es *por causa de esa auto-referencia* (que se manifiesta, entre otras cosas, en el hecho de que le pareció que A era lo mejor) que su deliberación es causalmente relevante para *y*. Por las mismas razones, si *y* escoge A en vez de B, no puede representar la elección como una orden del tipo: "Esto, haga A". Nosotros podemos ordenar que individuos hagan ciertas acciones, pero ordenar no es escoger (lo que se vuelve obvio por el hecho de que tengo que *escoger* la orden que daré, y que esa elección no está representada en la frase impersonal conteniendo el pronombre "esto").

La necesidad de la referencia a sí para la explicación de la deliberación es más evidente cuando la comparamos con otros estados que les atribuimos a las personas. Tomemos el caso de un deseo; los deseos son normalmente figurados reflexivamente en relación con un sujeto que desea algo, pero eso tal vez no sea una de sus condiciones necesarias; tal vez fuera posible simplemente *constatar*: "Deseo 1 surgió", "ahora, deseo 2". O sea, puede ser que sea una conexión contingente la que une el deseo a la representación de sí del agente que desea[37]. Además, muchos deseos involucran la representación del sujeto como *objeto* del deseo o como parte del *contenido* del pensamiento que lo expresa (el caso más obvio es el de "deseos de segundo orden", pero éstos están lejos de ser los únicos que involucran una representación del sujeto en ese sentido). Ahora bien, en el caso de la deliberación, ni la conexión con la representación de sí es contingente, ni esa representación del sujeto es apenas un contenido pensado. Aunque éste *también* sea un contenido, cuando me imagino en situaciones alternativas, *no se reduce* a un contenido; o mejor, el contenido en cuestión tiene que ser identifi-

37. Lo que es reforzado por la idea de deseos inconscientes. Más adelante, sin embargo, intentaré argumentar a favor de la tesis de que ciertos tipos de deseo no tienen una conexión meramente contingente con la conciencia de sí; lo mismo vale para la distinción entre dos tipos de elección, una consciente, otra inconsciente. Es interesante notar que la opción entre describir esa conexión como necesaria o contingente es, en mi opinión, lo que produjo problemas en la teoría de la identidad personal de Hume, señalados por él mismo en el Apéndice del *Tratado*: si esa reflexión fuera necesaria, tendríamos una especie de *hecho necesario*, lo que no tiene lugar en su teoría, pero si no fuera necesaria, entonces no habría cómo dar cuenta de la unidad de la conciencia (no desarrollaré aquí esa interpretación del texto, que, como se sabe, es un tema constante de debates entre los comentaristas).

cado por el sujeto como una representación de sí mismo todas las veces en que éste está embrollado en una deliberación. Sin esa referencia a sí, la deliberación y la elección no pueden ser figuradas, pues esa referencia es lo que explica el poder causal del agente en la producción de sus acciones.

La consecuencia inmediata de ese análisis es que *x* y *y* no pueden referirse a un sujeto interpretado de forma reduccionista. Si el reduccionismo fuera verdadero, el agente tendría que ser interpretado como no siendo nada más que una serie de eventos psicofísicos. Pero si el agente fuera reducible a esa serie, no sería posible pensar que *x* y *y* pudieran escoger A o B, o incluso que la persona que es la continuación de *x* escoja entre A o B. Un reduccionista radical podría decir que la única cosa que puede ser descrita es si estoy constituido por la serie A o por la serie B. Pero, si es así, el precio a pagar por tal teoría no es solamente admitir una teoría contra-intuitiva de identidad personal (algo con lo que el reduccionista está comprometido desde el principio), sino, además de eso, afirmar que la deliberación y la elección son apenas ilusiones. Con respecto al hecho de que el reduccionismo es una teoría neutra en lo que tiene que ver con la verdad del indeterminismo o del compatibilismo, esto es correcto, pero no elimina mi crítica, al contrario: apenas muestra que, en cualquier caso (en el caso de que la realidad sea correctamente descrita por la teoría indeterminista o por la teoría compatibilista), el reduccionismo no tiene un modelo adecuado para explicar la deliberación y la elección en ninguna de las teorías sobre la realidad. La teoría reduccionista no puede expresar las condiciones de la deliberación en sus propios términos, pues una persona, según aquélla, no es nada más que la serie en cuestión, es decir, es esa sucesión de propiedades (acciones y experiencias), y nada más. O sea, independientemente de la verdad del indeterminismo o del compatibilismo, la atribución de poder de elección al agente no puede ser explicada por el reduccionismo.

Luego es preciso no sólo que *x* y *y* crean ser los mismos que hacen A o B, sino que de hecho sean los mismos, si es que ellos deliberan efectivamente. O sea, no estoy afirmando únicamente, como Korsgaard, que el punto de vista autoral de un dado agente en relación con sus propias acciones es *pragmáticamente* ineliminable (*cf.* Korsgaard, 1996), sino, más bien, que la identidad real entre *x* y *y* está presupuesta en la deliberación y que esa identidad no puede ser expresada en términos impersonales. Korsgaard afirma que la "unidad pragmática" requerida por la deliberación es

> la unidad implícita en el *punto de vista* a partir del cual usted delibera y escoge. Puede ser que lo que realmente suceda cuando usted escoge es que lo más fuerte

> de sus deseos en conflicto gane. Pero no es ésta la manera en que usted piensa con respecto a eso cuando delibera. Cuando usted delibera, es como si hubiera algo sobre y por encima de todos sus deseos, algo que es *usted*, y que *escoge* a partir de cuál de ellos actuar [...] Eso no requiere que su capacidad de actuar esté localizada en una entidad que existe separadamente o involucre un hecho metafísico profundo. En vez de eso, se trata de una necesidad práctica que se le impone por la naturaleza del punto de vista deliberativo (*ibid.*, p. 370).

Siendo así, de la tesis de Korsgaard, así como de la tesis metafísica que defiendo, se sigue que no es posible formular una descripción "impersonal" de nuestras acciones. Esa consecuencia común tiene razones distintas, pragmáticas en un caso (Korsgaard) y metafísicas en otro (el mío); esa diferencia entre razones para justificar la misma tesis se explica por el hecho de que creo que debemos afirmar no solamente que un individuo que delibera se piensa como uno e idéntico, sino que esa referencia a sí es causa de efectos, es decir (asumiendo que un buen criterio para la atribución de existencia a algo es el reconocimiento de que ese algo tiene consecuencias causales), *existe*[38]. Ahora bien, como vimos, para el reduccionista, la descripción impersonal es necesariamente posible. Luego, este argumento presenta una crítica a un postulado central del reduccionismo.

Si esa crítica es correcta, una persona no es reducible a una serie de eventos más primitivos, o sea, es una entidad que no se confunde con los eventos psicofísicos que tiene. Pero, ¿la defensa del anti-reduccionismo implica afirmar que una persona es una "entidad separada" más allá de la serie de acciones y experiencias que tiene? ¿Hay alguna alternativa para explicar esa tesis, además de la suposición de que existe una sustancia cartesiana? Además, el argumento anti-reduccionista parece depender de la suposición de la verdad de una tesis metafísica que todavía no fue, ella misma, probada, a saber: que nosotros tenemos efectivamente un poder de elección. ¿Por qué habríamos de aceptar eso?

Comencemos por la última dificultad. Creo poder eliminarla mostrando que, en un cierto sentido, esa tesis no necesita ser "probada" para que mi argumento sea eficaz, pues negarla (al menos cuando esa negación asume una versión más radical) no tiene sentido.

De hecho, ¿qué significaría afirmar que no existe un poder de elección? Si eso significa afirmar que no existen deliberaciones genuinas, o sea, que ninguna deli-

38. Eso significa que el ser de una persona está constituido por esa auto-referencia, esto es, que su ser depende del ser percibido.

beración tiene eficacia causal[39], entonces esa tesis es, así me parece, simplemente absurda[40]. Dicho de otra forma: es posible afirmar que algunas de nuestras deliberaciones son ilusorias, racionalizaciones de determinaciones inconscientes, pero es imposible generalizar esa tesis para *todas* las deliberaciones. Y eso es absurdo porque, aunque el determinismo sea verdadero y el compatibilismo (entre libertad y determinismo) sea falso –lo que me parece ser también, por sí mismo, absurdo[41]–, aun en ese caso, deberemos afirmar que nuestras deliberaciones son causalmente eficaces, v.g., existen. Cualquiera que sea la verdad de la tesis metafísica compatibilista, es verdad que organismos vivos (en grados de complejidad variable) actúan según razones, v.g., escogen. Esas razones pueden ser externas o internas. Son externas cuando no es dada la representación de los motivos para el propio organismo del cual es verdadero afirmar que tales motivos son razones para actuar. Así, al explicar la acción de una ameba que absorbe una macromolécula que le sirve de alimento, debemos apelar a razones externas. Describir su comportamiento a partir de las razones que lo explican hace parte de nuestra idea de ese organismo, al cual le atribuimos las propiedades de ser sensible a estímulos de su medio ambiente y tener autocontrol, v.g., responder a esos estímulos a partir del modo como su estado interno fue afectado. Para animales superiores, esas razones no son meramente externas, como en el caso de las amebas, sino que son tales que el hecho de que el organismo se represente internamente para sí mismo esas razones explica al menos una gran parte de su eficacia causal[42]. Es importante notar que, entre los

39. Como ya señalé antes, estoy suponiendo aquí que "ser" implica no sólo "ser causado", sino también "ser causa de". Esa suposición es generalmente aceptada –por ejemplo, en filosofía de la mente, en las discusiones sobre prosecución.

40. Si eso significa una tesis más débil –por ejemplo, que no hay un poder de elección indeterminista, que escape a las leyes de la naturaleza–, entonces se trata de la discusión de la tesis compatibilista, cuya negación en nombre del indeterminismo puede hasta ser falsa, pero no es, estrictamente hablando, absurda.

41. Lo que me parece absurdo no es la verdad de la primera tesis y la falsedad de la segunda tesis tomadas separadamente, sino la verdad de su conjunción, pues esta última implica afirmar que la libertad es totalmente ilusoria, lo que me parece simplemente sin sentido. En este punto, como Aristóteles en *De Interpretatione*, 9, parto del hecho de que somos libres para evaluar los argumentos compatibilistas o incompatibilistas.

42. Sobre la idea de una evolución en el sistema de representaciones internas, que van desde la mera selección natural "ciega" hasta organismos que incorporan en sus previsiones artefactos producidos por ellos mismos, *cf.* Dennett (1995, Cap. 13, item 1, "The Role of Language in Intelligence", pp. 374-378), en el cual narra las transformaciones ocurridas desde lo que él denomina "criaturas darwinianas", pasando por las

ítems que son representados internamente por esos organismos superiores se encuentra la representación del propio organismo frente a situaciones alternativas. Esa referencia a sí hace parte del proceso de deliberación y explica su eficacia causal. Por tanto, por lo menos para animales superiores, es preciso garantizar que la representación de sí mismo como permaneciendo el mismo frente a situaciones alternativas existe sin duda, pues es precisamente esa representación lo que explica causalmente la existencia de los comportamientos animales. Otra manera de formular este punto sería afirmar, contrariamente a lo que fue sugerido antes, que no es posible que el determinismo sea verdadero y el compatibilismo falso, pues eso equivaldría a afirmar que no tenemos un poder de elección.

Esa referencia a sí se hace de un modo vago, como señalamos antes, de tal forma que no presupone estrictamente la representación perfecta de una identidad, sino apenas la imaginación de semejanzas[43], lo que es *suficiente* para explicar la eficacia causal de nuestras deliberaciones en el mundo. Deliberamos bajo la presión del tiempo y es *necesario* que sea así (*cf.* Dennett, 1995); luego, la "imagen de sí" que interviene causalmente en la generación de acciones es vaga, y no necesita suponer la representación cuidadosa de una identidad estricta. Una representación perfecta es por principio inaccesible desde el punto de vista de nuestras mentes finitas que, trabajando con un número limitado de informaciones, siempre proceden de forma aproximativa, y no desempeña (no *puede* desempeñar[44]) ningún papel causal relevante en la explicación de nuestros comportamientos. Aun así, debe haber una referencia a sí que sea al mismo tiempo incorregible y denote un hecho real, si es que hay deliberación real. Pero, como vimos, no tiene sentido decir que no hay deliberación real. Luego, en el caso de la deliberación ligada a la suposición de identidad personal que estuvimos examinando antes, podemos decir que

"criaturas skinnerianas" y "popperianas" hasta las "criaturas gregorianas". No discutiré aquí las implicaciones de la crítica de Dennett al "vocabulario intencional" en la descripción del comportamiento de organismos; basta apenas notar qué consideraciones sobre sistemas representativos serán utilizadas por él en su propuesta de compatibilización del determinismo con el poder de elección (*cf.* Dennett, 1984). Ciertamente, Dennett no concordaría con mis tesis sobre la identidad personal, pues, para él, una persona no es nada más que un "centro de gravedad narrativo", o sea, un resultado del punto de vista intencional.

43. Además, obviamente, de las otras características de la deliberación compatibles con el reduccionismo, como la noción de contrapartes (o, si hay identidad transmundana, la noción de hecceidade).

44. Si fuera una representación perfecta, no funcionaría para los propósitos de la deliberación de un ser que actúa en el tiempo.

el hecho de que *y* recuerde haber pensado[45] en A* y B* es una de las causas por las cuales él realiza A.

Hay todavía una cuestión central que permanece sin respuesta, en el caso que aceptemos la validez de mi crítica al reduccionismo, y que se expresa en las preguntas formuladas anteriormente: ¿Qué significa afirmar que una persona es irreducible a la serie de acciones y experiencias que tiene? ¿Hay alguna alternativa para explicar esa tesis, además de la suposición de que existe una sustancia cartesiana?

No podré responder a esa cuestión aquí, sino apenas indicar a partir de cuáles tesis podríamos pensar en alternativas para el reduccionismo que no nos comprometieran ni con el cartesianismo ni con el materialismo (los cuales, como señalé al iniciar este texto, son teorías poco plausibles para explicar la identidad personal). Todo lo que puedo decir en el momento es que, si el fenómeno de la deliberación no puede ser descrito en términos reduccionistas, esto no puede ser un hecho aislado, sino que debe ser un indicio o síntoma de un problema más general del reduccionismo, o, si queremos, de la verdad más general de la caracterización de la identidad personal a partir de la existencia de una entidad irreducible a la serie de deliberaciones que ella desarrolla y de acciones que ella realiza. Así, el sujeto de experiencias también debe ser caracterizado como irreducible a las experiencias que tiene. El conocimiento reflexivo del punto de vista por el cual experimentamos el mundo debe ser presupuesto en la descripción coherente de cualquier experiencia: en contraposición con lo que afirma el reduccionista, no sólo el significado de deícticos como "aquí" y "ahora" no es reducible a pronombres demostrativos como "esto", "éste", "en este", etc., sino que los deícticos presuponen que el sujeto que los enuncia haga una referencia a sí mismo como un "yo". Esa referencia a sí, siendo primitiva, implica también, creo, una tesis anti-perdurabilista, o sea, implica la tesis de que la totalidad de un individuo que tiene partes temporales está presente en todos los momentos en los cuales él existe.

Ese modo de existencia no tiene que ser el de existir como una sustancia simple, ni mucho menos como una especie de sustancia cartesiana, puramente mental y distinta de los procesos psíquicos (que serían sus modificaciones) y físicos (que serían de naturaleza diversa de la suya). Somos organismos complejos, seres vivos que deliberan teniendo a la vista finalidades relativas a nuestros organismos tomados como un todo. La referencia a ese todo constituye una parte

45. En lo que sigue, usaré el signo: * para designar las acciones A y B anticipadas por *x* en t_1 (del mismo modo que, antes, EA* designaba el tipo de experiencias y acciones que eran ejemplificadas por *y* como EA singulares).

importante de la explicación de varios fenómenos que ocurren con esos organismos; tenemos, así, una explicación perfectamente general y naturalista para la irreductibilidad de la personalidad: el tipo de entidad que una persona es tiene la característica especial según la cual su *ser* consiste, en parte, en *ser percibido*. Pero, al contrario de lo que puede parecer, no estoy defendiendo un idealismo o un fenomenalismo extremos; no estoy afirmando que una persona es una "apariencia". Lo que ocurre es que ciertos organismos complejos, compuestos de partes, piensan (y pensar es una propiedad prosiguiente a una base física, o sea, es una propiedad *reducible* a una base física), y, entre los organismos que piensan, algunos piensan en sí mismos como *una* totalidad organizada y, al pensar en eso, actúan de cierta forma. O sea, el hecho de representarse como unos es parte de la causa que produce aquellos efectos. Reducir esos fenómenos a sus elementos de base constituyentes (en última instancia, hasta las partículas físicas que los constituyen) significaría, como veremos, no sólo mudar de vocabulario, sino, incluso, no dar cuenta, de forma alguna, de aquella realidad. Siendo así, no se trata de afirmar, en ese caso, que ser es idéntico a ser percibido, sino que involucra ser percibido, lo que aleja esa afirmación de las tesis reduccionistas (o eliminativistas) según las cuales una persona sería apenas una manera (verbal) de referirse a ciertos organismos, pues lo que tal afirmación dice es que la eficacia causal depende de que el organismo en cuestión piense la totalidad que él es como *una* totalidad. Ahora bien, sólo lo real tiene efectos reales, o, dicho de otra forma, "ser" es "ser causa".

¿Qué significa entonces ser reduccionista? ¿El reduccionismo implica el eliminativismo? Tal vez un reduccionista podría afirmar que es reduccionista ontológico, y no semántico. O sea, la mención al todo compuesto sería necesaria para la verdad (y el sentido) de ciertos juicios, pero, de hecho, nada más que las partes organizadas existiría. El reduccionista podría entonces rechazar mi interpretación del estatuto ontológico de las personas como un organismo complejo que se refiere a sí mismo solamente como un todo. Según aquél, yo tendría que responder al menos las siguientes inquietudes: ¿Por qué un principio de sustitutividad del todo por la descripción de las partes organizadas que constituyen ese todo no mantendría la verdad (o incluso el sentido) de los juicios existenciales que contuvieran la mención a las partes o al todo? ¿Qué impediría la sustitutividad del todo por la descripción de sus partes organizadas? ¿La coreferencialidad no garantiza la eliminatividad, al menos en el plano ontológico?

Para responder estas preguntas es preciso discurrir brevemente sobre la relación entre parte y todo. Un objeto compuesto de partes tiene propiedades expresadas en un esquema conceptual que puede ser irreducible, en términos semánticos, a una descripción completa de la relación de las partes entre sí. O sea, hay ciertas

proposiciones verdaderas acerca del todo que no pueden ser expresadas cuando la referencia al todo es sustituida por una descripción de las partes y de sus relaciones (aunque admitamos que las relaciones en cuestión son las relaciones de constitución del todo), a pesar de que el nombre que designa el todo y la referida descripción de las partes se ocupen del *mismo* ente.

Tomemos como ejemplo la existencia de una montaña: puedo formular una explicación meteorológica que mencione la existencia de montañas, diciendo que una tormenta no alcanzó la ciudad porque una montaña cercana la detuvo. A pesar de que la montañas son compuestas de átomos y de moléculas, no puedo formular esa explicación meteorológica en términos de moléculas, átomos, partículas subatómicas, etc. (Probablemente este ejemplo es inadecuado, pues el meteorólogo también utiliza términos más refinados, pertenecientes a la teoría física, para referirse a entidades teóricas distintas de "nubes detenidas por montañas", que es como nosotros, en el sentido común, hacemos nuestras observaciones sobre el clima; el ejemplo, sin embargo, es intuitivo, y necesita sólo ser reformulado para adecuarse a la práctica científica más rigurosa)[46].

A pesar de que el cambio de significado de las expresiones referenciales ("montaña", "moléculas") modifica el contexto semántico de las proposiciones que las contienen, de tal forma que no tiene más sentido atribuir las mismas propiedades en los dos contextos, el reduccionista podría mantener su posición, afirmando que todo lo que existe y es expresado en el contexto semántico del todo (individuos, propiedades, relaciones, etc.) puede ser descrito en el vocabulario de las partes. O sea, ciertas proposiciones verdaderas no pueden ser expresadas en ese nuevo vocabulario, de tal forma que las propiedades (prosiguientes a las partes) que atribuimos al todo no pueden ser atribuidas a las partes, pero, aun así, todo lo que existe cuando formulo proposiciones sobre el todo y sus propiedades puede ser redescrito en el (reducido al) nuevo esquema conceptual de las partes (así, por ejemplo, las propiedades atribuidas al todo pueden ser reducidas a propiedades más elementales atribuidas a las partes).

46. La concepción contemporánea de átomos y de partículas subatómicas no nos permite decir propiamente que éstos son "partes" de objetos macroscópicos, como si fueran objetos muy pequeños unidos para formar un todo, pero con las mismas características de los objetos macroscópicos (como los ladrillos son partes del muro). Caracterizados como las cualidades de una cierta región del espacio, esas partículas componen los objetos, pero no son "partes" de ellos. Como quiera que sea, basta, para que mi argumento funcione, que pensemos vagamente en las partes como porciones mínimas de la materia que conservan ciertas propiedades del objeto mayor que ellas componen.

En el caso del poder de elección que atribuimos a las personas, sin embargo, tenemos buenas razones para ser anti-reduccionistas, o sea, para negar la posibilidad de expresar todo lo que existe en la actividad de deliberación en términos impersonales (series de eventos psicofísicos, estadios temporales de personas, etc.). Estoy asumiendo en este punto una premisa en relación con el reduccionismo, a saber: que el reduccionista con respecto a la identidad personal *debe* poder redescribir, en un vocabulario impersonal *pero psicológico*, las principales propiedades psicológicas que atribuimos a personas y que describimos en términos personales. Llamaré a esta afirmación presuposición psicológica. O sea, no se trata aquí de proponer un argumento contra el reduccionista eliminativista, que afirma poder reducir (o espera un día poder reducir) todos los predicados personales *y* los psicológicos a predicados físicos. El reduccionista en examen (del cual son ejemplos, Locke, Hume y Parfit) quiere redescribir, en su nuevo vocabulario, los principales eventos mentales que caracterizan a una persona en términos impersonales; ahora bien, la deliberación y la elección hacen parte de cualquier concepto mínimo de persona; entonces, deben ser posibles de una redescripción reduccionista, si es que el reduccionismo es verdadero.

Mi objetivo a lo largo del texto fue mostrar que, en el caso de la deliberación, tal reducción es imposible. La introducción de la relación parte-todo en el caso de las personas hace más claros los límites de la analogía con el caso de la previsión meteorológica que menciona "montañas": tanto montañas como personas son entidades compuestas de partes, y ambas son entidades que efectivamente existen (las proposiciones "existen montañas" y "existen personas" son verdaderas). Pero, en principio, todo lo que se dice acerca de las montañas y de sus propiedades (por ejemplo, la propiedad de "detener una tormenta") puede ser descrito en otro vocabulario, en el cual la proposición "existen montañas" es, no digamos refutada, sino simplemente superflua. O sea, el reduccionista acerca de montañas no tiene que defender la tesis de que "no existen montañas" (lo que quiera que eso signifique), sino apenas la de que es posible redescribir *toda* la realidad explicada por la tesis meteorológica en otro vocabulario. Ya en el caso de personas, si mi argumento es correcto, hay ciertas realidades (relativas a los eventos involucrados en la deliberación y en la elección) que no pueden ser redescritas en un vocabulario más primitivo; no se trata, para el reduccionista, de "optar" por describir persona en un vocabulario no-intencional (por ejemplo, en términos físicos, según los cuales, por definición, no se mencionarían los fenómenos intencionales), sino de no conseguir describir, en un vocabulario psicológico reduccionista, ciertos fenómenos que son, irrefutablemente, fenómenos psicológicos reales.

Pero alguien podría objetar que no deja de haber una analogía perfecta entre los casos aquí considerados de montañas y de personas; diría ese objetor: si yo *quiero* discurrir sobre una cierta realidad meteorológica, no puedo dejar por fuera la mención a "montañas"; ¿no sería una analogía perfecta con el concepto de persona decir que, si yo *quiero* discurrir sobre una cierta realidad psíquica, no puedo dejar por fuera la mención a "personas"? Al final, en los dos casos, estoy asumiendo que montañas y personas, fenómenos meteorológicos y psíquicos, son realidades efectivas, o sea, que hay proposiciones existenciales verdaderas sobre ellos. Pero hay un punto crucial que marca los límites de esa analogía: por la presuposición psicológica, el reduccionista *quiere* explicar ciertos fenómenos mentales *y* permanecer reduccionista; fue eso lo que mostré que es imposible[47].

Es posible reformular mi argumento, inicialmente presentado en términos epistémicos ("auto-referencia", "conciencia de sí", etc.), en términos ontológicos, o de la relación parte-todo, teniendo a la vista la pregunta sobre la naturaleza de las personas. Al tratar de la relación parte-todo, mencionamos cosas y propiedades de cosas. El poder de elección no es una propiedad de una cosa, sino una capacidad de ejercer un papel causal; según mi argumento, ni la capacidad, ni su ejercicio efectivo, pueden ser reducidos a términos más simples, que no hagan referencia al todo que es la persona. Es interesante notar en este punto que el poder de elección es expresado en términos de una *causalidad intencional*. La mera *causalidad* psíquica puede ser expresada en términos reduccionistas; por ejemplo, mi deseo inconsciente de x causó mi movimiento corporal en dirección a x. La mera intencionalidad también puede ser expresada en el vocabulario de la serie de eventos psíquicos; por ejemplo, el hecho de que varios pensamientos y sensaciones, referidos al mismo objeto, sean co-conscientes, puede ser explicado sin referencia a un "hecho irreducible" a la serie[48]. Siendo así, ¿por qué la *causalidad intencional* del poder de elección no puede ser explicada por el reduccionismo? ¿Qué hace que la conjunción de esas dos características de la deliberación (causalidad e intencionalidad) la haga inexpresable en términos reduccionistas?

47. Eso señala también los límites de mi crítica: ella alcanza apenas a aquel reduccionista que acepta la presuposición psicológica; no alcanza a un reduccionista eliminativista que pretendiera "traducir" todo el lenguaje psicológico a un vocabulario fisicalista, pues, en ese caso, podría simplemente ignorar la deliberación y la elección como fenómenos reales. Con todo, creo que es posible mostrar, por *otras* razones, que un reduccionismo eliminativista de ese tipo radical es insostenible; ciertamente, éste sería irrelevante si nuestro interés al discutir las tesis ontológicas reductivistas fuera un interés ético-existencial.

48. Sobre este último punto, *cf.* Parfit (1984, pp. 245-252).

La respuesta es evidente: se trata del hecho de que el ejercicio del poder de elección *y* los eventos que de él se siguen (y que son explicados por él) no son comprensibles a no ser como consecuencias de aquel todo complejo que *se considera* como haciendo A o B y que decide hacer A: sólo la *totalidad* de aquel proceso puede producir la elección de A. La auto-referencia al todo es necesaria en ese caso, pues *ser* una persona es pensarse como *una* persona.

Esto muestra que el dilema al que quieren condenarnos los reduccionistas es falso: *o bien* aceptamos la existencia de una sustancia cartesiana, *o bien* somos reduccionistas en relación con la identidad personal. ¿Qué afirma el reduccionista? ¿Que las personas son entidades compuestas? ¿Entonces, lo que niega es que las personas son entidades simples? Hay aquí una ambigüedad que permea la mayoría de las versiones más conocidas del reduccionismo: ora el reduccionismo se presenta como la tesis de que no somos una "entidad separada" de las partes que nos componen, ora como la tesis de que es posible (y deseable) explicar aquello que nos constituye como personas en términos impersonales[49]. El anti-reduccionista podría aceptar la primera tesis y rechazar la segunda: no precisa afirmar que somos una "entidad separada", distinta de las partes que nos componen. Somos una entidad compuesta, y no somos un "hecho adicional" a la totalidad de las partes organizadas según ciertas reglas, pero aun así hay ciertas capacidades que sólo existen cuando ese todo se refiere a sí mismo como *un* todo. Esas propiedades no "surgieron de la nada": emergieron de la correlación de las partes; pero, incluso en ese caso, la eliminación de la referencia al todo ya no permite capturar, en términos de propiedades más simples, lo que ocurre a ese nivel superior. No se trata más de decir que ciertos aspectos semánticos se perdieron; hay realidades que dejan de existir sin la (auto) referencia al todo.

Si mi crítica al reduccionismo es correcta, ha mostrado la imposibilidad de una descripción "impersonal" de nuestras acciones (o sea, mostró la necesidad de lo que Korsgaard llama punto de vista "autoral"). De estas consideraciones se sigue que debemos ser realistas en lo que tiene que ver con el concepto de persona: hay siempre una respuesta determinada para la pregunta acerca de si una acción o experiencia futura será o no será mía, gracias al realismo con respecto a la persona[50]. La conexión entre esas dos conclusiones es la siguiente: ya que la

49. Hume y Parfit parecen defender la tesis de que lo que realmente somos está mejor descrito en términos impersonales (el criterio de lo "mejor" es al mismo tiempo cognitivo y ético); Locke y Perry, por su vez, parecen más interesados en mostrar la génesis del ser de las personas a partir de elementos impersonales.

50. Como vimos, las tesis sobre la deliberación deben poder ser generalizadas para el caso de la atribución de experiencias.

perspectiva impersonal es falsa, el realismo es verdadero. Pero, si esto es así, Parfit está errado en cuanto a lo "que importa" en la supervivencia: aquello a lo que debemos dar valor no es solamente a la continuidad de nuestras experiencias (hasta el punto en que ellas tal vez, incluso, ya no sean *nuestras* experiencias), sino, más bien, al punto de vista único, irrepetible y factualmente incomunicable de un "yo" singular.

Bibliografía

Adams, Robert (1979): "Primitive Thisness and Primitive Identity", en: *The Journal of Philosophy*, Vol. LXXVI, No. 1, pp. 5-26.

Aristóteles (1984): *Catégories* y *De L'interprétation*. Trad. y notas de J. Tricot. Paris: Librairie Philosophique J. Vrin.

Chisholm, Roderick M. (1994): "On the Observability of the Self", en: Cassam, Quassim (ed.), *Self-Knowledge*, Oxford: Oxford University Press.

Dennett, Daniel (1995): *Darwin's Dangerous Idea. Evolution and the Meanings of Life*, New York, London: Simon & Schuster.

Dennett, Daniel (1984): *Elbow Room*, Cambridge, USA: The MIT Press.

Korsgaard, Christine (1996): "Personal Identity and the Unity of Agency: A Kantian Response to Parfit", en: *Creating the Kingdom of Ends*, Cambridge, USA: Cambridge University Press, pp. 363-397.

Hume, David (1978): *A Treatise on Human Nature*, ed. por L. A. Selby-Bigge, revisado por P. H. Nidditch. Oxford: Clarendon Press.

Leibniz, Gottfried W. (1988): *Discours de métaphysique et Correspondance avec Arnauld*, Georges le Roy (ed.), Paris: Librairie Philosophique J. Vrin.

Lewis, David (1995): *On the Plurality of Worlds*, Oxford, UK, Cambridge, USA: Blackwell, (1a. edición: 1986).

Locke, John (1975): *An Essay Concerning Human Understanding*, Peter H. Nidditch (ed.), Oxford: Oxford University Press,.

Merricks, Trenton (2000): "Perdurance and Psychological Continuity", en: *Philosophy and Phenomenological Research*, Vol. LXI, No. 1, pp. 195-198.

Moore, G. E. M. (1989): "Free Will", en: *Ethics*, Cap. VI. London: Williams and Norgate.

Parfit, Derek (1984): *Reasons and Persons*, Oxford: Clarendon Press.

Perry, John (1976): "The Importance of Being Identical", en: Rorty, Amélie O. (ed.), *The Identities of Persons*, Berkeley, Los Angeles, London: University of California Press, pp. 67-90.

Perry, John (1978): *A Dialogue on Personal Identity and Immortality*, Indianapolis: Hackett Publishing Company.

Perry, John (1975): "Personal Identity, Memory, and the Problem of Circularity", en: Perry, John (ed.), *Personal Identity*, Berkeley, Los Angeles, London: University of California Press, pp. 135-155.

Rea, Michael C. y Silver, David (2000): "Personal Identity and Psychological Continuity", en: *Philosophy and Phenomeno-logical Research*, Vol. LXI, No. 1, pp. 185-193.

Strawson, Peter (1992): "Freedom and Necessity", en: *Analysis and Metaphysics: An Introduction to Philosophy*, Oxford: Oxford University Press, pp. 133-142.