

(Versión para imprenta)

LA LÓGICA DE LA ACCIÓN: RAZONES Y CAUSAS*

Carlos J. Moya. Universidad de Valencia

La cuestión de la naturaleza y estructura de las razones y de su relación con la conducta incide de modo importante en diversos campos de reflexión filosófica. La respuesta que se dé a dicha cuestión condicionará inevitablemente las respuestas a problemas como la naturaleza de la acción y de la acción intencional, la relación de las ciencias humanas y sociales con las ciencias de la naturaleza, el lugar de los seres humanos en el mundo natural, la posibilidad del libre albedrío y de la responsabilidad moral, la racionalidad, la motivación y la interacción psicofísica. Permítaseme expresar la sospecha de que muchos de estos recalcitrantes problemas filosóficos podrían avanzar notablemente en el camino de su solución si, en lugar de ser abordados directamente, se llega a ellos dando un rodeo a través del problema, aparentemente secundario, de la naturaleza de las razones y su relación con nuestra conducta. En el presente trabajo nos proponemos, en primer lugar, examinar críticamente y analizar las dificultades a las que se enfrentan algunas de las respuestas contemporáneas más importantes a este problema, con especial referencia a las desarrolladas por Donald Davidson y Harry Frankfurt, y, en segundo lugar, defender, en conexión con dicho análisis, las bases de una alternativa inspirada en la concepción platónica de las razones, que presentamos, a grandes rasgos, en el primer apartado.

I. La concepción platónica de las razones.

La capacidad de actuar por razones ha sido considerada por distintos filósofos como un rasgo exclusivo de los seres humanos, que les concede un lugar de privilegio en el conjunto de la naturaleza y un valor y dignidad del que carecen otros seres. Ya en el *Fedón*, Platón establece, en boca de Sócrates, un claro contraste entre dos tipos de explicaciones. Si queremos saber por qué Sócrates se halla sentado en su celda, podemos apelar a la constitución anatómica de su cuerpo, a sus músculos, tendones y articulaciones que le permiten sentarse. Sin embargo, aun cuando estos factores sean necesarios para que Sócrates pueda llegar a estar sentado, no

constituyen las “causas de verdad”, las “verdaderas causas” de que lo esté, pues, como el propio Sócrates dice, “hace ya tiempo que estos tendones y estos huesos estarían en Mégara o en Beocia, arrastrados por la esperanza de lo mejor, si no hubiera creído que es más justo y más noble soportar la pena que la ciudad ordena, cualquiera que sea, antes que huir y desertar” (Platón, 1986, 107/98e-99a). Este pasaje de Platón tiene un gran interés. Si leemos “causas de verdad” como “razones”, Platón sostiene que una razón para actuar de un modo determinado puede ser (lo es, de hecho, en la situación que aquí se nos presenta) una creencia según la cual ese curso de acción es el mejor, o al menos mejor que sus *alternativas*. Una razón puede ser, según esto, una *creencia* con un contenido *valorativo*, una creencia sobre el valor de un determinado modo de actuar a la luz de determinados *criterios* de valoración, como la justicia o la belleza. Por otra parte, como queda claro en el contexto, Platón considera esta actuación por razones como la manifestación principal del poder causal de la mente y de su dominio sobre el cuerpo y la materia. Ambas tesis apuntan a cuestiones perfectamente vivas en la actualidad y representan tomas de posición substantivas en torno a ellas. Al sostener que las razones pueden consistir en creencias con contenido valorativo o normativo y que estas razones pueden explicar causalmente una acción, Platón se sitúa en un punto de vista contrario a la concepción humeana de las razones, según la cual una creencia, por sí misma, es causalmente inerte en la generación de la acción, de modo que una razón, si ha de poder explicar una acción, ha de incluir necesariamente un deseo (o una actitud análoga al deseo) independiente de la creencia en cuestión. Platón rechazaría resueltamente esta afirmación. Por otro lado, por lo que respecta al poder causal de la mente sobre el cuerpo, Platón sostendría, en casos como el que nos ocupa, que este poder causal deriva tanto de la *creencia* como de su *contenido*, es decir, de propiedades de carácter mental. Una explicación de carácter físico enuncia únicamente condiciones necesarias para que las causas mentales puedan operar y la acción se lleve a cabo, pero no condiciones suficientes de dicha acción. Aun cuando, para llevar a cabo la decisión de sentarse a esperar la pena impuesta por la ciudad, Sócrates necesite músculos, huesos y tendones (y, añadiríamos hoy, fibras nerviosas y cerebro), es su creencia de que eso es lo mejor y lo más justo lo que explica verdaderamente su decisión y su acción de sentarse y permanecer sentado. Si no hubiera tenido esa creencia, aun cuando todos los demás factores hubieran estado presentes, no estaría sentado en su celda, sino que habría escapado a Mégara o Beocia.

Claramente vemos que este tipo de explicación, aceptando que sea verdadera, concede al comportamiento de Sócrates tanto justificación racional como dignidad o valor moral. Estas dos cualidades no son independientes. Sólo de alguien capaz de ser sensible a razones y de actuar de un modo racionalmente justificado podemos predicar con sentido valor y dignidad moral.

II. Explicación y comprensión.

La justificación racional que las razones pueden conferir a nuestras acciones ha llevado a distintos autores a establecer una distinción tajante entre razones y causas, así como entre explicaciones por razones y explicaciones causales. Así, en polémica con el materialismo, Joseph Priestley escribe: “Cuando empiezo a moverme, lo hago por alguna razón y con respecto a algún fin. Pero, ¿quién puede imaginar a la materia moviéndose por argumentos, o a los silogismos y demostraciones clasificados entre las palancas y las poleas?” (Priestley, 1777, secc. 8, cit. en Toulmin, 1974, 20). La distinción entre las explicaciones por razones y las explicaciones nomológico-causales características de las ciencias de la naturaleza, expresada en términos de la distinción entre *Erklären* y *Verstehen*, entre explicación y comprensión, fue subrayada vigorosamente por la tradición decimonónica de las *Geisteswissenschaften*, representada por autores como Droysen, Dilthey o Windelband. Ya en el siglo XX, en polémica con la tesis de Hempel según la cual toda explicación científica tiene carácter nomológico-causal, William Dray sostenía que la explicación racional de la acción humana no pretende subsumir la acción bajo leyes, sino “mostrar que lo que se hizo era la cosa a hacer por las razones dadas, más que simplemente la cosa que suele hacerse en esas ocasiones, tal vez de acuerdo con ciertas leyes” (Dray, 1970, 124). La explicación por razones se caracteriza por la justificación racional que confiere a la acción. Nos presenta la acción como apropiada a la luz de las razones del agente, más que como una respuesta habitual o general al tipo de circunstancias en que dicho agente se hallaba, incluidas sus razones.

Paralelamente a esta tradición del *Verstehen* discurre la tradición del *Erklären*, representada por el empirismo británico y por el positivismo de los siglos XIX y XX, para la cual las explicaciones por razones poseen una estructura lógica sustancialmente idéntica a

las explicaciones causales de fenómenos físicos. Ambas formas de explicación descansan, explícita o implícitamente, en regularidades o leyes que conectan propiedades o clases de fenómenos. Así, en el contexto de la discusión del problema de la libertad y la necesidad, escribe Hume: “Pues, ¿qué se entiende por libertad cuando se aplica a acciones voluntarias? Desde luego no queremos decir que las acciones tienen tan poca conexión con motivos, inclinaciones y circunstancias que las unas no se siguen de los otros y que las unas no nos permiten inferir la existencia de los otros. Pues se trata de cuestiones de hecho manifiestas y reconocidas” (Hume, 1980, 119). Es decir, para Hume, nuestras explicaciones de la conducta voluntaria humana descansan en la constancia con que a determinados tipos de “motivos, inclinaciones y circunstancias” suceden determinados tipos de comportamiento, del mismo modo que nuestras explicaciones de fenómenos físicos descansan en la constancia con que fenómenos de determinados tipos se suceden los unos a los otros. Las explicaciones por razones son explicaciones causales como cualesquiera otras y se fundan en leyes psicológicas y psicofísicas que conectan tipos de fenómenos mentales entre sí y con tipos de acciones. Siendo causa y efecto, razón y acción son fenómenos dispares y distintos, “existencias separadas”, en términos de Hume, y sólo la constancia con que razones y acciones de determinados tipos se suceden unas a otras nos permite explicar una acción apelando a una razón como causa de aquélla. La línea humeana encuentra sus principales continuadores en el positivismo de Auguste Comte y John Stuart Mill, en el siglo XIX, y en el empirismo lógico del siglo XX. La formulación más rigurosa del concepto humeano de la explicación causal es, seguramente, el modelo nomológico-deductivo de la explicación científica desarrollado por Hempel (cf., p. ej., Hempel, 1979, 233-246), complementado posteriormente, por el mismo Hempel, con un modelo no deductivo que introduce en las premisas de la explicación leyes de carácter probabilista (cf. Hempel, 1966).

La discusión entre ambas tradiciones desemboca frecuentemente en un punto muerto. Los partidarios de la tradición interpretativa reprochan a sus adversarios la incapacidad de éstos para dar cuenta de la normatividad de las explicaciones por razones, de la justificación racional que confieren a la acción. Si estas explicaciones se fundan en último término en regularidades fácticas de coexistencia y sucesión, este fundamento no puede generar los rasgos normativos propios de la justificación racional. Esta última no sería, en todo caso, sino la forma ilusoria que adquiere nuestra familiaridad con la sucesión regular de razones y

acciones de determinados tipos. Por su parte, los partidarios de la tradición explicativa objetan a sus oponentes que, aun cuando, a la luz de determinadas razones y circunstancias, una forma de actuar sea razonable, ello no es todavía suficiente para explicar que el agente actúe de esa forma. Como Hempel señala en una crítica a Dray, del hecho de que un agente se halle en cierta situación, que incluye sus razones, y de que en esa situación sea apropiado hacer X, lo único que se sigue es que es apropiado que el agente haga X, no que de hecho lo haga (cf. Hempel, 1966). Para derivar esto último necesitamos precisamente regularidades o leyes sobre el modo en que las personas actúan, o es probable que actúen, en situaciones de ciertos tipos, incluyendo en ellas sus razones. Para Hempel, pues, Dray no consigue dar cuenta del carácter propiamente explicativo de las explicaciones mediante razones.

III. Justificación racional y 'conexión lógica': la escuela wittgensteiniana.

Lo que subyace a la insistencia de la tradición del *Verstehen* en la normatividad característica de las explicaciones mediante razones es la idea de que este tipo de explicaciones se hallan en un 'espacio lógico' distinto del que ocupan las explicaciones nomológico-causales. Esta idea recibió un impulso nuevo a partir de las importantes reflexiones de Wittgenstein sobre la conducta regida por reglas. En el marco de estas reflexiones, los conceptos de reglas y razones tendían a oponerse a los de leyes y causas. El autor que extrajo consecuencias más detalladas de la filosofía de Wittgenstein para el viejo debate entre *Verstehen* y *Erklären* fue Peter Winch (1972). Para este autor, entre los conceptos de ley y causa, por un lado, y los de regla y razón, por otro, no hay meramente una relación de oposición, sino también de dependencia: los últimos son más básicos que los primeros. La noción de ley involucra el concepto de 'lo mismo' y este último sólo es inteligible sobre la base del concepto de regla. Qué sea 'lo mismo' (el mismo tipo de propiedad o de suceso) en un contexto dado es algo que no puede establecerse salvo por relación a una regla (cf. Winch, 1972, 31). Una inspiración análoga podemos hallar en el intento de Georg Henrik von Wright de mostrar que el concepto de acción (intencional) es más básico que el de causa (humeana), en el sentido de que no podríamos tener este último si no fuésemos capaces de intervenir intencionalmente en el curso del acontecer (cf. Von Wright, 1979, cap. 2). Para Winch, la explicación mediante

razones no apunta a la predicción de la acción o a su subsunción bajo leyes, sino a su comprensión y justificación racional. Así, Winch escribe:

Supongamos que N, catedrático universitario, dice que va a suspender sus clases de la próxima semana porque piensa viajar a Londres: he aquí un enunciado de intención para el cual se da una razón. Pero N no *infiere* su intención de suspender las clases de su deseo de ir a Londres, como podría deducirse el inminente destrozo del cristal del hecho de que alguien tiró una piedra o de la fragilidad del mismo. N no da tal razón como una *evidencia* de la validez de esa predicción con respecto a su conducta futura... Por el contrario, está *justificando* su intención. Su enunciado no presenta esta forma: “Dado que tales y cuales factores causales están presentes, este será, por consiguiente, el resultado”, ni tampoco esta otra: “Dado que tengo tal disposición, el resultado será que yo haga esto”; su forma es: “En vista de tales y cuales consideraciones, lo que haré es algo razonable” (Winch, 1972, 78).

Para Winch, pues, una razón es sustancialmente distinta de una causa. La explicación de la acción o la intención mediante razones, a diferencia de la explicación causal de un suceso, es esencialmente una justificación racional de tal acción o intención. Una razón explica la acción correspondiente al presentarla como razonable, pero esa forma de explicación no es causal ni tiene como objetivo la predicción del comportamiento.

Sin embargo, vale la pena preguntarse por qué se supone que justificación racional y explicación causal son incompatibles, por qué aquello que justifica cierta forma de actuar no podría también ser causa de esa acción y servir para explicarla. En tiempos recientes algunos filósofos han acuñado, no sin cierta ironía, el término *hilefobia* para aludir al rechazo o temor visceral que la materia, y sobre todo el materialismo, despierta en determinados autores. Por nuestra parte, quisiéramos proponer el término *etiofobia* para referirnos a un rechazo o temor semejantes que ciertos autores sienten hacia la aplicación (o la mera aplicabilidad) del concepto de causa en el ámbito de la acción humana. Creemos que la etiofobia es detectable en muchos representantes de la tradición del *Verstehen*, incluido Winch. Es como si la presencia de la causalidad en el ámbito de la acción amenazase inexorablemente la libertad, la espontaneidad o la creatividad de los seres humanos, convirtiéndolos en engranajes inertes de un entramado puramente mecánico. Sin embargo, si estamos interesados en la libertad y en la creatividad, haremos bien en librarnos de la etiofobia y en tratar de fundamentar la posibilidad de que nuestras razones tengan influencia causal sobre nuestras decisiones y éstas

sobre nuestras acciones. De hecho, la etiofobia no forma parte de la concepción platónica de las razones: las razones son, para Platón, las verdaderas causas de nuestras acciones. La libertad, la espontaneidad y la creatividad de los seres humanos sólo se ven amenazadas si sus acciones tienen exclusivamente causas físicas, pero no si sus causas son mentales. La insistencia de Winch y otros pensadores, como Melden (1961) y Taylor (1964), en que las razones no causan nuestras acciones podría minar las intenciones humanistas de estos autores, puesto que, si están en lo cierto, o bien las acciones carecen de causas o bien sus causas son factores de carácter no racional. Y ambas posibilidades sitúan nuestras acciones más allá del control que podemos ejercer sobre ellas con nuestras deliberaciones y decisiones.

Si pretendemos defender la especificidad y el valor de los seres humanos, no constituye, en nuestra opinión, una buena estrategia rechazar el carácter causal de las razones. Nos inclinamos, más bien, por una defensa de la tesis platónica según la cual las razones son las “verdaderas causas” de nuestras acciones. Mostrar cómo es posible que lo sean, frente a los argumentos contrarios a dicha posibilidad, constituye una tarea filosófica de la mayor importancia en el contexto de una defensa del humanismo.

Sin embargo, no son únicamente razones de estrategia las que aconsejan rechazar la tesis de Winch y otros pensadores según la cual la noción de causa es ajena a la justificación racional de nuestras decisiones y acciones. En otros trabajos (Moya 1998a, 1998b) hemos defendido, por el contrario, que nuestro concepto cotidiano de justificación racional involucra el concepto de causa. Concretamente, una razón no puede justificar racionalmente una decisión (o una acción) a menos que la cause: la causalidad es una condición necesaria de la justificación racional. Pensemos en el ejemplo de Winch. Supongamos que, aunque N dice a sus alumnos que suspenderá sus clases de la próxima semana porque quiere viajar a Londres, en realidad no pretende hacer tal viaje, aun siendo cierto que hace tiempo que desea hacerlo, de modo que la verdadera explicación de que suspenda sus clases es, digamos, que se encuentra algo deprimido y quiere tomarse unos días de descanso en su casa. En este caso, la razón que ofrece a sus alumnos para suspender sus clases no justifica racionalmente su decisión, y no lo hace precisamente porque esa razón (su deseo de viajar a Londres) no causa su decisión. En realidad, la condición causal se requiere no sólo en el contexto del razonamiento práctico, sino también del razonamiento teórico. Supongamos que alguien extrae de ciertas premisas una conclusión que es de hecho correcta simplemente porque esta

conclusión le ha sido sugerida por otra persona, no porque él mismo haya considerado el contenido de las premisas y haya percibido que la conclusión se deriva del mismo. Su extracción de la conclusión, aun cuando ésta se deduzca de hecho de las premisas, no está justificada racionalmente, y de nuevo parece ser la ausencia de los antecedentes y nexos causales adecuados lo que da cuenta de esa falta de justificación racional.

La razón de fondo por la que Winch y otros pensadores wittgensteinianos rechazan el carácter causal de las razones y se cierran a sí mismos la vía platónica que aquí defendemos tiene que ver con la concepción de la mente desarrollada por Wittgenstein en polémica con el cartesianismo. Para Wittgenstein, entre los estados mentales y el comportamiento no existe una relación externa, sino una relación interna o constitutiva: determinadas formas de comportamiento, en particular el llamado comportamiento expresivo, no representan meros indicios empíricos de la presencia de un estado mental, sino criterios constitutivos del mismo. Así, el significado de los términos psicológicos (y los conceptos que expresan) incluye esencialmente la referencia al comportamiento que constituye el criterio de uso correcto de tales términos. El comportamiento fija el contenido experiencial (fenomenológico) e intencional de los estados mentales, de modo que este contenido no es independiente del comportamiento. En este marco, de claras afinidades con el conductismo lógico, la mente y el comportamiento no satisfacen la condición humeana según la cual causa y efecto son sucesos distintos e independientes, “existencias separadas”. Así, las razones, puesto que están constituidas por creencias, deseos y otros estados mentales, no pueden ser causas de las acciones.

Esta perspectiva general sobre la mente representa el verdadero fundamento filosófico del llamado “argumento de la conexión lógica” entre razones y acción, cuya conclusión es el rechazo del carácter causal de las razones. La estructura general del argumento es la siguiente. Como Hume mostró, causa y efecto son sucesos lógicos y conceptualmente independientes; pero las razones no son lógicas ni conceptualmente independientes de las acciones correspondientes; por lo tanto, las razones no son causas de las acciones. Un caso claro de esta dependencia lo constituye la explicación de una acción (por ejemplo, ir al cine) apelando al deseo de llevarla a cabo (el deseo de *ir al cine*). La razón, en este caso el deseo, contiene ya, como su objeto intencional, el concepto de la acción, por lo que no puede ser causa de ésta. Lo mismo podría decirse de las intenciones o de los pares deseo/creencia (cf.,

entre otros, Melden, 1961, 114; Taylor 1964, 33). En el marco de este rechazo del carácter causal de las explicaciones por razones, la interpretación y explicación de las acciones humanas se concibe como la progresiva elaboración de una imagen global del agente o agentes, a cuya luz tales acciones se presentan como razonables o racionalmente justificadas. Como Winch lo expresa, refiriéndose a la historia: “La explicación histórica no es la aplicación de generalizaciones y teorías a casos particulares: es la búsqueda de relaciones internas” (Winch, 1972, 123). Sin embargo, como Davidson sostuvo (Davidson, 1982, 9), a menos que pongamos en juego la idea de causa, es difícil ver, desde esta concepción, sobre qué bases podríamos decidir, ante dos conjuntos de razones que guardan “relaciones internas” igualmente coherentes con una acción y nos presentan dicha acción como igualmente razonable, cuál de los dos ofrece una explicación verdadera de la acción y (si, como hemos sostenido más arriba, la causalidad es una condición necesaria de la justificación racional) la justifica racionalmente. Una ilustración muy simple de esta consideración nos la proporciona la modificación que más arriba imaginamos del ejemplo de Winch. Tanto “suspenderé mis clases de la próxima semana porque quiero viajar a Londres” como “suspenderé mis clases de la próxima semana porque quiero tomarme un descanso” nos presentan la suspensión de las clases como razonable en alguna medida; en cierto sentido, la primera explicación muestra dicha suspensión como más razonable incluso que la segunda; pero, en el caso que construimos, no justifica racionalmente la acción precisamente porque el deseo allí mencionado no fue causalmente efectivo en la decisión de suspender las clases; sobre esta base, podemos decir que la primera explicación, a diferencia de la segunda, ofrece únicamente una 'racionalización', no una explicación verdadera de la acción.

IV. La síntesis de Davidson.

Una importante propuesta sobre la naturaleza de las razones y de su relación con la acción, en la que encontramos un intento de integrar y dar cuenta tanto del elemento de justificación racional propio de la tradición del *Verstehen* como del elemento explicativo-causal subrayado por la tradición del *Erklären*, la encontramos en los ensayos de Davidson sobre filosofía de la acción (Davidson, 1982). Trataremos de mostrar, sin embargo, que la propuesta presenta dificultades en ambos respectos y que, en último término, no es capaz de

fundamentar las dos ideas centrales de la concepción platónica de las razones, a saber, que éstas pueden consistir en creencias valorativas y que son las “verdaderas causas” de la acción.

Los ensayos de Davidson sobre la filosofía de la acción presentan al menos dos concepciones sucesivas de la estructura de las razones, aun cuando esta evolución no siempre es tenida suficientemente en cuenta en la bibliografía secundaria. En la primera de estas concepciones, que aparece en el renombrado artículo “Acciones, razones y causas” (Davidson, 1982, 3-19), se nos presenta una imagen muy sobria de la estructura de las razones y del razonamiento práctico. Las dificultades internas de esta primera concepción y su incapacidad para explicar algunos fenómenos conduce a Davidson a su segunda concepción, en la que el razonamiento práctico y las razones presentan una estructura más compleja. Sin embargo, como veremos, esta segunda concepción resulta excesivamente lastrada por el peso de la primera, que se mantiene en sus rasgos esenciales para ser únicamente complementada, en lugar de sustituida por aquélla.

Veamos ahora los rasgos fundamentales de lo que estamos llamando la primera concepción davidsoniana de las explicaciones por razones. Frente a la tradición del *Verstehen*, Davidson sostiene, como es sabido, que tales explicaciones son un tipo particular de explicaciones causales, que se distinguen de otras por la peculiar función de justificación racional de la acción que llevan a cabo. Esta función no es suficiente para afirmar que no se trata de explicaciones causales. Puede ser considerada simplemente como su diferencia específica dentro del género común de la explicación causal. Davidson reconoce, pues, la presencia del elemento de justificación racional en las explicaciones por razones, pero se niega a aceptar, frente a Winch o Dray, que esa característica obligue a concebirlas como explicaciones no causales. El modo en que Davidson da cuenta de ese elemento de justificación racional conlleva también el reconocimiento de una característica fuertemente subrayada por la tradición wittgensteiniana en su ataque al carácter causal de las razones, a saber, la existencia de una conexión 'lógica' o conceptual entre razón y acción. De nuevo, la estrategia davidsoniana consiste en negar que dicha característica nos obligue a aceptar que no estamos ante explicaciones causales y que las razones no puedan constituir causas de la acción. Según Davidson, la justificación racional que proporcionan las explicaciones por razones deriva de la estructura misma que una razón ha de tener para que pueda explicar con verdad una acción. Aunque, en su apariencia superficial, una razón puede presentar formas

muy diversas, sólo puede explicar realmente una acción si es o permite fácilmente construir lo que Davidson denomina una “razón primaria”. Y una razón primaria ha de consistir en la combinación de una actitud favorable (cuyo paradigma es el deseo) hacia acciones que tengan cierta propiedad y una creencia según la cual una determinada acción posee la propiedad en cuestión. En palabras de Davidson,

R es una razón primaria por la que un agente llevó a cabo la acción *A* bajo la descripción *d* sólo si *R* consiste en una actitud favorable del agente hacia acciones que tengan cierta propiedad y en una creencia del agente según la cual *A*, bajo la descripción *d*, tiene esa propiedad (Davidson, 1982, 5).

Esta estructura de las razones da cuenta, según Davidson, del elemento de justificación racional propio de las explicaciones por razones. A la luz de una razón primaria, la acción se nos presenta como apropiada o razonable, a saber, como un medio que el sujeto juzga adecuado para lograr algún objetivo que desea o considera atractivo. Advertamos que la razón desempeña este papel de justificación de la acción debido, en parte, a que incluye, en el contenido de la creencia, el concepto mismo de la acción para la que es una razón, como podemos comprobar en el texto que acabamos de citar. La justificación requiere, pues, una 'conexión lógica' entre razón y acción en este sentido preciso. Pero además, en esta primera concepción, la acción está conectada lógicamente con la razón en un sentido más fuerte, a saber, como la conclusión, lógicamente válida, de dos premisas, correspondientes a las dos actitudes del agente. Las premisas del argumento, un 'silogismo práctico', tendrían la siguiente forma: 1) Es deseable cualquier acción mía que tenga la propiedad P y 2) la acción A, así descrita, tiene la propiedad P. En realidad, lo que puede deducirse de estas premisas es el juicio siguiente: la acción A, así descrita, es deseable. Pero, para Davidson, siguiendo al propio Aristóteles, el agente extrae este juicio de modo práctico, a saber, llevando a cabo la acción. La acción llevada a cabo por razones corresponde, pues, a un juicio sobre la deseabilidad de dicha acción. Así, acerca de esta su primera concepción, el propio Davidson escribe:

Acepté la idea según la cual los contenidos proposicionales del deseo y de la creencia explicativos proporcionarían premisas de las que podría deducirse la deseabilidad de la acción.

En mi concepción, no se trataba de que el agente dedujera primero la consecuencia y luego actuase. En lugar de ello, abracé la idea de Aristóteles según la cual la extracción de la conclusión podía identificarse con la acción (Davidson, 1982, 195).

La estructura lógica de una explicación por razones es, pues, extremadamente simple: consiste en un argumento deductivo con dos premisas y una conclusión. La primera premisa corresponde a una actitud conativa del agente, normalmente un deseo, que cualquier acción de un determinado conjunto le permitirá satisfacer. De ahí su actitud favorable hacia ese conjunto de acciones. La segunda premisa corresponde a una actitud cognitiva, la creencia de que una determinada acción pertenece a ese conjunto. La conclusión es un juicio de deseabilidad que corresponde a la acción. Este es el modo en que Davidson integra en su concepción el elemento de justificación racional subrayado por la tradición del *Verstehen*.

Sin embargo, Davidson arguye que un sujeto puede tener una razón primaria para llevar a cabo una acción, y llevar a cabo esta acción, sin que esa sea la razón *por la cual* la llevó a cabo. Así, para que una razón explique una acción, además de constituir una razón primaria, ha de ser la razón por la cual el agente actuó. Y el modo más natural de expresar este requisito es exigir que la razón *cause* la acción. De ahí la segunda condición que una razón ha de cumplir para explicar una acción: “Una razón primaria de una acción es su causa” (Davidson, 1982, 12).

La introducción de este requisito causal, independiente del requisito de justificación, se debe, en nuestra opinión, a una concepción de la justificación racional, que Davidson comparte con la tradición del *Verstehen*, según la cual dicha justificación consiste únicamente en la existencia de relaciones lógicas apropiadas entre razones y acción y no requiere una relación causal entre ellas. Según hemos apuntado más arriba, este concepto de la justificación no corresponde a nuestro concepto cotidiano, que exige la influencia causal de la razón sobre la acción. Una vez aceptado este último, podemos decir que lo único que se requiere para que una razón explique una acción es que la justifique racionalmente, sin que sea necesario añadir un requisito causal independiente, ya que este requisito está ya incluido en la justificación como condición necesaria de la misma (cf. Moya, 1998b). En el caso de Davidson, sin embargo, la independencia entre la relación de justificación racional y la relación causal entre razones y acción, en unión con una concepción nomológica de la causalidad, le lleva a buscar el fundamento de la relación causal entre las razones y la acción

en un ámbito conceptual (de carácter físico o neurofisiológico) ajeno al contenido intencional de tales razones, que como tal acaba siendo causalmente inerte en la generación de la acción. Así, Davidson no consigue dar cuenta de la tesis platónica, que se halla también incorporada en nuestras intuiciones cotidianas, según la cual las razones como tales, en virtud de sus propiedades mentales, son las “verdaderas causas” de nuestras acciones. Y si el concepto causal de la justificación que hemos defendido es correcto, esta deficiencia acaba minando también, en el seno de la propuesta davidsoniana, la propia justificación que las razones aportan a las acciones humanas. En efecto, si no llevamos a cabo nuestras acciones *debido al* contenido de nuestras razones y *en virtud de* dicho contenido, es difícil ver en qué sentido están racionalmente justificadas: recordemos el caso, al que aludíamos más arriba, de una conclusión extraída con independencia de la consideración del contenido de las premisas. La justificación racional, subrayada por la tradición del *Verstehen* en su concepción de las explicaciones por razones, y su carácter explicativo-causal, destacado por la tradición del *Erklären*, resultan, pues, meramente yuxtapuestos, más que verdaderamente reconciliados, en la filosofía davidsoniana de la acción.

Pero la concepción davidsoniana de las explicaciones por razones resulta asimismo insatisfactoria porque no da cuenta adecuadamente del otro aspecto central de la propuesta platónica, a saber, que las razones pueden consistir en creencias valorativas. Trataremos de fundamentar esta afirmación en el próximo apartado.

V. Deseos y juicios de deseabilidad

Aparentemente, la afirmación que cierra el apartado anterior resulta injustificada, puesto que, como hemos visto, el componente conativo o desiderativo de la razón, en la primera concepción davidsoniana, puede entenderse como un juicio sobre la deseabilidad de cualquier acción que tenga cierta propiedad, y éste es un juicio valorativo. Asimismo, la acción corresponde a un juicio sobre la deseabilidad de una determinada acción. Pero es precisamente esta asimilación del deseo a un juicio de deseabilidad lo que resulta insatisfactorio, puesto que no permite distinguir entre deseos y juicios de deseabilidad ni asignar a estos últimos un papel distinto e independiente del deseo en la generación de la acción. Es en este sentido en el que la concepción davidsoniana de la acción constituye una

reformulación de la teoría humeana de la motivación, según la cual sólo un deseo, pero no una creencia, puede motivar a un agente a llevar a cabo una acción. Este es el fundamento de la afirmación de Hume según la cual la razón no puede ser sino la esclava de las pasiones. En la concepción davidsoniana, la creencia sólo motiva de modo subsidiario, dependiente del deseo, al reconocer en una acción la propiedad que la hace objeto del deseo o, lo que es lo mismo en esta concepción, del juicio de deseabilidad, actuando de este modo al servicio de la satisfacción del deseo.

En una primera aproximación, sin embargo, no es imposible desear llevar a cabo una acción determinada y juzgar o creer, con todo, que no es deseable (o apropiado, o bueno, etc.) llevarla a cabo, ni es tampoco imposible creer o juzgar que una acción es deseable y no tener, sin embargo, deseos de llevarla a cabo. La distinción entre deseos y creencias valorativas resulta crucial para que la acción humana pueda estar racionalmente justificada en un sentido que sobrepasa el marco de la primera concepción davidsoniana. Veamos ahora cómo se plantea esta cuestión en el pensamiento de Davidson.

La insuficiencia de la primera concepción davidsoniana de la explicación por razones se torna patente en casos de conflicto de deseos, e incluso en algunos casos en los que resulta plausible atribuir a un agente más de un deseo. En algunos de estos casos, la primera concepción davidsoniana conduce a contradicciones. Para verlo, apliquemos esta concepción al caso en que me planteo si ir o no al dentista. Presumiblemente, deseo mantener sana mi dentadura y creo que ir al dentista es un buen modo de hacerlo. En el marco de la primera concepción davidsoniana, mis premisas son las siguientes: 1) Considero deseable cualquier acción mía que tenga como consecuencia probable mantener sana mi dentadura, y 2) creo que ir al dentista tendrá probablemente esa consecuencia. Mi conclusión es: considero deseable ir al dentista, conclusión que extraigo prácticamente yendo al dentista. Sin embargo, también deseo evitar el dolor y creo que no ir al dentista es un modo de lograrlo. Mis premisas son ahora: 1) Considero deseable cualquier acción mía que tenga como consecuencia evitarme dolor y 2) creo que no ir al dentista me evitará dolor. Mi conclusión es: considero deseable no ir al dentista, conclusión que extraigo prácticamente no yendo al dentista. Ahora bien, uniendo ambas conclusiones llegamos a una contradicción, a una imposibilidad, a saber, que voy y no voy al dentista (cf. Moya, 1990, 146). La primera concepción davidsoniana de la explicación por razones encierra, pues, la posibilidad de contradicciones y no puede, por

lo tanto, ser correcta.

Davidson llegó a advertir la incorrección de su primera concepción al abordar el problema de la debilidad de la voluntad y de la acción incontinente. Actuar de modo incontinente es actuar en contra del propio mejor juicio. Si un sujeto desea fumar y lo hace, pero juzga que a pesar de todo sería mejor no fumar, actúa de modo incontinente. La primera concepción davidsoniana de las razones no puede dar cuenta de casos tan simples como éste, dando lugar a contradicciones análogas a la del ejemplo anterior. Así, Davidson se vio llevado a revisar dicha concepción. Veamos cómo describe él mismo esta revisión:

En “¿Cómo es posible la debilidad de la voluntad?” abandoné el modelo deductivo y modifiqué mi consideración del carácter lógico del contenido proposicional de un deseo. La premisa que tiene alguien que desea ver lo que hay en la página siguiente no es: “Cualquier acto mío que tenga una buena probabilidad de permitirme ver lo que hay en la página siguiente es deseable”, sino más bien algo como lo siguiente: “Cualquier acto mío, en la medida en que tenga una buena probabilidad de permitirme ver lo que hay en la página siguiente, es deseable”. Esta premisa se combina efectivamente con el contenido de una creencia según la cual dar vuelta a la hoja tiene la característica deseable, pero lo que se puede deducir es únicamente: “Dar vuelta a la hoja es deseable en la medida en que tiene una buena probabilidad de permitirme ver lo que hay en la página siguiente”. Semejante conclusión, argüí, no podría corresponder o ser identificada con la acción, puesto que es compatible con que el agente sepa que la acción (debido a otras características) es altamente indeseable. Las acciones, insistí, corresponden a juicios “incondicionales” (all-out) como “Dar vuelta a la hoja es deseable”, y tales juicios no se siguen en la lógica normal de las premisas proporcionadas por nuestros deseos y creencias (Davidson, 1982, 196).

Davidson sostiene, pues, que la acción requiere un juicio de deseabilidad incondicional, no relativizado a ciertas consideraciones. Pero al revisar su opinión sobre el contenido del deseo y la creencia del agente (sobre la estructura de una razón), advierte ahora que ese juicio (y con ello la acción) ya no puede derivarse válidamente de dicho contenido: un juicio incondicional no se sigue válidamente de uno condicional.

La nueva propuesta de Davidson, su segunda concepción de la explicación por razones y el razonamiento práctico, tiene, a grandes rasgos, la siguiente estructura. La premisa que corresponde al deseo es ahora un juicio de deseabilidad relativo a ciertos aspectos de la acción, del tipo “cualquier acción mía es deseable *en la medida en que* tenga

la propiedad P”, el cual es compatible con la consideración de determinadas acciones que tengan la propiedad P como no deseables debido a otras de sus propiedades. La premisa que corresponde a la creencia se mantiene como en el primer modelo. Tenemos ahora, en la matriz de la decisión, conjuntos de premisas que atribuyen deseabilidad relativa a acciones en función de determinadas propiedades, así como creencias sobre tipos de acciones que poseen tales propiedades. Estas premisas ya no pueden dar cuenta de la acción, pues de ellas sólo se deduce que determinadas acciones poseen una deseabilidad *prima facie*, relativa a alguna de sus propiedades. El juego de estas consideraciones a favor y en contra de determinadas formas de actuar se resuelve mediante la formación de un juicio de deseabilidad comparativo (*all-things-considered judgment*) según el cual, a la luz de todas estas consideraciones, una de estas formas de actuar resulta preferible a sus alternativas. Este es todavía un juicio condicional, relativo al conjunto de las consideraciones que el agente tiene en cuenta. Pero se requiere aún un juicio incondicional (*all-out judgment*), desligado de estas consideraciones, según el cual una determinada acción es, sin más, deseable, o más deseable que sus alternativas. Este juicio se identifica ahora, según Davidson, no directamente con la acción, sino con la intención o decisión de llevarla a cabo: “Un juicio incondicional según el cual cierta acción es deseable, o mejor, un juicio incondicional según el cual cierta acción es más deseable que cualquier alternativa al alcance del agente, no es distinto de la intención: es idéntico a ésta” (Davidson, 1982, 197). La intención, a su vez, si el mundo coopera, causa la acción.

El nuevo modelo ofrece recursos para dar cuenta, sin contradicción, de casos de conflicto de deseos y de casos de incontinencia o debilidad de la voluntad. Estos últimos son casos en los cuales el agente juzga que en conjunto A es preferible a B y, sin embargo, hace B y no A. El nuevo modelo davidsoniano puede dar cuenta de estos casos. El juicio según el cual A es preferible a B es un juicio de deseabilidad condicional, relativo al conjunto de las razones y consideraciones que el agente tiene en cuenta. Ahora bien, que el agente haga B se explica concibiendo esta acción como efecto de un juicio de deseabilidad incondicional, no relativo, según el cual B es deseable, sin más. Estos dos juicios no son contradictorios entre sí, dado que uno de ellos es condicional y el otro no, ni el modelo tiene como consecuencia que el agente lleva a cabo dos acciones incompatibles. Aun así, esta perspectiva no deja de atribuir al agente cierto grado de irracionalidad, como corresponde a nuestra opinión intuitiva

sobre estos casos. La irracionalidad deriva de que el juicio incondicional no concuerda con el juicio condicional, basado en la consideración, por parte de agente, del conjunto de sus razones, a diferencia de lo que constituye una acción continente: en ésta el sujeto juzga deseable y decide hacer aquello que juzga más deseable a tenor de todas sus razones.

En otro trabajo (cf. Moya, 1990, 148ss), pusimos en cuestión la concepción de la intención en Davidson como un juicio de deseabilidad. Lo que ahora quisiera plantear, sin embargo, es una cuestión distinta, a saber, si la introducción que Davidson lleva a cabo, en su segundo modelo de la explicación por razones y el razonamiento práctico, de dos nuevos tipos de juicios de deseabilidad, condicionales y no condicionales, entre la acción y las premisas correspondientes a las creencias y deseos, le llevan a conceder a los juicios o creencias valorativas un papel independiente de los deseos en la generación de la acción intencional. Dicho de otro modo, la cuestión es ahora si el segundo modelo davidsoniano representa un abandono de la concepción humeana de la motivación y la acción que informa su primer modelo.

El segundo modelo davidsoniano manifiesta un claro progreso con respecto al primero, al evitar el surgimiento de contradicciones y poder dar cuenta de casos que éste no podía tratar adecuadamente. Por otra parte, en lo que respecta a la cuestión que nos ocupa, el segundo modelo presenta importantes diferencias con respecto al primero. En este último, los juicios de deseabilidad no eran independientes de los deseos, sino que constituían, por decirlo así, deseos expresados de otro modo, como actitudes cognitivas, a fin de mostrar la relación entre las razones y la acción en forma de un silogismo práctico. En el segundo modelo, sin embargo, tanto los juicios de deseabilidad condicionales, relativos a todas las consideraciones que el agente tiene en cuenta, como los no condicionales parecen ser independientes de los deseos de los que parte el agente y constituir así fuentes autónomas de motivación. No hay, en la matriz de la decisión, deseos que correspondan a los nuevos juicios de deseabilidad que Davidson introduce. Sin embargo, cabe argüir que la autonomía de estos juicios con respecto a los deseos es sólo aparente.

Por lo que respecta a los juicios de deseabilidad condicional, relativos a todas las consideraciones que el agente tiene en cuenta, tales juicios representan, en el segundo modelo davidsoniano, un resultado del juego recíproco entre los deseos y las creencias (las razones, en términos davidsonianos) de los que parte el agente y son función de ellos. Esta afirmación

resultará más plausible si tomamos en consideración la teoría de la decisión racional y la comparamos con el modelo davidsoniano. En términos de esta teoría, que Davidson considera en todo momento una representación adecuada del razonamiento práctico y la toma racional de decisiones, la decisión racional vendría determinada por el resultado final de un cálculo llevado a cabo por el agente a partir de la utilidad esperada que, subjetivamente, asigna a distintas alternativas de acción posibles. Esta utilidad esperada es el producto de la deseabilidad asignada al resultado o resultados de cada forma de actuar y de la probabilidad asignada a dicho(s) resultado(s) dado ese curso de acción. El paso final de este cálculo es la asignación a un determinado curso de acción, entre aquellos que el agente considera abiertos para él, del grado más alto de utilidad esperada. Esta asignación determinaría la decisión racional de actuar de ese modo. Claramente, esta asignación es función de la utilidad esperada que el agente asigna a cada curso de acción y dicha utilidad esperada es función, a su vez, de los deseos y creencias del agente, a saber, de lo deseable que el agente considera el resultado o resultados de cada modo de actuar y de la probabilidad que asigna al logro de dicho(s) resultado(s) dada esa forma de actuar.

El fundamento filosófico de la teoría de la decisión es humeano. La deseabilidad de ciertos resultados para un agente es un dato último, un punto de partida sobre el que la razón no tiene jurisdicción. La función de la razón no es, y no puede ser, dictar fines o establecer valores, sino mostrar al agente el camino para alcanzar los fines y valores previamente establecidos por sus deseos. La razón carece de fuerza motivadora autónoma. Pues bien, no hay nada en la segunda concepción davidsoniana de la explicación por razones y del razonamiento práctico que ponga en cuestión de modo substancial la concepción humeana que inspira la teoría de la decisión. Los juicios de deseabilidad condicional, relativos al conjunto de consideraciones que el agente toma en cuenta, no son realmente independientes de los deseos de los que parte el agente y que constituyen (junto con sus creencias acerca de qué acciones poseen la propiedad objeto de deseo) sus razones iniciales en favor y en contra de determinados cursos de acción. El papel de estos juicios en el segundo modelo davidsoniano es estrictamente análogo al papel que en la teoría de la decisión desempeñan los juicios sobre el curso de acción que posee mayor utilidad esperada. Así como estos últimos son función de la utilidad esperada que el agente asigna a cada curso de acción, determinada a su vez por la deseabilidad de ciertos resultados y las creencias sobre su

probabilidad, aquéllos son asimismo función de los deseos iniciales del agente y de sus creencias sobre las acciones que le pueden permitir satisfacerlos.

Finalmente, por lo que respecta a los juicios de deseabilidad incondicionales, cuando concuerdan con los juicios condicionales, heredan de éstos su dependencia de los deseos del agente. Y cuando divergen de ellos, como en los casos de incontinencia, dependen de algún deseo o juicio de deseabilidad condicional presente en la matriz inicial, eliminando simplemente de él la cláusula de relatividad.

VI. La propuesta de Frankfurt: deseos de primer y de segundo orden.

Aun cuando el segundo modelo davidsoniano del razonamiento práctico carece del carácter deductivo presente en el primero, no representa, sin embargo, un cambio substancial por lo que respecta a la consideración de los deseos como los principales factores causales de la acción intencional. En ambos modelos, las premisas doxásticas, correspondientes a las creencias del agente, tienen un carácter meramente instrumental con respecto a las premisas correspondientes a los deseos: indican simplemente el modo en que éstos pueden ser satisfechos, al identificar las acciones que poseen la propiedad objeto de deseo. La creencia de que determinadas acciones poseen cierta propiedad es por sí misma causalmente inerte, a menos que esa propiedad sea objeto de deseo por parte del agente. En ambos modelos, la razón sigue siendo esclava de las pasiones, como lo era para Hume.

En un contexto algo distinto, aunque relacionado, a saber, la discusión de los problemas de la libertad y la responsabilidad moral, Harry Frankfurt argumentó que la concepción de la libertad en el compatibilismo clásico de Hume, Mill o Ayer, según la cual actuar libremente es actuar de acuerdo con la propia decisión, determinada a su vez por los propios deseos y motivos, capta tal vez el concepto de libertad de acción, pero no el de libertad de la voluntad o libre albedrío (cf. Frankfurt, 1982). La propuesta del compatibilismo clásico constituye una concepción minimalista de la libertad, según la cual ser libre es ser capaz de hacer lo que uno quiere y actuar libremente es ejercer esa capacidad, haciendo

efectivamente lo que uno quiere. Si entendemos la libertad en estos términos, no podemos explicar por qué atribuimos a las personas adultas normales, pero no a los niños pequeños, libre albedrío y responsabilidad moral, ya que también los niños pueden actuar, y actúan a veces, según quieren y deciden hacerlo. En nuestra opinión, la concepción davidsoniana de la explicación por razones y el razonamiento práctico tiene como consecuencia una perspectiva sobre la naturaleza de la libertad y la responsabilidad moral muy cercana a la del compatibilismo clásico. De hecho, en el artículo “Freedom to Act” (Davidson, 1982, 63ss.), Davidson nos presenta asimismo una concepción minimalista de la libertad que equipara la acción libre con la acción intencional, entendida ésta en términos de su concepción de la explicación por razones. Una acción intencional es aquella que está racionalmente justificada, y causada del modo apropiado, por las razones del agente, constituidas, como sabemos, por deseos y creencias. La distinción entre los dos modelos davidsonianos no da lugar en este punto a diferencias decisivas. Esta concepción de la libertad está expuesta, como la de Hume o Ayer, a las críticas de Frankfurt.

La propuesta de Frankfurt para remediar las dificultades del compatibilismo clásico y dar cuenta del concepto de libre albedrío consiste en distinguir, junto a los deseos y juicios de deseabilidad que tienen como objeto acciones, y que él denomina “deseos de primer orden”, deseos y juicios de deseabilidad de segundo orden, cuyo objeto son deseos de primer orden. El libre albedrío consiste para Frankfurt en la capacidad de actuar por deseos que el agente juzga a su vez deseables como motivos de su acción, por deseos que el agente desea que le muevan a actuar. Del mismo modo que un sujeto actúa libremente cuando hace lo que desea hacer, actúa con voluntad libre cuando actúa por un deseo que él desea que le mueva a actuar. La carencia de deseos de segundo orden hace inaplicable el concepto de libre albedrío, según la caracterización que del mismo ofrece Frankfurt, a determinadas criaturas. Así, la apelación a este nivel de actitudes conativas reflexivas permite a Frankfurt distinguir entre personas adultas y niños pequeños por lo que respecta al libre albedrío y la responsabilidad moral.

A la luz de la propuesta de Frankfurt, el modelo davidsoniano se nos presenta claramente como un modelo compatibilista clásico: todos los deseos y juicios de deseabilidad que aparecen en él son, en la terminología frankfurtiana, deseos y juicios de deseabilidad de primer orden, puesto que su objeto son siempre acciones. En este aspecto, pues, no sobrepasa

los límites e insuficiencias que Frankfurt descubre en el compatibilismo clásico de corte humeano. En relación con ello, es interesante destacar una diferencia adicional entre los modelos de Frankfurt y Davidson: mientras que en éste el juicio por el que el sujeto atribuye a un modo de actuar mayor deseabilidad que a sus alternativas es función de las razones del agente, y en especial de sus juicios de deseabilidad *prima facie*, en el caso de Frankfurt los deseos o juicios de deseabilidad de segundo orden, siendo actitudes reflexivas sobre los deseos de primer orden, no son necesariamente función de éstos, sino que tienden a expresar más bien el ideal personal y los valores que el sujeto mantiene, la clase de persona que quiere ser.

Aun así, es dudoso que la propuesta de Frankfurt rebase realmente los límites de la concepción humeana, reconociendo fuentes de motivación y de finalidad distintas del deseo. Como distintos intérpretes han subrayado, los deseos de segundo orden son, a fin de cuentas, deseos y, del mismo modo que un sujeto se encuentra a sí mismo deseando llevar a cabo ciertas acciones y no otras, se encuentra también a sí mismo deseando ser movido a actuar por unos deseos y no por otros, y encontrando atractivos ciertos valores e ideales personales y no otros. En principio, una acción causada por deseos de segundo orden no tiene por qué ser más libre o autónoma, y tampoco más racional, que una acción causada por deseos de primer orden. De hecho, no consideramos más libre o autónomo, o más racional, a un puritano que rechaza determinados tipos de deseos como motivos de su acción y actúa según ese rechazo que a un sujeto que aprueba esta clase de deseos y actúa según ellos. Hay deseos de segundo orden que consideramos irracionales, como el deseo de una persona anoréxica de no tener el deseo de comer. Juzgamos este rechazo del deseo de comer, y seguramente también el rechazo del puritano de sus deseos naturales, como erróneo y patológico. Lo que estos juicios muestran es que el mero hecho de que un deseo tenga como objeto, bien una acción, bien un deseo, siendo así de primer o de segundo orden, no es el factor decisivo para la libertad y autonomía de un sujeto, y tampoco para la racionalidad de sus decisiones y acciones. Lo que echamos en falta en estos casos no son actitudes reflexivas conativas, que en ambos casos están presentes, sino la justificación y racionalidad de las mismas. Pero la pregunta por dicha justificación y racionalidad no es legítima para la concepción humeana de la motivación, que inspira tanto la propuesta de Davidson como la de Frankfurt. En esta concepción, los deseos son la fuente última de las razones, de modo que no hay razones para

tener o no tener deseos que no deriven a su vez de otros deseos.

Sin embargo, una persona puede ser víctima de sus deseos de segundo orden no menos que de sus deseos de primer orden, y puede tener deseos irracionales de ambos tipos, como vemos en los ejemplos del anoréxico o el puritano. Las propuestas de Davidson y Frankfurt no permiten explicar nuestros juicios sobre estos casos, pues no permiten la posibilidad de juicios sobre la racionalidad de los deseos mismos excepto sobre la base de otros deseos. Así, en la concepción davidsoniana, un deseo puede ser irracional únicamente en un sentido instrumental: si un sujeto desea realmente lograr algo, y cree que un modo de actuar es el mejor medio de conseguirlo, es irracional por su parte no desear actuar de ese modo. En la concepción de Frankfurt, un deseo puede ser irracional únicamente a la luz de un deseo de orden superior. Un modo natural de dar cuenta de nuestro juicio sobre el anoréxico o el puritano consiste en decir que sus creencias valorativas, que dan lugar a sus deseos de segundo orden, son erróneas o, sencillamente, falsas. La falsedad de estas creencias da cuenta de la irracionalidad de dichos deseos. Pero este modo de explicar nuestro juicio no tiene cabida en la concepción humeana de las razones, puesto que presupone, en primer lugar, en contra de la posición sostenida por Hume, que es legítimo juzgar, de modo sustantivo y no meramente instrumental, sobre la racionalidad de los deseos mismos y, en segundo lugar, que las creencias (valorativas) pueden constituir fuentes de motivación, bien por sí mismas, bien a través de deseos que generan en el sujeto.

VII. Razones y creencias valorativas.

Volvamos ahora, desde la perspectiva que hemos alcanzado, al texto del *Fedón* platónico que citamos al comienzo de este trabajo. La situación y la decisión de Sócrates que nos presenta Platón recibe, en nuestra opinión, una representación incorrecta si es interpretada desde la concepción davidsoniana del razonamiento práctico. Si atribuimos a Sócrates, como parece plausible hacerlo, un deseo natural de conservar su vida, lo que tenemos aquí es un conflicto de razones y, en particular, de deseos. Sócrates juzga deseable conservar su vida y cree que escapar a Megara o Beocia es un buen modo de lograrlo. Al mismo tiempo, juzga deseable preservar las leyes de la ciudad y cree que permanecer en prisión, arriesgándose a ser condenado a muerte, es un modo de satisfacer ese deseo. Pero escapar y permanecer en

prisión son incompatibles. Como resultado de sopesar ambas razones, Sócrates forma un juicio según el cual, a la luz de todas esas razones, es más deseable permanecer en prisión que escapar y, de modo continente, juzga deseable sin más permanecer en prisión, formando así la intención de hacerlo y actuando en consecuencia. Lo que esta representación de la situación oscurece crucialmente es el distinto carácter, origen y justificación de las razones que Sócrates considera para tomar su decisión, razones que aquí se nos presentan meramente como dos conjuntos de deseos (juicios de deseabilidad) y creencias instrumentales en conflicto. La explicación que da el propio Sócrates de su decisión apela a su *creencia valorativa*, no instrumental, según la cual “es más justo y más noble soportar la pena que la ciudad ordena, cualquiera que sea, antes que huir y desertar”. La razón que ofrece Sócrates como lo *verdadera causa* de su decisión tiene un origen, una naturaleza y una justificación distinta del deseo de preservar su vida, deseo que, sin duda, también tiene. Se trata de una creencia valorativa cuya génesis y justificación no es meramente fáctica, sino normativa y racional; no es un deseo con el que de hecho le sucede hallarse, ni un mero juicio sobre lo que de hecho le parece deseable. Esta distinción queda completamente en la sombra en la representación que se deriva de la concepción davidsoniana. Y algo muy semejante cabría decir de la interpretación que se desprendería de la teoría de la decisión.

Por lo que respecta a la propuesta de Frankfurt, la representación de la situación sería la siguiente. Sócrates tiene, de hecho, dos deseos en conflicto: desea escapar y también desea permanecer en prisión. Pero Sócrates es la clase de hombre al que, en las circunstancias del caso, le gustaría ser movido por deseos del segundo tipo, y no del primero, de modo que, además de estos dos deseos de primer orden, tiene también un deseo de segundo orden: Sócrates desea que sea el segundo deseo, y no el primero, el que le mueva a actuar. Y este deseo de segundo orden causa su acción, que es así expresión de su voluntad libre. En otro sentido, la representación de la situación desde la propuesta frankfurtiana es también insatisfactoria, y por razones muy semejantes a la de Davidson: tampoco en la propuesta de Frankfurt encontramos el reconocimiento de una fuente de razones y motivación distinta del deseo, de modo que, en este marco, la creencia de Sócrates según la cual lo más justo es quedarse en prisión sólo explicaría realmente su acción si fuese una expresión alternativa, en forma doxástica, del deseo de segundo orden indicado. De nuevo, la distinta naturaleza de los deseos, incluso de segundo orden, y la creencia valorativa que Sócrates ofrece como

explicación de su acción queda aquí completamente oscurecida.

Ninguna de las propuestas que hemos considerado permite la posibilidad de que el agente establezca activamente sus propios fines y motivos mediante la reflexión racional, con independencia de los deseos con los que de hecho se encuentra, y ninguna de ellas permite que la justificación que el agente puede dar de su acción termine en algo distinto de un deseo o una actitud conativa que de hecho tiene. Por ello no pueden dar cuenta adecuadamente del caso de Sócrates ni conceder a su decisión el valor que vemos que realmente tiene. En el marco de las propuestas indicadas, en último término ninguna acción puede tener mayor valor que otra, puesto que, al final, su origen y justificación no puede ser sino un deseo, un estado pasivo y no racional del sujeto. Puede asombrarnos que, en ciertos casos, un deseo predomine sobre otro, pero nada más.

Lo que distingue la creencia valorativa de Sócrates, que constituye, en la situación descrita por Platón, la verdadera causa de su conducta, es, al menos en parte, la justificación que podría ofrecer de su verdad si fuese retado a hacerlo. Sócrates podría ofrecer, en favor de su creencia, argumentos y consideraciones racionales, arraigadas en un proceso personal de formación reflexiva de sí mismo basado en la búsqueda racional de la verdad, en especial acerca de cuestiones normativas. De hecho, si consideramos los diálogos socráticos como una exposición fiel del pensamiento de Sócrates, partes importantes de estos diálogos estarían conectadas con la argumentación que Sócrates podría ofrecer en defensa de su creencia y de su decisión. No vemos la decisión y la actuación de Sócrates como la obra de un fanático doctrinario, sino como la decisión racional de un hombre libre. Y esto tiene que ver con la formación de la creencia causante de su decisión a través de su propia actividad racional y con la consiguiente justificación que podría ofrecer de su verdad. La formación de creencias valorativas sobre lo que es más o menos importante, deseable o valioso no tiene por qué ser el resultado de un cálculo a partir de valores y deseos que un sujeto encuentra pasivamente en sí mismo; puede ser el producto de un arte vital y racional capaz de crear y configurar activamente los propios fines y valores. Esto es al menos parte de lo que distingue a un hombre libre como Sócrates de, por tomar un hecho histórico reciente, los fanáticos que estrellaron el avión que pilotaban contra las Torres Gemelas de Nueva York. No vemos su acción como expresión de libertad y racionalidad. Y esto tiene que ver con el hecho de que, o mucho nos equivocamos, o la justificación de su decisión descansaría pronto en creencias

incapaces de soportar un escrutinio y una argumentación racional en favor de su verdad. Parte también de la diferencia, no desconectada de la anterior, es que la decisión de Sócrates sólo culminó en su propia muerte, sin arrastrar a otros consigo a la tumba.

VIII. Resumen y conclusiones.

La concepción platónica de las razones ha sido el punto de partida y el hilo conductor del presente trabajo. De acuerdo con esta concepción, las razones para actuar pueden consistir en creencias valorativas, independientes de los deseos fácticos de un sujeto. Las razones, por otra parte, constituyen, en cuanto tales, las “verdaderas causas” de las acciones. Hemos tratado de defender estas dos tesis platónicas frente a otras perspectivas, pues estamos convencidos —aun cuando una fundamentación detallada de esta convicción sobrepasaría los límites del presente trabajo— de que sólo esa concepción de las razones puede fundamentar el humanismo y dar cuenta de la especificidad y el valor de los seres humanos, de la posibilidad de la libertad y la responsabilidad moral.

La tradición del *Verstehen*, en la que incluimos la escuela wittgensteiniana, subraya la función normativa, de justificación, que desempeñan las razones, en detrimento de su papel causal, entrando así en conflicto con la segunda tesis platónica. Sin embargo, si se pretende —como lo pretende esta tradición— defender el humanismo, lo importante no es abstraer la conducta humana a la causalidad, sino mostrar que puede estar causada por razones en sentido platónico.

La tradición del *Erklären*, representada emblemáticamente por Hume, insiste en cambio en el carácter causal de las razones y de la explicación que proporcionan de las acciones humanas, pero rechaza la posibilidad de que las razones consistan en creencias valorativas, entrando así en conflicto con la primera tesis platónica. De acuerdo con la concepción humeana de las razones, sólo los deseos pueden causar realmente la acción.

La filosofía de la acción de Donald Davidson y su concepción de las razones constituye un intento de síntesis y mediación entre ambas tradiciones, aunando el aspecto normativo de justificación racional subrayado por la primera y la función explicativo-causal de las razones en la que insiste la segunda. No obstante, en nuestra opinión, el intento resulta

finalmente fallido en varios aspectos. En primer lugar, la perspectiva davidsoniana no rebasa realmente los límites de la concepción humeana de las razones y entra, como ésta, en conflicto con la primera tesis platónica. Y, en segundo lugar, no consigue mostrar que las razones, en cuanto tales, puedan constituir las “verdaderas causas”, la explicación verdadera de la acción, entrando así en conflicto con la segunda de las tesis platónicas.

Examinamos a continuación el modo en que Harry Frankfurt trata de remediar los defectos de la concepción humeana de las razones, aplicada a la comprensión de la libertad, apelando a la importancia de los deseos de segundo orden en la generación de la acción. El resultado de nuestro examen ha sido que la propuesta de Frankfurt no constituye tampoco una alternativa real a la concepción humeana y no ofrece un fundamento suficiente de la libertad y el valor de los seres humanos.

Una concepción satisfactoria de las razones y el razonamiento práctico ha de poder dar cuenta de los juicios intuitivamente correctos acerca de determinados casos, como el que tendemos a dar acerca de la decisión de Sócrates, o acerca de ejemplos como el del anoréxico o el puritano. También debe fundamentar la posibilidad del propio desarrollo, libertad y felicidad. En ambos aspectos, la concepción platónica, aun necesitada de perfeccionamiento, nos ha parecido claramente superior a las concepciones rivales que hemos tenido en cuenta.

REFERENCIAS BIBLIOGRAFICAS

Davidson, D. (1982), *Essays on Actions and Events*, Clarendon Press, Oxford.

Dray, W. (1970), *Laws and Explanation in History*, Clarendon Press, Oxford.

Frankfurt, H. G. (1982), “Freedom of the Will and the Concept of a Person”, en G. Watson (ed.), *Free Will*, Oxford University Press, Oxford, 81-95.

Heil, J. y Mele, A. (eds.) (1993), *Mental Causation*, Clarendon Press, Oxford.

Hempel, C. G. (1966), “Explanation in Science and in History”, en W. Dray (ed.),

Philosophical Analysis and History, Harper & Row, N. York, 95-126.

Kim, J. (1993), "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?", en Heil y Mele (eds.), 1993, 19-26.

Liz, M. (1995), "Causalidad y contenido mental", en F. Broncano (ed.), *La mente humana*, Trotta, Madrid, 207-243.

McLaughlin, B. P. (1993), "On Davidson's Response to the Charge of Epiphenomenalism", en Heil y Mele (eds.), 1993, 27-40.

Melden, A. I. (1961), *Free Action*, Routledge & Kegan Paul, Londres.

Moya, C. J. (1990), *The Philosophy of Action*, Polity Press, Cambridge.

Moya, C. J. (1998a), "Reason and Causation in Davidson's Theory of Action Explanation": *Crítica*, 30, 29-43.

Moya, C. J. (1998b), "Justificación, causalidad y acción intencional": *Theoria* 13, 349-365.

Parfit, D. (1997), "Reasons and Motivation": *Proceedings of the Aristotelian Society*, Suppl. Vol. 71, 99-130.

Platón (1986), *Diálogos III: Fedón, Banquete, Fedro*, Gredos, Madrid.

Sosa, E. (1993), "Davidson's Thinking Causes", en Heil y Mele (eds.), 1993, 41-50.

Taylor, Ch. (1964), *The Explanation of Behaviour*, Routledge & Kegan Paul, Londres.

Toulmin, S. (1974), "Razones y causas", en R. Borger y F. Cioffi (eds.), *La explicación en las ciencias de la conducta*, Alianza, Madrid, 19-50.

Von Wright, G. H. (1979), *Explicación y comprensión*, Alianza, Madrid.

Winch, P. (1972), *Ciencia social y filosofía*, Amorrortu, Buenos Aires.