

(Pre-print)

LA NATURALIZACION DE LA RESPONSABILIDAD MORAL*

Carlos J. Moya. Universitat de València

En su importante artículo, "Epistemología naturalizada", Quine propuso concebir la epistemología como una parte de la ciencia, y en especial de la psicología empírica. La psicología en la que pensaba Quine era la psicología conductista, una psicología que prescindía de conceptos intencionales y se ocupaba sólo de "la constitución física y la conducta de organismos".¹ La admiración de Quine hacia la física y su desconfianza hacia lo intencional le inclinaba hacia una ontología fisicalista en la que no cabían las actitudes proposicionales. El programa quiniano de naturalización presenta un sesgo fuertemente científicista y proclive al eliminativismo con respecto a los estados intencionales en general. Sin embargo, aunque una actitud favorable hacia la ciencia parece consustancial a las propuestas de naturalización de la filosofía, ni el científicismo ni el eliminativismo son realmente esenciales a ellas. Formas más moderadas de la naturalización pueden admitir que los problemas filosóficos presentan peculiaridades que los distinguen de los problemas de carácter científico y que los conceptos intencionales de la llamada "psicología del sentido común", como los conceptos de creencia, deseo o intención, son indispensables para la psicología empírica y otras ciencias sociales. Lo que, en cambio, parece un requisito mínimo de una concepción naturalizada de un determinado ámbito de la reflexión filosófica es la exigencia según la cual dicha reflexión debe proceder mediante conceptos admitidos, o al menos admisibles, por las ciencias empíricas. No estoy seguro de que Quine hubiera

* Este trabajo fue presentado originalmente en el Simposio "La naturalización de la filosofía: problemas y límites", celebrado los días 10 y 11 de mayo de 2004 en homenaje al profesor Josep Lluís Blasco. Agradezco a los participantes en el simposio su atención y comentarios. Sin embargo, creo expresar el sentir de todos ellos al decir que, por mi parte, habría preferido que este simposio no hubiera tenido lugar, porque ello habría significado que el profesor Blasco estaría todavía entre nosotros.

¹ Quine, W. O., *Word and Object*, MIT Press, Cambridge MA, 1960, p. 221 (tr. castellana de Manuel Sacristán en ed. Labor).

considerado aceptable una concepción tan laxa de la naturalización de la reflexión filosófica, pero así es como propongo entender esta propuesta en lo que sigue.

En este trabajo me ocuparé de un objeto tradicional de reflexión filosófica: la cuestión de la responsabilidad moral y la cuestión relacionada de la libertad o el libre albedrío. Con respecto a estas cuestiones, las ciencias sociales, y en particular la psicología, son especialmente pertinentes. Una aproximación naturalizada a estos problemas puede emplear conceptos como *creencia, deseo, decisión, intención, motivo, razón, yo (self)*, e incluso *valor*. Estos son, pienso, conceptos admitidos, o al menos admisibles, por la psicología y otras ciencias sociales. Con respecto al concepto de valor, la psicología evolutiva, por ejemplo, emplea este concepto al estudiar el desarrollo moral del niño. Una concepción naturalizada de la responsabilidad moral debería poder dar cuenta de esta propiedad mediante el uso de estos conceptos y, eventualmente, de conceptos procedentes de otras ciencias empíricas, sociales o naturales.

Requisitos tan permisivos como éstos han sido sin embargo violados por algunas teorías filosóficas, pasadas y presentes, de la libertad y la responsabilidad moral, especialmente por teorías de carácter libertarista. Así, algunas de estas teorías usan un concepto de yo, como sujeto de la libertad y la responsabilidad moral, entendido como un núcleo incondicionado de decisión, una entidad distinta e independiente de determinaciones psicológicas como deseos y motivos y capaz de enjuiciarlas desde fuera. El yo substancial de Descartes y el yo nouménico kantiano serían tal vez ejemplos de este tipo de perspectiva. Lo que conduce, al menos en parte, al uso de este concepto de yo es el temor de que la determinación causal de nuestras decisiones por nuestras creencias, deseos y motivos mine las bases de la libertad y la responsabilidad moral, en la medida en que tales estados psicológicos nos vienen dados y no están bajo nuestro control voluntario. Así, sólo la existencia de un yo separado del juego de nuestras creencias, deseos y motivos y de sus relaciones causales parece capaz de justificar las atribuciones de responsabilidad moral. Otro ejemplo pertinente serían las teorías que postulan un concepto de relación causal *sui generis*,

cuyos *relata* no son dos cambios o sucesos particulares, sino una entidad, un yo agente, por un lado, y un cambio o suceso, una decisión o una acción, por otro. Estas teorías filosóficas, conocidas como "teorías de la causalidad del agente", parecen violar también los requisitos mínimos indicados para una concepción naturalizada de la libertad y la responsabilidad moral.

La postulación de este tipo de entidades o relaciones no es, sin embargo, gratuita. Estas teorías no naturalizadas se ven llevadas a adoptar esos compromisos ontológicos debido a lo que, aparentemente, es una condición necesaria de la justificación y la verdad de las atribuciones de responsabilidad moral, a saber, el *control* del agente sobre sus decisiones y acciones. Desde luego, para que una atribución de responsabilidad moral esté justificada, el agente ha de tener un grado apropiado de control cognitivo y volitivo sobre su acción. Ha de llegar a ella mediante el uso de capacidades normales de razonamiento práctico, de modo que su decisión sea sensible a sus razones. Pero esto no parece suficiente: el agente ha de tener además un control adecuado sobre las raíces de su acción y ser responsable de ellas. Esta exigencia adicional es la ultimidad en el control sobre la acción. Dos, pues, parecen ser los aspectos constitutivos del control necesario para la responsabilidad moral: la racionalidad y la ultimidad. Podemos denominar esta condición "control último". Este requisito deriva de la naturaleza a la vez parcialmente causal y fuertemente valorativa de las atribuciones de responsabilidad moral. El objeto propio de estas atribuciones es el *agente* mismo, cuyo valor y mérito como persona se ve potencialmente aumentado o menoscabado en tanto que se le considera como autor y origen de una acción o una consecuencia de esta que son juzgadas buenas o malas. Así, para que las atribuciones de responsabilidad moral estén justificadas, el agente ha de ser realmente merecedor de ellas, lo cual requiere que aquello por lo que se le considera responsable dependa en último término de él. De otro modo, su valor personal podría verse disminuido o aumentado debido a factores completamente ajenos a su control e influencia. No es suficiente con que el agente sea una causa instrumental o intermedia de su acción: ha de ser la causa última, el iniciador y autor genuino de esa acción. La capacidad de ser iniciadores u orígenes genuinos de sus acciones es un aspecto esencial de la libertad o

libre albedrío que atribuimos a muchos seres humanos y un fundamento central del valor y la dignidad que les concedemos. Robert Kane ha denominado esta condición de control último "responsabilidad última" y Galen Strawson "autodeterminación genuina".

Supongamos que el control o la responsabilidad última es realmente un requisito de la responsabilidad moral. En ese caso, el logro de una concepción naturalizada de esta última, en el sentido que hemos dado a esta expresión, requerirá el logro de una concepción naturalizada del primero. Esta tarea presenta, sin embargo, grandes dificultades. Un modo de expresar estas dificultades sería el siguiente. Desde una perspectiva naturalista, los seres humanos somos configuraciones contingentes formadas a partir de factores físicos y biológicos presentes previamente a nuestro nacimiento y sobre los cuales no hemos tenido control alguno. Sin embargo, muchos de nuestros rasgos psicológicos básicos vienen determinados por esta herencia física y biológica. No tenemos tampoco control alguno sobre el entorno físico y social en el que pasamos los primeros años de nuestra vida, y este entorno moldea asimismo de una manera decisiva nuestro yo y determina nuestras oportunidades de formación y desarrollo personal. Podemos, desde luego, proceder a una configuración adicional de nuestra personalidad y nuestros valores, pero esta configuración adicional ha de partir de factores como deseos, valores y motivos que nos vienen simplemente dados y que no hemos podido elegir. ¿Cómo podría entonces un agente poseer sobre sus propias acciones un grado de control suficiente para ser verdaderamente merecedor de atribuciones de responsabilidad moral y de la fuerte valoración personal que las acompaña? Una concepción naturalizada de la responsabilidad moral ha de ofrecer una respuesta razonada a esta pregunta. En este contexto, los compatibilistas han incidido sobre el componente racional del control que los agentes han de tener sobre sus propias decisiones, con la esperanza de mostrar que la satisfacción de este aspecto puede satisfacer al mismo tiempo la intuición que subyace al componente de ultimidad. Los libertaristas, por su parte, insisten más bien en esta ultimidad del control, en la importancia de elecciones o decisiones causalmente indeterminadas, capaces de configurar el propio yo, con la esperanza de mostrar que tales

decisiones pueden satisfacer también el componente racional del control y no ser en último término arbitrarias.

Para determinadas formas de compatibilismo, como la representada por Hume, un sujeto ejerce el control adecuado sobre sus propias acciones, y puede ser así moralmente responsable de ellas, cuando actúa por sus propios motivos y deseos. Este modo de actuar, que puede ser llamado libre, se contrapone a la actuación causada por una coacción o fuerza externa, o por una compulsión patológica. En el primer caso, cabe considerar al sujeto como origen de su acción, en la medida en que esos motivos y deseos por los que actúa son parte del complejo psicológico que constituye su yo. El hecho de que este complejo psicológico tenga causas externas al control del sujeto, como su herencia biológica o su entorno social, no afectaría a su libertad y a su responsabilidad moral por sus acciones: a fin de cuentas, es él mismo la causa de su acción. Pero esta concepción de la libertad y el control necesarios para la responsabilidad moral no satisface la intuición que subyace al requisito de ultimidad, porque hay casos en que las acciones de un sujeto surgen de sus deseos y de su yo y sin embargo no le consideramos moralmente responsable de tales acciones porque juzgamos que no es realmente responsable de tener esos deseos y de ser como es. Ciertos casos de sujetos con una infancia traumática serían aquí pertinentes. Nuestra reacción inicial frente a un sujeto que ha cometido actos detestables puede ser de indignación, pero esta reacción se ve fuertemente atemperada si nos enteramos de que en su infancia sufrió toda clase de vejaciones y maltratos que marcaron indeleblemente su personalidad.

Otras formas de compatibilismo, como la representada por Harry Frankfurt, tratan de dar cuenta de la intuición que subyace a esta exigencia de ultimidad insistiendo en que la responsabilidad moral requiere un alto grado de complejidad psicológica. Para que un agente sea moralmente responsable de una acción suya, no es suficiente que esta acción esté causada por su propio carácter, por sus propios deseos y motivos. Es necesario también que el sujeto sea responsable de ese carácter, de esos deseos y motivos por los que actúa, de modo que éstos sean realmente suyos. Aun cuando determinados deseos, motivos y rasgos

de carácter le vengan dados, un sujeto puede aún adoptar ante ellos una actitud reflexiva de aceptación o de rechazo. Tenemos deseos que no deseamos tener, o que no deseamos que nos lleven a actuar, aun cuando a veces lo hagan. Al aceptar ciertos motivos e identificarse con ellos, el sujeto se hace al mismo tiempo responsable de las decisiones y acciones que resultan de ellos. En el caso de Frankfurt, y a diferencia de Hume, el yo está constituido por las actitudes reflexivas ("voliciones de segundo orden", en terminología de Frankfurt) que un sujeto adopta hacia los deseos y motivos por los que actúa. Así, para Frankfurt, "en la medida en que una persona se identifica con los motivos de sus acciones, asume la responsabilidad de esas acciones y adquiere responsabilidad moral por ellas".² Así, aun cuando no tengamos control inicial sobre nuestros motivos y nuestro carácter, podemos, retrospectivamente, lograr ese control al identificarnos reflexivamente con ellos.

Es dudoso, sin embargo, que la propuesta de Frankfurt satisfaga realmente la intuición que subyace a la exigencia de ultimidad en el control como requisito de la responsabilidad moral. Pensemos de nuevo en un sujeto con una infancia fuertemente traumática. Supongamos que este sujeto adopta reflexivamente una actitud de aceptación hacia su carácter y motivos. No parece que esta actitud reflexiva lo convierta sin más en un agente moralmente responsable de sus acciones, porque las mismas circunstancias que explican su carácter y motivos, y nos llevan a negar que sea responsable de ser como es y de tener los motivos que tiene, pueden explicar asimismo su actitud reflexiva de identificación con esos motivos y carácter y llevarnos a negar que sea realmente responsable de esa actitud. El control reflexivo de Frankfurt no satisface realmente la exigencia de ultimidad que parece necesaria para la responsabilidad moral.

Mientras que Frankfurt elabora la noción de control sobre la base de una jerarquía de deseos o actitudes volitivas y de un ajuste entre ellas, Gary Watson interpreta esta noción sobre la base de una distinción entre dos fuentes distintas e independientes de motivación, a saber, deseos y valores. Así como para Frankfurt una persona controla su acción en la medida

² Frankfurt, H. G., 'Three Concepts of Free Action', en *The Importance of What We Care About*, Cambridge University Press, Cambridge, 1988, p. 54.

en que la lleva a cabo por un motivo por el que desea ser movido y con el cual se identifica, para Watson una persona controla su acción en la medida en que es movida a actuar por sus propios valores. Es la armonía entre el sistema motivacional y el sistema axiológico lo que constituye el control sobre las propias acciones que requiere la responsabilidad moral. Por otra parte, así como, para Frankfurt, el yo está constituido esencialmente por las actitudes reflexivas o voliciones de segundo orden, para Watson está constituido esencialmente por valores. Así, cuando una acción está causada por estos valores, tiene al mismo tiempo su origen en el yo y el agente puede ser considerado, legítimamente, como moralmente responsable de ella.

Aun cuando las propuestas de Frankfurt y de Watson difieren en contenido, son sin embargo estructuralmente muy semejantes. Ambos subrayan que la responsabilidad moral requiere una estructura psicológica con un alto grado de complejidad y conciben el control de un sujeto sobre sus acciones en términos de un ajuste interno entre distintos aspectos de dicha estructura psicológica. Ambas propuestas pueden calificarse de ahistóricas, en tanto que no consideran pertinente, para la responsabilidad moral, el origen efectivo de los factores psicológicos involucrados en el control del sujeto sobre sus acciones, tales como las voliciones de segundo orden o los valores. Lo importante para la responsabilidad moral es que el agente posea una estructura psicológica con la complejidad y las relaciones internas requeridas, y no el modo en que esa estructura se ha generado. No importa, por ejemplo, que el proceso que ha dado lugar a esa estructura sea un proceso causal determinista: la responsabilidad moral es, para ambos autores, compatible con el determinismo. Por otra parte, se trata de teorías puramente formales e internistas. Lo pertinente para el control de un agente sobre sus acciones y su responsabilidad moral es la posesión de las actitudes psicológicas correspondientes y su ajuste interno, y no su contenido, su corrección o su verdad.

Estas semejanzas hacen que ambas propuestas se enfrenten también a dificultades semejantes. La intuición que nos lleva a juzgar que ciertos sujetos con una infancia fuertemente traumática no son responsables de ser como son y de tener los deseos y motivos

que tienen no se ve debilitada por la adición de actitudes reflexivas (Frankfurt) o de valores (Watson): en determinados casos, no tendemos tampoco a considerar a estos sujetos responsables de esas actitudes y valores, ni de las acciones que resultan de ellas. No vemos a estos sujetos como el origen genuino de sus acciones. La intuición que subyace a la exigencia de ultimidad no queda realmente salvaguardada por el control reflexivo y axiológico sobre los motivos de las acciones propugnado por estas teorías. Otro modo de expresar esta objeción es señalar que sujetos sometidos a procesos ocultos de condicionamiento sistemático parecen en principio capaces de cumplir las condiciones que estas teorías exigen para la responsabilidad: los ciudadanos del Mundo Feliz de Huxley, en tanto que posean la complejidad psicológica y el ajuste interno exigidos por estas teorías, serían agentes con plena responsabilidad moral. Pero esta consecuencia parece claramente inaceptable.

Sin embargo, la carencia de control *último* sobre sus acciones podría no ser la única explicación de nuestro juicio según el cual sujetos con una infancia fuertemente traumática o ciudadanos del Mundo Feliz no son moralmente responsables de sus acciones. Una explicación alternativa, propuesta por Susan Wolf, sería que los sujetos en esas circunstancias no son capaces de formar un sistema de valores correcto y no distorsionado. Wolf está de acuerdo con Watson en que la capacidad de actuar de acuerdo con los propios valores es necesaria para la responsabilidad moral, pero no considera este requisito suficiente. Se requiere también la capacidad de formar valores objetivamente correctos. Wolf denomina esta propuesta "la perspectiva de la Razón". Así, según la propuesta de Wolf, "un individuo es responsable si, y sólo si, es capaz de configurar sus acciones sobre la base de sus valores y es capaz de formar sus valores sobre la base de lo que es Verdadero y Bueno".³ La condición añadida por Wolf es explícitamente normativa. Su carácter naturalizado depende, pues, de la posibilidad de ofrecer una concepción naturalizada de la normatividad y de la corrección de los juicios axiológicos. Para Wolf, el problema de los agentes con infancias traumáticas o sometidos a condicionamiento oculto no tiene que ver con el origen último de sus valores o

³ Wolf, S., *Freedom Within Reason*, Oxford University Press, Oxford, 1990, p. 75.

de sus actitudes reflexivas, sino con su incapacidad para formar valores objetivamente correctos: "Una víctima de una infancia traumática (o depravada), por ejemplo, puede ser tan inteligente como una persona que ha crecido en un entorno más normal, pero, debido a un conjunto de experiencias lamentablemente sesgado, sus valores pueden estar distorsionados. Es capaz de razonar, digamos, pero no de actuar conforme a la Razón".⁴

Para Wolf, la cuestión del control y origen último de los valores de un sujeto puede reducirse a cuestiones acerca del contenido y la corrección de esos valores. Si un agente puede actuar sobre la base de sus valores y formar valores objetivamente correctos, es un agente libre y moralmente responsable, y viceversa, con independencia del origen de esa capacidad y esos valores.

La explicación de Wolf de nuestro juicio en ciertos casos de sujetos con una infancia traumática o sometidos a condicionamiento oculto puede tener cierta plausibilidad, debido a que un resultado probable en tales casos es un sistema distorsionado o incoherente de valores. Pero este resultado no es necesario. Es concebible que un sujeto forme sus valores a través de un proceso de condicionamiento sistemático y que estos valores sean correctos y no distorsionados. Pensemos, por ejemplo, en la posibilidad de lo que llamaré un Mundo Feliz Platónico,⁵ regido por benevolentes filósofos-reyes que conocen la Verdad y el Bien, y en el que los ciudadanos son ocultamente condicionados para formar valores objetivamente correctos y para actuar según ellos. Estos ciudadanos parecen satisfacer plenamente las condiciones necesarias y suficientes que Wolf establece para la responsabilidad moral, de modo que, a la luz de esta teoría, serían agentes moralmente responsables. Pero esta consecuencia es difícilmente aceptable. Y lo que explica este juicio no puede ser que los sujetos posean valores erróneos o incoherentes. Parece ser más bien nuestra percepción de que tales sujetos no son el origen genuino de sus propios valores y no poseen control *último* sobre ellos. Sus acciones pueden ser moralmente buenas, pero ellos no merecen realmente

⁴ *Ibid.*, p. 75-6.

⁵ Cf. al respecto mi manuscrito, no publicado, *Moral Responsibility: The Ways of Scepticism*, cap. 3, secc. 5.

alabanza por ellas. Así, pues, la propuesta de Wolf, como la de Frankfurt o Watson, no consigue tampoco dar cuenta de las intuiciones que subyacen a la exigencia de ultimidad.

Un intento reciente de dar cuenta de la condición de control último desde una perspectiva naturalizada, de inspiración compatibilista, lo debemos a John Martin Fischer y Mark Ravizza. Según estos autores, el control necesario para la responsabilidad moral incluye, como en el caso de las propuestas anteriores, un requisito de control racional. El modo en que ellos interpretan este requisito tiene que ver con el 'mecanismo' con el que el agente ha llegado efectivamente a su decisión y acción. Para ser moralmente responsable de una acción suya, el sujeto ha de llegar a esta acción a través de un mecanismo que responda adecuadamente a razones. Fischer y Ravizza denominan este requisito "capacidad moderada de respuesta a razones". Que un mecanismo de deliberación y toma de decisiones es moderadamente capaz de responder a razones significa, *grosso modo*, que, al menos en algunos casos, y manteniendo fijado ese mecanismo, si hubiera razones suficientes para actuar de un modo diferente de como el sujeto actúa de hecho, el sujeto reconocería estas razones y actuaría de otro modo. Así, en términos de este requisito, un cleptómano severo o un sujeto afectado por una compulsión o una fobia grave no serían moralmente responsables de aquellas acciones a las que han llegado debido a esas patologías. En términos más generales, si un sujeto decidiría y actuaría del mismo modo en que ahora lo hace aun cuando, en cualquier circunstancia, hubiera razones decisivas para decidir y actuar de otro modo, este sujeto no satisface el requisito indicado. El requisito es ciertamente plausible y apunta a algo que claramente parece necesario para la responsabilidad moral por una acción, a saber, una capacidad normal para la deliberación, el razonamiento práctico y la toma de decisiones. Sin embargo, para Fischer y Ravizza, este requisito no es todavía suficiente. Además de ser moderadamente sensible a razones, el 'mecanismo' por el que un agente llega a una acción ha de ser genuinamente propio del agente. Este requisito es el que no parecen satisfacer los ciudadanos del Mundo Feliz. El mecanismo con el que deliberan y deciden es atribuible al equipo que dirige el proceso de condicionamiento de los ciudadanos, más que a los

ciudadanos mismos. Si el primer requisito es puramente estructural, este segundo requisito tiene carácter histórico: guarda relación con el *origen* de los factores psicológicos que conducen a la acción, más que con su estructura y organización interna. En este aspecto histórico, la propuesta de Fischer y Ravizza se distingue de las de Frankfurt, Watson y Wolf. La cuestión del origen de los factores involucrados en la producción de la acción ha de ser tratada por sí misma, con independencia de la cuestión de la naturaleza y estructura interna de tales factores. La unión de ambos aspectos, estructural e histórico, constituye lo que Fischer y Ravizza denominan el "control guía", el cual, dadas ciertas condiciones cognitivas, es necesario y suficiente para la responsabilidad moral.

La especificación de las condiciones históricas que han de ser satisfechas para que el 'mecanismo' que conduce a la acción sea propio del agente toma la forma de una descripción del proceso por el que un niño llega a convertirse en un agente moral. En este proceso cabe distinguir tres estadios. En primer lugar, el niño recibe un *entrenamiento* moral, siendo tratado por sus padres o educadores *como si* fuese responsable. Este proceso lleva al niño a considerarse a sí mismo como objeto apropiado de atribuciones de culpabilidad y alabanza, y esta actitud es esencial en la *asunción de responsabilidad*. Sobre la base de esta actitud, el niño está ya dispuesto a *ser considerado responsable*. Según Fischer y Ravizza, estos distintos estadios "se superponen unos a otros y componen un ciclo que ha de repetirse continuamente en el proceso de llegar a ser un agente moralmente responsable en sentido pleno".⁶

Un paso crucial en este proceso es la asunción de responsabilidad. Fischer y Ravizza conciben este paso, más que como una acción, como la adquisición por parte del sujeto de un conjunto de creencias. El sujeto llega a creer que es un agente, cuyas elecciones y acciones tienen efectos sobre el mundo; que, si eligiera y actuara de otro modo, los efectos serían también distintos; que es un objeto apropiado de atribuciones de responsabilidad moral, etc. A través de este proceso, el sujeto se responsabiliza de los motivos de sus acciones y las hace suyas.

⁶ Fischer, J. M. y Ravizza, M., *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge, 1998, p. 210.

Fischer y Ravizza pueden dar cuenta de nuestro juicio acerca de ciertos casos de sujetos con infancias traumáticas, bien en términos del requisito de respuesta a razones o del requisito histórico. Por otra parte, en la medida en que la teoría incluye un componente histórico, parece en mejor situación que las propuestas anteriores para afrontar el problema del Mundo Feliz. Ante esta cuestión, Fischer y Ravizza podrían argüir que, aun cuando el mecanismo de deliberación y toma de decisiones que estos ciudadanos emplean sea adecuadamente sensible a razones, no sería propiamente suyo, debido al condicionamiento que lo ha originado. Según esta teoría, el mecanismo que un agente emplea es propiamente suyo suponiendo que se haya originado a través del proceso de entrenamiento y asunción de responsabilidad al que nos hemos referido. Sin embargo, no hay ninguna razón para pensar que los ciudadanos del Mundo Feliz no se han formado como agentes a través de un proceso de este tipo. Del mismo modo que estos ciudadanos han sido condicionados para tener los valores que tienen, pueden perfectamente haber sido ocultamente condicionados para entrenar a sus hijos como agentes morales y llevarlos a adquirir las creencias involucradas en la asunción de responsabilidad por sus motivos y acciones. Así, en términos de esta teoría, los mecanismos de deliberación y decisión de los ciudadanos del Mundo Feliz serían propiamente suyos. Estos sujetos pueden, pues, satisfacer las condiciones que, según Fischer y Ravizza, son necesarias y suficientes para la responsabilidad moral y ser por ello agentes moralmente responsables. Pero, una vez más, esta consecuencia de la teoría contradice abruptamente nuestro juicio natural sobre estos casos. La propuesta naturalizada de Fischer y Ravizza no satisface tampoco nuestras intuiciones sobre la ultimidad del control necesario para la responsabilidad moral. Lo que subyace a nuestro juicio según el cual los ciudadanos del Mundo Feliz no son moralmente responsables es que, en un sentido profundo, no son realmente el origen de sus propias acciones.

Las teorías que hasta ahora hemos presentado son de inspiración compatibilista, es decir, tratan de mostrar la posibilidad de la responsabilidad moral bajo el supuesto del determinismo causal. Los principales intentos de desarrollar concepciones naturalizadas de la

libertad y la responsabilidad moral han procedido de hecho del campo compatibilista. Ello no es de extrañar si pensamos que el compatibilismo nace, de la mano de Hobbes y Hume, en una época en que la ciencia ofrece una imagen estrictamente determinista del universo. Esta puede ser también la razón por la cual el incompatibilismo libertarista o libertarismo, que considera la libertad y la responsabilidad moral incompatibles con el determinismo, ha tendido históricamente a adoptar una actitud, si no de oposición, al menos de indiferencia hacia las ciencias empíricas, lo que le ha valido frecuentes acusaciones de oscurantismo y misterio. Sin embargo, esta actitud no es consustancial al libertarismo. La posibilidad de que las leyes físicas no sean deterministas ha llevado a muchos libertaristas a abandonar la desconfianza o indiferencia hacia las ciencias y a ver en ellas un apoyo, más que un obstáculo, a su defensa de la libertad y la responsabilidad moral. Uno de los principales defensores actuales del libertarismo, Robert Kane, ha insistido en la necesidad de acercar el libertarismo a las ciencias empíricas y de hacerlo comprensible a una mentalidad científica. Así, según Kane, "quien esté dispuesto a defender la libertad libertarista en la época moderna ha de... hacer... esfuerzos para mostrar como el libre albedrío incompatibilista puede tener un lugar en la imagen científica del mundo".⁷ Para Kane, un libertarismo razonable no debería comprometerse con entidades o relaciones con las que no se comprometan también otras teorías, en especial compatibilistas. Así, en nuestros términos, la pretensión de Kane es desarrollar una concepción naturalizada de la libertad y la responsabilidad moral de inspiración libertarista.

El libertarismo de Kane deriva de su convicción según la cual hay un aspecto de la libertad que es incompatible con el determinismo causal y que se requiere para que un agente sea realmente merecedor de la valoración involucrada en las atribuciones de responsabilidad moral, a saber, que dicho agente sea la fuente u origen último de la acción correspondiente. Como Kane escribe, "es la preocupación por el *merecimiento objetivo* por nuestras acciones y carácter lo que lleva de modo natural a una preocupación por la cuestión de si esas acciones

⁷ Kane, R., *The Significance of Free Will*, Oxford University Press, Oxford, 1996, pp. 212-13.

y caracteres tienen sus fuentes últimas en nosotros".⁸ Este aspecto está, pues, directamente relacionado con el requisito de control último y con la concepción de un ser que posee libre albedrío como un ser que es origen e iniciador genuino de sus acciones. Y esta idea ocupa un lugar central en el modo en que Kane concibe el libre albedrío, que él define como "el poder que tienen los agentes de ser creadores (u originadores) últimos y sostenedores de sus propios fines y propósitos".⁹ Es el ejercicio de este poder lo que confiere a un agente verdadero mérito o demérito por sus acciones y justifica las atribuciones de responsabilidad moral, con el carácter fuertemente valorativo que les es propio.

Para Kane, un agente tiene la responsabilidad última de sus decisiones y acciones sólo si es responsable de los factores que explican que las lleve a cabo. Así, si, en un momento dado, su carácter, valores y motivos constituyen una explicación suficiente de su decisión y acción, el agente ha de ser a su vez personalmente responsable de tener ese carácter, valores y motivos para ser moralmente responsable de la decisión y acción que resultan de ellos. Y para ser responsable de ese carácter y motivos, ha de haberlos elegido voluntaria y racionalmente. Pero ¿cómo podría haberlos elegido voluntaria y racionalmente a menos que ya existiera dotado de un carácter y unos motivos previos? Kane es consciente de que su concepción del control o responsabilidad última "parece llevar a un regreso vicioso... El regreso se detendría con acciones que no fuesen explicadas por nuestro carácter y motivos (o por cualquier otra cosa), pero entonces, ¿en qué sentido seríamos responsables de esas acciones?".¹⁰

Para responder a esta pregunta, Kane apela a la existencia, en la vida de una persona, de elecciones a través de las cuales dicha persona forma su propio carácter y motivos. Kane llama a estas elecciones "voliciones autoformadoras". Si estas elecciones han de detener el regreso vicioso que amenaza la posibilidad del control último, deben satisfacer determinadas condiciones. En primer lugar, los factores presentes en el momento de llevarlas a cabo no han

⁸ *Ibid.*, p. 98.

⁹ *Ibid.*, p. 4.

¹⁰ *Ibid.*, p. 37.

de constituir una explicación suficiente de las mismas. Así, la elección no ha de estar determinada por el pasado y las leyes de la naturaleza, y en particular por los rasgos psicológicos, motivos y valores preexistentes del agente. La elección ha de estar abierta y genuinamente indeterminada, ha de poder "ocurrir o no, dado *exactamente el mismo pasado y leyes naturales*",¹¹ pues de otro modo el agente no podría ser su origen *último*. Y, en segundo lugar, el agente ha de tener control racional y volitivo de la elección; la elección, en cualquier sentido en que se lleve a cabo, ha de ser fruto de la voluntad racional del agente, pues de otro modo la elección contendría un elemento de arbitrariedad o irracionalidad incompatible con el control y la responsabilidad del sujeto por ella. La responsabilidad moral no puede tener como fundamento actos arbitrarios o irracionales.

¿Puede haber elecciones que satisfagan estas condiciones? Kane incluye entre los casos centrales de este tipo de voliciones autoformadoras las elecciones morales y prudenciales. Estas elecciones "involucran conflictos entre lo que un agente cree que se debería hacer y lo que el agente desea o quiere hacer".¹² No se trata de casos Buridan, en los que un agente se enfrenta a razones equivalentes para decidir en uno u otro sentido. En casos como éstos, cualquier elección será arbitraria. Se trata más bien de que el agente se halla frente a conjuntos de razones *inconmensurables* en favor de dos alternativas, por ejemplo, normas morales o prudenciales, por un lado, y deseos, por otro. En estos casos, "el agente se encuentra dolorosamente escindido entre puntos de vista internos en conflicto que representan *visiones distintas e inconmensurables de lo que desea en la vida* o de aquello que desea llegar a ser".¹³ Estas elecciones son 'autoformadoras' porque configuran y modelan el yo del agente, su carácter, motivos y voluntad, de modo que el agente puede así ser considerado responsable último de su propio yo, carácter y motivos y, con ello, de ulteriores acciones que puedan derivar de estos factores.

¹¹ *Ibid.*, p. 106.

¹² *Ibid.*, p. 126.

¹³ *Ibid.*, p. 199.

A pesar de su interés, la propuesta de Kane se enfrenta a importantes dificultades. En una volición autoformadora, como hemos visto, el agente se enfrenta a una elección que ha de llevar a cabo sobre la base de razones recíprocamente inconmensurables. Pero si las razones son realmente inconmensurables, ¿cómo puede el agente optar racional o razonadamente por una de las alternativas? Una elección de este tipo puede llevarse a cabo racionalmente si el agente dispone de algún criterio adicional (un meta-criterio, cabría decir) que le lleve a situar un conjunto de razones por encima del otro. Este criterio puede ser, por ejemplo, una volición de segundo orden, según la cual prefiere ser movido a actuar por razones, digamos, morales que por razones de autointerés, o un sistema de valores que sitúe las primeras por encima de las segundas. Pero si el agente posee ya ese criterio adicional, surge entonces la pregunta de cómo ha llegado a tener ese criterio. Si ha llegado a él a través de previas elecciones entre razones inconmensurables, el problema de la racionalidad se plantea de nuevo en relación con estas otras elecciones. Y si se ha encontrado a sí mismo teniendo ese criterio adicional, no posee entonces control último con respecto a esta elección, pues la lleva a cabo sobre la base de un criterio que él mismo no ha elegido; la elección, pues, no es realmente autoformadora. Por otra parte, ya hemos indicado que, para Kane, una elección autoformadora habrá de ser fruto de la voluntad racional del agente, en cualquier sentido en que se lleve a cabo. Pero esta condición no se cumple si el agente afronta la elección con un criterio adicional, porque en tal caso una elección conforme con ese criterio será fruto de la voluntad racional del agente, pero una elección contraria al mismo no lo será. Así, una elección autoformadora entre razones inconmensurables no puede llevarse a cabo sobre la base de un meta-criterio o criterio adicional, sino exclusivamente sobre la base de las razones inconmensurables que el agente contempla. Pero entonces es difícil ver cómo el agente podría controlar racionalmente esa elección. El papel que desempeñaría un criterio adicional es ocupado ahora por la pura *decisión* del agente: "El agente *hará* que un conjunto de razones o motivos prevalezca sobre el otro allí y entonces *decidiendo*".¹⁴ Pero entonces el

¹⁴ *Ibid.*, p. 133.

libertarismo de Kane tiende a aparecer como puro *decisionismo*. El carácter último de la elección se consigue a costa del control racional del agente sobre ella.

En cierto modo, el problema al que se enfrenta el libertarismo de Kane, en relación con el requisito de control último, es el opuesto del que vimos surgir en relación con el compatibilismo de Frankfurt, Watson o Wolf. El énfasis de estos últimos autores en el control cognitivo y racional no era suficiente para satisfacer la exigencia de que un agente sea fuente u origen último de sus acciones. La propuesta de Kane puede satisfacer esta última exigencia, pero a costa de sacrificar el control racional. Esta parece ser una consecuencia de la tensión interna entre los dos aspectos involucrados en el requisito de control último: la ultimidad del origen y el control racional. Estos dos aspectos parecen empujar en direcciones opuestas: el refuerzo del control racional sobre las decisiones y acciones conlleva un debilitamiento del agente como origen último de las mismas y viceversa. Pero si el control último es realmente necesario para la responsabilidad moral, es crucial mostrar que dicha tensión interna es una mera apariencia o que, si no lo es, puede en todo caso ser resuelta. De otro modo, el escepticismo sobre la posibilidad de la responsabilidad moral, una posición no carente de defensores en la actualidad, será el resultado más probable.

Ninguna de las propuestas de naturalización de la responsabilidad moral que hemos presentado es realmente satisfactoria. Cabría pensar en rechazar la perspectiva naturalizada en un intento de lograr una justificación adecuada de la responsabilidad moral y privar de base al escepticismo. Sin embargo, dados los requisitos mínimos de lo que aquí hemos entendido por "naturalización", no creo que esa alternativa sea muy prometedora. Hemos de seguir trabajando por mostrar cómo es posible la responsabilidad moral en el marco general de un naturalismo razonablemente amplio, alejado de cualquier veleidad científicista o reductivista. Una sugerencia, que aquí no puedo desarrollar, sería desplazar la voluntad, bajo sus distintas formas, como la decisión o la elección, del lugar central que hasta ahora ha venido ocupando en la reflexión filosófica sobre estas cuestiones y atender preferentemente a las creencias que subyacen a nuestras decisiones y elecciones. Tal vez cabría mostrar que

podemos tener, sobre esas creencias, un control que sea a la vez racional y lo suficientemente profundo para satisfacer nuestras intuiciones acerca del control último sobre nuestras acciones como requisito de la responsabilidad moral.