

---

**RAZÓN ECONÓMICA Y  
RAZÓN MORAL. EN TORNO  
A DAVID GAUTHIER**

## ¿POR QUÉ CONTRACTUALISMO?\*

*A medida que la voluntad de verdad gane así autoconciencia -y no se puede dudar de ello- la moralidad irá pereciendo poco a poco: este es el gran espectáculo, con un centenar de actos, que le está reservado a Europa durante los próximos doscientos años: el más terrible, el más problemático y, tal vez, también el más esperanzado de todos los espectáculos.*

Nietzsche<sup>1</sup>

### I

**L**a moralidad se ve enfrentada con una crisis de fundamentación. El contractualismo ofrece la única solución plausible de esta crisis. Estas dos proposiciones definen mi tema. Lo que sigue es su elaboración. Es posible que Nietzsche haya sido el primero -pero no el único- en reconocer la crisis a la que me refiero. Veamos algunas afirmaciones recientes: «La hipótesis que deseo formular es que en el mundo real en que vivimos, el lenguaje de la moralidad se halla en... [un] estado de grave desorden... hemos perdido en gran medida -si no por completo- la comprensión, tanto teórica como práctica, de la moralidad.» (Alasdair MacIntyre)<sup>2</sup>. «Los recursos de la mayor parte de la moderna filosofía moral no se ajustan adecuadamente al mundo moderno.» (Bernard Williams)<sup>3</sup>. «No existen valores objetivos... (Pero)

---

\* La versión original de este trabajo fue presentada en un seminario sobre la teoría moral contractualista en la Universidad de Western Ontario y será publicada en una compilación de ensayos basada en aquel seminario. Algunos fragmentos aparecen también en «Morality, Rational Choice and Semantic Interpretation - A reply to mi critics», en *Social Philosophy & Policy*. Agradezco a Annette Baier, Paul Hurley y Geoffrey Sayre McCord sus comentarios sobre un primer borrador. También agradezco a las personas que participaron en la discusión en la Western Washington University, en la Universidad de Arkansas, y en la Universidad de California (Santa Cruz), sus comentarios a una conferencia vinculada con este tema.

<sup>1</sup> *On the Genealogy of Morals*, trad. de Walter Kaufmann y R. J. Hollingdale, Random House, Nueva York, 1967, Tercer Ensayo, sección 27, pág. 161.

<sup>2</sup> *After Virtue*, Notre Dame, 1981, pág. 2.

<sup>3</sup> *Ethics and the Limits of Philosophy*, Harvard, Cambridge, Mass., pág. 197.

la principal tradición de la filosofía moral europea incluye la pretensión contraria.» (J. L. Mackie)<sup>4</sup>. «Las hipótesis morales no ayudan a explicar por qué las personas respetan lo que respetan. Por ello, la ética es algo problemático y el nihilismo debe ser tomado en serio... Una versión extrema del nihilismo sostiene que la moralidad es simplemente una ilusión... Según esta versión, deberíamos abandonar la moralidad, del mismo modo que un ateo abandona la religión tras haber decidido que los hechos religiosos no pueden contribuir a explicar las observaciones.» (Gilbert Harman)<sup>5</sup>.

Escojo estas afirmaciones para señalar algunos rasgos de la crisis con la que se enfrenta la moralidad. Ellas sugieren que el lenguaje moral responde a una visión del mundo que hemos abandonado: una visión del mundo como algo teleológicamente ordenado. Sin esta visión, ya no entendemos verdaderamente las exigencias morales que continuamos formulando. Aquellas afirmaciones sugieren que existe una inadecuación entre lo que la moralidad presupone -valores objetivos que contribuyen a explicar nuestra conducta- y los estados psicológicos -deseos y creencias- que, dada nuestra visión actual del mundo, nos brindan realmente la mejor explicación. Esta falta de adecuación amenaza con socavar la idea misma de una moralidad como algo más que una curiosidad antropológica. Pero, ¿cómo podría darse esto? ¿Cómo podría *perecer* la moralidad?

## II

Para poder continuar, tengo que ofrecer una mínima caracterización de la moralidad que se ve enfrentada con una crisis de fundamentación. Se trata de la moralidad de la restricción justificada. Desde el punto de vista del agente, las consideraciones morales se presentan como restricciones de sus elecciones y acciones, independientemente de sus deseos, propósitos e intereses. Más adelante, añadiré a esta caracterización otros rasgos pero, de momento, ello basta, pues pone claramente de manifiesto qué es lo que se cuestiona: el fundamento de la restricción. Este fundamento parece estar ausente de nuestra visión actual del mundo

---

<sup>4</sup> *Ethics: Inventing Right and Wrong*, Penguin, Harmondsworth, Middlesex, págs. 15, 30.

<sup>5</sup> *The Nature of Morality*, Oxford, Nueva York, 1977, pág. 11.

Y, por lo tanto, nos preguntamos: ¿qué razón puede tener una persona para reconocer y aceptar una restricción que es independiente de sus deseos e intereses? Puede admitir que una tal restricción podría estar *moralmente* justificada; tendría una razón para aceptarla si tuviera una razón para aceptar la moralidad. Pero, ¿qué justifica que se preste atención a la moralidad y no se la deje de lado como un apéndice de creencias trasnochadas? Preguntamos, y parece que no encontramos respuesta. Pero, antes de proseguir, deberíamos tomar en cuenta tres objeciones.

La primera consiste en cuestionar la idea de restricción. ¿Por qué la moralidad ha de ser considerada como algo que restringe nuestras elecciones y acciones? ¿Por qué no deberíamos decir más bien que la persona moral escoge con la mayor libertad, porque lo hace a la luz de una concepción verdadera de sí misma y no a la luz de las falsas concepciones que, con tanta frecuencia, predominan? ¿Por qué no establecer un vínculo entre la moralidad y la autocomprensión? Se podría considerar que Platón y Hume apoyaron este punto de vista. Pero Hume sería, en el mejor de los casos, un aliado parcial, ya que su concepción de la «virtud, en todos sus encantos genuinos y más atractivos, ... cuando no habla de austeridades y rigorismo, sufrimiento y autorrenuncia inútiles» sino que más bien hace que «sus partidarios ... en cada momento de su existencia, [sean,] dentro de lo posible, alegres y felices» queda oscurecida al admitir que «en el caso de la justicia, ... un hombre -vistas las cosas desde una cierta perspectiva- puede, a menudo, parecer un perdedor a causa de su integridad»<sup>6</sup>. Ciertamente, Platón va más allá, al insistir en que únicamente el hombre justo posee un alma sana pero, por muy heroica que sea la defensa que Sócrates hace de la justicia, todos tendemos a pensar que Glaucón y Adimanto llegaron a un acuerdo más por la persuasión que por la razón y que no se ha demostrado que el hombre injusto es necesariamente el perdedor<sup>7</sup>. No pretendo, en absoluto, proseguir esta línea de pensamiento. La moralidad, tal como la concebimos nosotros -herederos de la moral cristiana y kantiana- restringe las empresas a las que nos conducirían incluso nuestros deseos meditados («*considered*»). Y ello no es simple o íntegramente una restricción del autointerés: las afecciones que la moralidad refrena

---

II. <sup>6</sup> David Hume, *An Enquiry Concerning the Principles of Morals*, 1751, Sección IX, Parte II.

<sup>7</sup> Véase Platón, *República*, esp. los libros II y IV.

incluyen afecciones sociales, tales como el favoritismo y la parcialidad, para no hablar de la crueldad.

La segunda objeción a la opinión, según la cual la restricción moral está insuficientemente fundamentada consiste en cuestionar la pretensión de que opera independientemente -más que a través- de nuestros deseos, intereses y afectos. La moralidad, pueden decir algunos, se refiere al bienestar de todas las personas o, tal vez, de todas las criaturas sensibles<sup>8</sup>. Y entonces se podría argumentar o bien -siguiendo a Hume- que la moralidad surge en y de nuestra identificación simpática con nuestros congéneres, o bien que reside directamente en el bienestar, y que nuestras afecciones tienden a tener una disposición favorable al respecto. Pero, por supuesto, no todas nuestras afecciones. Y así, nuestros sentimientos de simpatía entran en una oposición característica con otros sentimientos, en relación con los cuales funcionan como restricciones.

Esta es una caracterización muy cruda, pero bastará para el presente argumento. Este punto de vista admite que la moralidad, tal como la entendemos, carece de fundamentos puramente *racionales*, pero nos recuerda que no por ello nos deja de interesar el bienestar de nuestros congéneres. La moralidad se funda en las preocupaciones generalizadas, de simpatía, dirigidas hacia los demás, que la mayoría de nosotros tenemos; estas preocupaciones ponen freno al autointerés y también al favoritismo y a la parcialidad con los que, a menudo, tratamos a los demás. Sin embargo, si la moralidad depende -para su relevancia práctica y para su eficacia motivacional- enteramente de nuestros sentimientos de simpatía, no puede aspirar a la fuerza prescriptiva con la que ha sido investida en la perspectiva cristiana y kantiana a las que me he referido, y que Glaucón y Adimanto exigían que Sócrates defendiera en el caso de la justicia. Pues el que se nos recuerde que algunas veces nos preocupamos de nuestros congéneres y estamos dispuestos a reprimir otros deseos para poner de manifiesto esta preocupación, no nos dice nada que pueda guiarnos en aquellos casos en los que manifiestamente no nos preocupamos por ellos, o no nos preocupamos lo suficiente; no nos dice nada que permita defender las exigencias que la moralidad nos impone en los casos difíciles. Es verdad que no todas las situaciones en las que la preocupación por los demás entra en

---

<sup>8</sup> Algunos estarían dispuestos a extender la moralidad a los seres no sensibles pero, por más simpatía que tenga por los derechos de los tranvías y de las locomotoras a vapor, propongo dejar de lado esta opinión.

conflicto con el autointerés son casos difíciles, pero la moralidad, tal como la entendemos normalmente, alude a estos casos, mientras que su sustituto humano o naturalista, no.

Estas observaciones se aplican al intento positivo reciente más firme de crear una teoría moral: el de John Rawls. Pues el intento de describir nuestra facultad moral o -más específicamente, según Rawls- nuestro sentido de la justicia, en términos de principios plausibles a la luz de nuestra teoría psicológica más general y coherentes con «nuestros juicios meditados en un equilibrio reflexivo»<sup>9</sup>, no dará ninguna respuesta a por qué, en aquellos casos en los que nuestro interés en ser justos es nulo o insuficiente, deberíamos, sin embargo, seguir los principios. John Harsanyi, cuya teoría moral es, en algunos aspectos, una variante utilitarista de la construcción contractualista de Rawls, lo reconoce explícitamente: «Lo único que podemos probar mediante argumentos racionales es que todo aquel que desee servir nuestros intereses humanos comunes de un modo racional debe obedecer estos mandatos»<sup>10</sup>. Pero, aunque la moralidad puede presentarse al servicio de nuestros intereses humanos comunes, no se ofrece tan sólo a quienes desean servirlos.

La moralidad es una restricción que, tal como lo reconociera Kant, no debe suponerse que depende únicamente de nuestros sentimientos. Y, por ello, no podemos apelar a los sentimientos para dar respuesta a la cuestión de su fundamentación. Pero la tercera objeción consiste en desechar directamente esta cuestión, rechazando la idea misma de una crisis de fundamentación. Nada justifica la moralidad, pues la moralidad no necesita justificación. En la moralidad, al igual que en otros asuntos, nos encontramos *in mediis rebus*. Formulamos, aceptamos y rechazamos, justificamos y criticamos juicios morales. La tarea de la teoría moral consiste en sistematizar esta práctica para, de este modo, proporcionarnos una comprensión más profunda de lo que es la justificación moral. Pero no hay fundamentos extra-morales de la justificación moral, de la misma manera que no existen fundamentos extra-epistémicos de los juicios epistémicos. En la moral, como en la ciencia, el fundacionalismo es un proyecto en bancarota.

---

<sup>9</sup> John Rawls, *A Theory of Justice*, Harvard, Cambridge, Mass., 1971, página 51.

<sup>10</sup> John C. Harsanyi, «Morality and the theory of rational behaviour», en Amartya Sen & Bernard Williams (eds.), *Utilitarianism and beyond*, Cambridge, 1982, pág. 62.

Afortunadamente no tengo que defender el fundacionalismo *normativo*. Uno de los problemas de la aceptación de la justificación moral como parte de nuestra práctica habitual consiste en que, tal como lo he sugerido, ya hemos dejado de aceptar la visión del mundo de la cual depende. Pero, tal vez, un problema más directamente acuciante sea el hecho de que tenemos, al alcance de la mano, un modo alternativo para justificar nuestras elecciones y acciones. En su forma más austera y, en mi opinión, más defendible, consiste en mostrar que las elecciones y las acciones maximizan la utilidad esperada de la gente, siendo la utilidad una medida de la preferencia meditada. En su versión menos austera, consiste en mostrar que las elecciones y las acciones no satisfacen un requisito definido subjetivamente, tal como la utilidad, sino los intereses objetivos del agente. Como no creo que tengamos intereses objetivos, pasaré por alto esta segunda posibilidad. Pero esto no importa. Pues la idea es clara: contamos con un modo de justificación que no requiere la introducción de consideraciones morales<sup>11</sup>.

Llamaré neutralmente a este modo alternativo, no moral, de justificación, justificación deliberativa. Ahora bien, tanto la justificación moral como la deliberativa se dirigen a los mismos objetos: nuestras elecciones y acciones. ¿Qué sucede si entran en conflicto? Y, ¿qué le diremos a la persona que ofrece una justificación deliberativa de sus elecciones y acciones y se niega a ofrecer ninguna otra? Por supuesto, podemos decir que su comportamiento no posee justificación *moral*, pero esto parece que carece de todo asidero, a menos que elija entrar en el esquema moral. Y es posible que insista en que este ingreso carece de toda justificación deliberativa, al menos para él.

Si la moralidad perece, no por ello perece la empresa justificatoria con respecto a la elección y la acción. Más bien, lo que perece es un modo de justificación, que puede parecer que ahora pende sin fundamento. Pero no sólo sin fundamento, pues es difícil negar que la justificación deliberativa es más claramente básica, que no puede ser evitada en la medida en que somos agentes racionales, de forma tal que si la justificación moral entra en conflicto con ella, la moralidad no parece sólo carecer de

---

<sup>11</sup> Desde luego, si pensamos en la moralidad tal como se expresa en algunas de nuestras afecciones y/o intereses, se incorporarán consideraciones morales en la medida en que ellas estén realmente presentes en nuestras preferencias. Pero esto supondría aceptar el naturalismo, que he dejado de lado por considerarlo inadecuado.

fundamento, sino oponerse a lo que racionalmente es más fundamental.

La justificación deliberativa está relacionada con el sentido profundo que tenemos de nosotros mismos. Lo que distingue a los seres humanos de otros animales y ofrece la base de la racionalidad, es la capacidad de representación semántica. Una persona puede, a diferencia de su perro, representarse a sí misma un estado de cosas y considerar específicamente si es real o si quiere o no que lo sea. Uno puede representarse a sí mismo el contenido de las propias creencias, deseos o preferencias, pero, al representarlos, se los pone en relación recíproca. Uno se representa a sí mismo que el Real Madrid ganará la Copa del Mundo, y que un equipo de la liga italiana ganará la Copa del Mundo, y que el Real Madrid no es un equipo de la liga italiana. Y, al reconocer un conflicto entre esas creencias, uno encuentra que la racionalidad se nos impone. Nótese que las dos primeras creencias podrían ser reemplazadas por preferencias, con el mismo efecto.

Como al representarnos nuestras preferencias nos damos cuenta del conflicto entre ellas, el paso de la representación a la elección se vuelve complicado. De algún modo, tenemos que establecer una especie de coherencia entre nuestras preferencias y deseos en conflicto. Y no existe más que un candidato plausible como principio de coherencia: un principio maximizador. Ordenamos nuestras preferencias en relación a la decisión y la acción, de modo tal que podamos escoger en un sentido que maximice la expectativa de satisfacción de nuestras preferencias. Y, al actuar de este modo, demostramos que somos agentes racionales, inmersos en la deliberación y en la justificación deliberativa. Simplemente no se requiere nada más para que haya racionalidad práctica.

Así, pues, no puede evitarse la crisis de fundamentación de la moralidad señalando la existencia de una práctica de justificación dentro del marco moral y negando que cualquier fundamento extra-moral es relevante. Pues ya está presente un modo de justificación extra-moral que no existe al lado de la justificación moral, sino que está ligado a la forma como unificamos nuestras creencias y preferencias; adquirimos así el sentido profundo de nosotros mismos. No hay por qué suponer que esta justificación deliberativa deba ser entendida en sí misma fundacionalmente. Lo único que necesitamos suponer es que la justificación moral plausiblemente no sobrevive al conflicto con ella.

### III

Al explicar por qué no es posible rechazar la idea de una crisis de fundamentación en la moralidad como resultado de una apelación fuera de lugar a una idea de fundacionalismo, filosóficamente desacreditada o sospechosa, comencé a exponer el carácter y las dimensiones de la crisis. He sostenido que la moralidad tiene ante sí un modo de justificación alternativo, conflictivo, más profundo, relacionado con nuestro sentido profundo de nosotros mismos, que vale para todo el ámbito de la elección y la acción y que evalúa cada *acción* en términos de los intereses reflexivamente sustentados por su *agente*. Me parece que está fuera de duda la relevancia de los intereses del agente para la justificación práctica. La relevancia de toda otra cosa, excepto en la medida en que afecte los intereses del agente, me parece sumamente dudosa. Si los intereses reflexivamente sustentados por el agente, sus preferencias, deseos y propósitos, son, conjuntamente con sus creencias meditadas, parte constitutiva de su autoconcepción, no puedo ver ninguna vía remotamente plausible de argumentar sobre su relevancia que no esté similarmente relacionada con su sentido de sí mismo. Y, en verdad, no puedo ver forma alguna de introducir algo que sea relevante para la justificación práctica, excepto a través de la autoconcepción del agente. Mi afirmación de este individualismo práctico no es un argumento concluyente, pero el peso de la prueba recae sobre quienes quisieran sostener una posición contraria. Que aporten los argumentos, si pueden.

La justificación deliberativa no refuta la moralidad. En verdad no le ofrece a la moralidad la cortesía de una refutación. Ignora la moralidad y, aparentemente, ocupa su lugar. Se apodera de la palestra de la justificación y, al parecer, no deja a la moralidad lugar a dónde agarrarse. Permítaseme que presente una comparación controvertida. La religión se enfrenta -y se ha enfrentado- con una crisis de fundamentación similar. La religión exige el culto de un ser divino que ordena teleológicamente el universo. Pero se ha visto confrontada con un modo alternativo de explicación. Aunque la emergencia de una teoría cosmológica, basada en una causalidad más eficiente que teleológica, advirtiera lo que iba a suceder, la suplantación de la teleología

en la biología -debido al éxito de la teoría de la evolución al ofrecer un modo de explicación que daba cuenta en términos causal-eficientes de la *apariencia* de un orden teleológico entre los seres vivientes- puede parecer que anuncia la muerte de la religión como empresa intelectualmente respetable. Pero la biología evolucionista y, de modo más general, la ciencia moderna no refutan la religión. Más bien la ignoran, sustituyendo sus explicaciones por otras ontológicamente más simples. Quizás la religión, entendida como algo que afirma el culto justificable de un ser divino, no sobreviva a su crisis de fundamentación. ¿Puede sobrevivir la moralidad, entendida como la afirmación de restricciones justificables a la elección, independientemente de los intereses del agente?

Parecería que la moralidad tiene tres vías para escapar del aparente destino de la religión. Una sería encontrar para los hechos morales o las propiedades morales un papel explicatorio que los atrincherara ante cualquier consideración de justificación<sup>12</sup>. Se podría argumentar entonces que todo tipo de justificación que ignorara consideraciones morales sería ontológicamente defectuoso. Menciono esta posibilidad tan sólo para dejarla de lado. Sin duda, hay personas que aceptarían las restricciones morales a sus elecciones y acciones si ignorásemos esto. Pero nuestra explicación de su comportamiento no tiene por qué obligarnos a adoptar su punto de vista. Aquí la comparación con la religión debería ser clara e incontrovertida. No podríamos explicar muchas de las prácticas de las personas religiosas sin hacer referencia a sus creencias. Pero el hecho de caracterizar, por ejemplo, como un acto de culto lo que una persona religiosa hace, no nos obliga a suponer que realmente existe el objeto del culto, aunque sí nos obliga a suponer que ella cree que tal objeto existe. De manera similar, caracterizar lo que un agente hace como el cumplimiento de un deber, por ejemplo, no nos obliga a suponer que existe deber alguno, aunque sí nos obliga a suponer que él cree que existen deberes. El escéptico, que no acepta ninguna de estas dos posiciones, puede tratar el papel manifiesto de la moralidad en la explicación de una manera similar al de la religión. Por supuesto, no considero que el paralelismo pueda, en última instancia, ser mantenido, ya que coincido con el escéptico religioso pero no con el escéptico moral. Pero, para establecer

---

<sup>12</sup> Esto respondería al desafío a la moralidad que encontramos en la cita anterior de Gilbert Harman.

un papel explicatorio de la moralidad, uno tiene primero que demostrar sus credenciales justificatorias. No se puede suponer que posea un papel explicatorio previo.

La segunda vía consistiría en reinterpretar la idea de justificación mostrando que, entendida más plenamente, la justificación deliberativa es incompleta y que tiene que ser complementada de forma tal que deje lugar para la moralidad. Existe una larga tradición en la filosofía moral -derivada principalmente de Kant- que está comprometida con esta empresa. No es este el momento para embarcarse en una crítica de aquello que -esperando nuevamente lograr una caracterización neutral- llamaré justificación universalista. Pero la crítica puede estar fuera de lugar. El éxito de la justificación deliberativa puede bastar. Pues las aseveraciones teóricas de su incompletitud parecen fracasar ante el simple reconocimiento práctico de que funciona. Por supuesto, manifiestamente, la justificación deliberativa no funciona para proporcionar un lugar a la moralidad. Pero suponer que debe hacerlo para ser totalmente adecuada o completa como modo de justificación, sería presuponer lo que está en discusión, es decir, que la justificación moral es defendible.

Si existieran valores objetivos independientemente de los propios deseos y objetivos y si, independientemente de los propios propósitos reales, uno fuera una parte de un orden objetivamente teleológico, tendríamos entonces razón en insistir en la inadecuación del marco deliberativo. Un orden objetivamente teleológico introduciría consideraciones relevantes para la justificación práctica, que no dependerían de la autoconcepción del agente. Pero la suplantación de la teleología en nuestras explicaciones físicas y biológicas cierra esta posibilidad, de la misma manera que cierra la posibilidad de la explicación religiosa.

Me centro ahora, pues, en la tercera vía para resolver la crisis de fundamentación de la moralidad. El primer paso consiste en aceptar la justificación deliberativa y reconocer que el lugar de la moralidad debe hallarse dentro, y no fuera, de su marco. Ahora bien, esto hace surgir de inmediato dos problemas. Ante todo, parecerá que el intento de establecer cualquier restricción a la elección y la acción, dentro del marco de una deliberación que apunta a la satisfacción máxima de las preferencias meditadas del agente, tiene que resultar imposible. Pero, aun si se duda de esto, parecerá que el intento de establecer una restricción *independientemente de las preferencias del agente*, dentro de un marco de este tipo, está al borde de la insensatez. Sin embargo, esta es precisamente la tarea aceptada por mi tercera vía. Y, a diferencia de sus predecesoras, creo que puede tener éxito; en

verdad, mi reciente libro, *Morals by Agreement* muestra cómo puede tener éxito<sup>13</sup>.

No me extenderé en un argumento que ahora es conocido, al menos por algunos lectores, y que, en todo caso, puede encontrarse en ese libro. Pero permítaseme esbozar brevemente aquellos rasgos de la racionalidad deliberativa que le permiten restringir la elección maximizante. La idea clave es que en muchas situaciones, si cada persona elige aquello que, dadas las elecciones de los demás, maximizaría su utilidad esperada, entonces el resultado será mutuamente desventajoso en comparación con alguna alternativa en la que todos podrían estar mejor<sup>14</sup>. El equilibrio que se obtiene cuando la acción de cada persona es la mejor respuesta a las acciones de los demás es incompatible con la optimalidad (paretiana) que se da cuando nadie podría estar mejor sin empeorar la situación de algún otro. Dada la ubicuidad de tales situaciones, cada persona puede percibir el beneficio que implica para ella misma el participar con sus congéneres en prácticas que exigen a cada cual que se abstenga de procurar maximizar directamente su propia utilidad cuando tal restricción mutua es mutuamente ventajosa. Por supuesto, nadie puede tener una razón para aceptar ninguna restricción unilateral de su comportamiento maximizador; cada cual se beneficia de, y sólo de, la restricción aceptada por sus congéneres. Pero si uno se beneficia más de una restricción impuesta a los demás que lo que uno pierde al ser restringido, uno puede tener una razón para aceptar una práctica que requiera que cada cual, incluyendo a uno mismo, asuma tal restricción. Podemos imaginar que una práctica tal puede lograr el acuerdo unánime entre personas racionales que eligen las condiciones bajo las cuales estarían dispuestas a interactuar. Y este acuerdo es la base de la moralidad.

Consideremos un ejemplo simple de una práctica moral que exigiría el acuerdo racional. Supongamos que cada uno de nosotros estaría dispuesto a ayudar a sus congéneres sólo si o bien uno espera beneficiarse por prestar ayuda, o bien uno tiene un interés directo en el bienestar de los demás. Por lo tanto, en muchas situaciones, las personas no ayudarían a las demás, aun

---

<sup>13</sup> Véase David Gauthier, *Morals by Agreement*, Oxford, 1986, esp. los capítulos V, VI.

<sup>14</sup> El ejemplo clásico de este tipo de situación es el Dilema del Prisionero: véase *Morals by Agreement*, págs. 79-80. De un modo más general, podría decirse que estas situaciones, en términos económicos, suponen un fracaso del mercado. Véase, por ejemplo, «Market Contractarianism», en Jules L. Coleman, *Markets, Morals and the Law*, Cambridge, 1988, cap. 10.

en el caso en que el beneficio del receptor excediera en gran medida el costo del dador, porque no existiría ninguna disposición para que el dador participe en el beneficio. Cada cual esperaría, entonces, estar mejor si cada cual prestara ayuda a sus congéneres, sin que importe su propio beneficio o interés, cada vez que el costo de la ayuda fuera bajo y el beneficio de recibir ayuda, considerable. Cada cual aceptaría de ese modo una restricción a la persecución directa de sus propios intereses, no unilateralmente, pero sí en el caso de que se diera una aceptación similar por parte de los otros. La reflexión nos lleva a reconocer que aquellos que pertenecen a grupos cuyos miembros se adhieren a esta práctica de asistencia mutua gozan de beneficios en la interacción que les están negados a los otros. Podemos entonces imaginar que tal práctica es racionalmente aceptable por cada cual.

Esta razón de la restricción acordada no hace referencia al contenido de las preferencias de cada cual. El argumento depende simplemente de la *estructura* de la interacción, de la forma como los esfuerzos de cada persona para satisfacer sus propias preferencias afecta a la satisfacción de todos los demás. Así, pues, la razón de cada persona para aceptar una práctica de restricciones mutuas es independiente de sus deseos, propósitos e intereses particulares, aunque no, por supuesto, del hecho de que ella tiene tales intereses. Creo que la idea de un agente puramente racional, que actúa sólo motivado por la razón, no es inteligible. La moralidad no ha de ser entendida como una restricción a la satisfacción de preferencias no racionales surgida tan sólo de la razón. Más bien, un agente racional es aquel que actúa para lograr la satisfacción máxima de sus preferencias, y la moralidad es una restricción a la forma como él actúa, que surge de los efectos de la interacción con otros agentes.

El «Tonto», de Hobbes, hace aquí su conocida entrada en escena e insiste en que, por más racional que pueda ser para una persona acordar con sus congéneres prácticas que prometen ventajas mutuas, sin embargo, es racional seguir tales prácticas sólo si ello conduce directamente a la satisfacción máxima de sus preferencias<sup>15</sup>. Pero, entonces, tales prácticas no imponen ninguna restricción real. El efecto de acordarlas o aceptarlas puede tan sólo ser el de cambiar los resultados esperados de las elecciones posibles, haciendo que para la persona sea racional aquello que, a falta de esta práctica, no maximizaría la utilidad. Las prácticas ofrecerían sólo verdadera prudencia, pero no verdadera moralidad.

---

<sup>15</sup> Véase Hobbes, *Leviathan*, Londres, 1651, cap. 15.

El tonto comete un doble error. En primer lugar, no entiende que la aceptación real de prácticas morales, tales como, por ejemplo, ayudar a los demás, o mantener las promesas, o decir la verdad, sólo es posible entre quienes están dispuestos a cumplirlas. Si mi disposición de cumplimiento sólo se extiende hasta donde llegan mis intereses o preocupaciones en el momento de actuar, entonces tú serías el verdadero tonto si interactúas conmigo de una manera tal que exige un cumplimiento más riguroso. Si, por ejemplo, es racional mantener las promesas sólo si el hacerlo maximiza directamente la utilidad, entonces entre personas cuya racionalidad es de conocimiento común sólo se harán aquellas promesas que requieren este cumplimiento limitado. Y, con ello, se perderían oportunidades de ventaja mutua.

Consideremos el siguiente ejemplo de la forma como las promesas facilitan el beneficio mutuo. Jones y Smith poseen granjas adyacentes. Aunque son vecinos y no alientan sentimientos de recíproca hostilidad, tampoco son amigos, de modo tal que ninguno de los dos experimenta ninguna satisfacción ayudando al otro. Sin embargo, reconocen que, si recogen sus cosechas juntos, los resultados serán mejores para ambos que si las recogen por separado. La semana próxima, la cosecha de Jones estará a punto para la siega; la de Smith lo estará quince días después. Una vez hecha la cosecha, Jones se retirará, venderá su granja y se irá a Florida, donde es difícil que se encuentre con Smith o con otros miembros de su comunidad. A Jones le gustaría prometer a Smith que, si le ayuda a cosechar la semana próxima, él ayudará a Smith a cosechar dentro de dos semanas. Pero tanto Jones como Smith saben que dentro de dos semanas, el ayudar a Smith no supondría más que un coste para Jones. Incluso si Smith le ayuda, no tiene nada que ganar al devolver la ayuda, ya que no lo motivan ni el interés por Smith ni, en estas circunstancias, la preocupación por su propia reputación. Por tanto, si Jones y Smith saben que Jones actúa directamente para maximizar la realización de sus preferencias, saben también que Jones no ayudará a Smith. Por ello, Smith no ayudará a Jones, incluso si Jones finge prometer ayuda a cambio de la que reciba. Sin embargo, Jones estaría mejor si hiciera y cumpliera una promesa tal. Y lo mismo cabe decir de Smith.

El segundo error del tonto, que se sigue del primero, debería ser claro: es incapaz de reconocer que, en circunstancias plausibles, las personas que están auténticamente dispuestas a un cumplimiento más riguroso de las prácticas morales que el que se seguiría de sus intereses en el momento de actuar, pueden esperar estar mejor que las que no están dispuestas a hacerlo.

Pues las primeras, a quienes llamaré maximizadores restringidos, serán socios bien recibidos en tareas de cooperación mutuamente ventajosas, en las que cada cual confía en la adhesión voluntaria de los demás; las otras personas, los maximizadores directos, serán excluidos de ellas. Por lo tanto, los maximizadores restringidos pueden esperar oportunidades más favorables que sus congéneres. Aunque al ayudar a los demás, al mantener sus promesas y al cumplir otras prácticas morales, renuncian parcialmente a la satisfacción que podrían obtener de sus preferencias, por contar con mejores oportunidades, estarán mejor, en última instancia, que aquellos que siempre maximizan la utilidad esperada.

Al identificar la moralidad con aquellas restricciones que obtendrían acuerdo entre personas racionales que eligen los términos de su interacción, me muevo en el plano de la reconstrucción racional. No supongo que hayamos acordado realmente las prácticas y principios morales existentes. Y tampoco supongo que todas las prácticas morales existentes obtendrían nuestro acuerdo en caso de que se planteara la cuestión. No todas las prácticas morales existentes necesitan ser justificables; no necesitan ser de un tipo tal que deberíamos cumplirlas voluntariamente. En verdad, ni siquiera supongo que las prácticas que debemos cumplir voluntariamente necesitan ser aquellas que contarían con nuestro acuerdo real. Supongo que las prácticas morales justificables son aquellas que obtendrían nuestro acuerdo *ex ante*, en una situación pre-moral adecuada. Son aquellas que deberíamos haber acordado como términos constitutivos de nuestra interacción futura si hubiéramos estado, *per impossibile*, en una posición que nos permitiera decidir estos términos. Por lo tanto, un acuerdo hipotético nos proporciona así un «test» de la justificabilidad de nuestras prácticas morales existentes.

#### IV

Muchas cuestiones podrían ser planteadas con respecto a esta versión, pero aquí deseo considerar sólo una de ellas. He sostenido que las prácticas morales son racionales -a pesar de que restringen el intento de cada persona de maximizar su propia utilidad- en la medida en que ellas serían objeto de un acuerdo unánime *ex ante*. Pero, para refutar al Tonto, tengo no sólo que defender la racionalidad del acuerdo, sino también la de su cumplimiento, y la defensa del cumplimiento amenaza debilitar el argumento en favor del acuerdo, de manera tal que el

título de este trabajo debería ser *¿Por qué restricción?* y no *¿Por qué contractualismo?* Es racional disponerse a aceptar ciertas restricciones a la maximización en las elecciones y en las acciones si y sólo si al hacerlo uno maximiza la utilidad esperada. ¿Cuál es, pues, la relevancia del acuerdo, especialmente del acuerdo hipotético? ¿Por qué debería ser racional disponerse a aceptar sólo aquellas restricciones que serían objeto de un acuerdo mutuo en una situación pre-moral apropiada, más que aquellas restricciones que encontramos en nuestras prácticas morales existentes? Ciertamente es la aceptación de estas últimas la que hace que una persona sea bien recibida en la interacción con sus congéneres. Pues el cumplimiento de la moralidad existente es lo que ellos esperan y toman en cuenta cuando eligen socios con quienes cooperar.

Comencé con un desafío a la moralidad: ¿cómo puede ser racional para nosotros aceptar sus restricciones? Puede ahora parecer que lo que he mostrado es que, en verdad, es racional aceptar restricciones, pero aceptarlas sin que importe el que ellas puedan ser plausiblemente consideradas como morales. Puede parecer que la moralidad no tiene nada que ver con mi argumento: lo que he mostrado es que es racional estar dispuesto a cumplir cualesquiera restricciones que son generalmente aceptadas y esperadas, sin que importe su naturaleza. Pero ésta no es mi concepción.

Para mostrar la relevancia del acuerdo con respecto a la justificación de las restricciones, supongamos una sociedad existente en la que los individuos más o menos reconocen y cumplen el conjunto dado de prácticas que restringen sus elecciones en relación con lo que serían si tomaran directamente en cuenta sólo sus deseos, propósitos e intereses. Supongamos que una disposición a cumplir estas prácticas existentes sea *prima facie* ventajosa, ya que las personas que no están así dispuestas pueden esperar ser excluidas por sus congéneres de oportunidades deseables. Sin embargo, las prácticas mismas no tienen, o al menos no necesitan tener, ninguna base de acuerdo. Y no necesitan satisfacer ninguna pauta intuitiva de equidad o de imparcialidad, características que podemos suponer relevantes para la identificación de las prácticas como genuinamente morales. A pesar de que podemos decir que estas prácticas constituyen la moralidad de la sociedad en cuestión, no necesitamos considerar que ellas están moralmente justificadas o que son moralmente aceptables. Simplemente son prácticas que restringen el comportamiento individual de una manera que cada cual encuentra racional aceptar.

Supongamos ahora que nuestras personas, como maximizadores

racionales de la utilidad individual, se ponen a reflexionar sobre las prácticas que constituyen su moralidad. Por supuesto, evaluarán las prácticas en relación con su propia utilidad, pero sabiendo que sus congéneres hacen lo mismo. Una pregunta que entonces tiene que surgir es ésta: ¿por qué estas prácticas? Pues reconocerán que el conjunto de las prácticas morales reales no es el único conjunto posible de prácticas restrictivas que conducirían a resultados óptimos mutuamente ventajosos. Reconocerán la posibilidad de órdenes morales alternativos. En este punto, no bastará decir que, de hecho, cada persona puede esperar beneficiarse de una disposición a cumplir las prácticas existentes. Pues las personas se preguntarán también a sí mismas: ¿no puedo beneficiarme más, no abandonando simplemente toda moralidad y desconociendo toda restricción, sino rechazando parcialmente las restricciones existentes en favor de un conjunto alternativo? Una vez que se ha planteado esta cuestión, la situación se transforma: el orden moral existente tiene que ser evaluado, no sólo frente a su simple incumplimiento, sino frente a lo que podríamos llamar cumplimiento alternativo.

Para hacer esta evaluación, cada cual comparará sus perspectivas bajo las prácticas existentes con aquéllas que podría anticipar de un conjunto que, en las actuales circunstancias, podría esperar como resultado de una negociación con sus congéneres. Si sus perspectivas mejorarían con tal negociación, entonces tendrá un incentivo real, aunque no necesariamente suficiente, para exigir un cambio en el orden moral establecido. De modo más general, si hay personas cuyas perspectivas mejorarían con una renegociación, entonces el orden existente será manifiestamente inestable. Sin duda, aquellos cuyas perspectivas empeorarían con la renegociación tendrán un claro incentivo para resistir, para apelar al *status quo*. Pero su apelación será débil, especialmente entre las personas que no se dejan embaucar por consideraciones ideológicas espurias sino que se orientan hacia la maximización de la utilidad individual. Así, aunque en el mundo real comenzamos con un conjunto existente de prácticas morales como restricciones a nuestro comportamiento maximizador, la reflexión nos conduce a la idea de un conjunto modificado que obtendría el acuerdo de cada cual, y este conjunto modificado tiene -y se reconocerá que tiene- una estabilidad de la que carece la moralidad existente.

La capacidad reflexiva de los agentes racionales les conduce de lo dado a lo acordado, de las prácticas y principios existentes que imponen restricciones, a aquéllas que obtendrían el consentimiento de cada persona. Sostengo que la misma capacidad reflexiva

conduce de aquellas prácticas que serían acordadas en las circunstancias sociales existentes, a aquellas que recibirían acuerdo *ex ante*, pre-moral y pre-social. Así como el *status quo* demuestra ser inestable cuando entra en conflicto con lo que sería acordado, así también lo que sería acordado demuestra ser inestable cuando entra en conflicto con lo que hubiera sido acordado en un contexto pre-social adecuado. Pues, así como las prácticas existentes tienen que parecer arbitrarias en la medida en que no respondan a aquello que una persona racional acordaría, así también lo que tal persona acordaría en las circunstancias existentes tiene que parecer arbitrario en relación con lo que aceptaría en una condición pre-social.

Lo que una persona racional acordaría en circunstancias existentes depende, en gran medida, de su posición de negociación frente a sus congéneres. Pero su posición de negociación está significativamente afectada por las instituciones sociales existentes y, con ello, por las prácticas morales comúnmente aceptadas e incorporadas en esas instituciones. Así, aunque el acuerdo pueda muy bien proporcionar prácticas diferentes de aquellas incorporadas en las instituciones sociales existentes, sin embargo, será influenciado por aquellas prácticas que no son en sí mismas el producto de un acuerdo racional. Y esto tiene que poner en tela de juicio la racionalidad de las prácticas acordadas. La arbitrariedad de las prácticas existentes tiene que infectar todo acuerdo cuyos términos están afectados significativamente por ellas. Aunque, en sí mismo, el acuerdo racional es una fuente de estabilidad, sin embargo, esta estabilidad es socavada por la arbitrariedad de las circunstancias en las cuales se lleva a cabo. Para evitar esta arbitrariedad, las personas racionales pasarán del acuerdo real al hipotético, considerando cuáles prácticas ellas hubieran acordado en una posición no estructurada por las instituciones existentes y las prácticas que ellas representan.

El contenido de un acuerdo hipotético está determinado por una apelación a la racionalidad igual de las personas. Personas racionales aceptarán voluntariamente un acuerdo sólo en la medida en que ellas perciban que es igualmente ventajoso para cada una de ellas. Ciertamente, a cada cual le gustaría aceptar un acuerdo más ventajoso para ella misma que para los congéneres, pero como ninguna aceptará un acuerdo percibido como menos ventajoso, los agentes cuya racionalidad es conocida reconocerán la futilidad de pretender o insistir en obtener más y minimizarán sus costes de negociación coordinándose en el punto de igual ventaja. Ahora bien, la extensión de la ventaja es determinada de una doble manera. Primero, hay una ventaja interna al

acuerdo. En este respecto, la expectativa de igual ventaja es asegurada por la equidad procesal. El paso de las prácticas morales existentes a aquellas que resultan de un acuerdo real lleva a personas racionales a una situación procesalmente equitativa, en la cual cada una percibe que las prácticas acordadas son aquellas que es igualmente racional para todas aceptar, dadas las circunstancias en las cuales el acuerdo es alcanzado. Pero aquellas mismas circunstancias pueden ser cuestionadas en la medida en que se perciba que son arbitrarias: el resultado, en parte, del cumplimiento de prácticas restrictivas que no aseguran ellas mismas la expectativa de igual ventaja y, por ello, no reflejan la racionalidad igual de las personas que las cumplen. Para neutralizar este elemento de arbitrariedad, para que las prácticas morales sean plenamente aceptables, tienen que ser concebidas como constituyendo el resultado posible de un acuerdo hipotético bajo circunstancias que no están afectadas por instituciones sociales que, en sí mismas, carecen de aceptabilidad plena. La igual racionalidad exige la consideración, tanto de circunstancias externas como de procedimientos internos.

Pero, ¿cuál es la relevancia práctica de este argumento? Sería absurdo pretender que el mero conocimiento del mismo o hasta su aceptación, habrá de conducir a la sustitución de las prácticas morales existentes por aquellas que establecería un acuerdo pre-social. Sería irracional para cualquiera renunciar a los beneficios del orden moral existente simplemente porque se da cuenta que le proporciona más que lo que podría esperar de un acuerdo puramente racional con sus congéneres. Y sería irracional para cualquiera aceptar una pérdida de utilidad a largo plazo, negándose a cumplir el orden moral existente, simplemente porque se da cuenta que tal cumplimiento le proporciona menos que lo que podría esperar de un acuerdo puramente racional. Sin embargo, estos conocimientos transforman -o quizás ponen de manifiesto- el carácter de las relaciones entre personas que están sujetas a las restricciones existentes, de forma tal que algunas de estas relaciones son reconocidas como coactivas. Estos conocimientos constituyen la eliminación de la falsa conciencia y resultan de un proceso de reflexión racional que lleva a las personas a algo que, en mi teoría, es similar a la situación ideal de diálogo de Jürgen Habermas<sup>16</sup>. Sin un argumento para defenderse en el diálogo abierto con sus congéneres, aquellos que están

---

<sup>16</sup> Véase Raymond Geuss, *The Idea of a Critical Theory: Habermas & the Frankfurt School*, Cambridge, 1981, págs. 65 y sigs.

más que igualmente aventajados pueden esperar mantener su posición privilegiada sólo si pueden coaccionar a sus congéneres para que la acepten. Y, por supuesto, esto puede ser posible. Pero la coacción no es acuerdo y carece de toda estabilidad inherente.

La estabilidad juega un papel clave en la vinculación del cumplimiento con el acuerdo. Conscientes de los beneficios que han de obtener de las prácticas restrictivas, las personas racionales buscarán aquéllas que invitan a un cumplimiento estable. Ahora bien, el cumplimiento es estable si surge del acuerdo entre personas, cada una de las cuales considera que los términos del acuerdo son lo suficientemente favorables para ella misma como para que sea racional aceptarlas y que no son tan favorables para las otras, que sería racional para ellas aceptar términos menos favorables para ellas y más favorable para ella misma. Un acuerdo que proporcione términos igualmente favorables para todos invita, como ningún otro, a un cumplimiento estable.

## V

Al sostener que las prácticas morales, para lograr el cumplimiento voluntario estable de individuos racionales, tienen que ser los objetos de un acuerdo hipotético adecuado, he agregado algo más a la caracterización inicial mínima de la moralidad. La moralidad no sólo restringe nuestras elecciones y acciones sino que lo hace de manera imparcial, reflejando la racionalidad igual de las personas sujetas a la restricción. A pesar de que no es parte de mi argumento mostrar que los requisitos de la moralidad contractualista satisfarán el «test» rawlsiano de coherencia con nuestros juicios meditados en el equilibrio reflexivo, sería, sin embargo, erróneo tratar las restricciones a la maximización directa de utilidad racionalmente acordadas como constituyendo una moralidad más que como reemplazando la moralidad, si no fuera por la adecuación entre su contenido y nuestros puntos de vista morales pre-teóricos. Sugiero que la adecuación reside en la imparcialidad requerida para el acuerdo hipotético.

Así pues, la crisis fundacional de la moralidad se resuelve mostrando la racionalidad de nuestro cumplimiento de las restricciones a la persecución de nuestros deseos, propósitos e intereses acordadas mutua y racionalmente. Aunque desprovista de una base de valores objetivos o de un orden objetivamente teleológico, y confrontada con un modo más fundamental de justificación, la moralidad sobrevive al incorporarse ella misma a ese

modo. Las consideraciones morales tienen el mismo *status* y juegan el mismo papel en la explicación del comportamiento que las otras razones reconocidas por un deliberador racional. Nos queda una versión unificada de la justificación en la cual las elecciones y acciones de un agente son evaluadas en relación con sus preferencias, con los intereses que son parte constitutiva de su sentido de sí mismo. Pero, como la moralidad obliga al agente, independientemente del contenido particular de sus preferencias, tiene la fuerza prescriptiva que le confieran las concepciones cristiana y kantiana.

Al incorporar la moralidad a la justificación deliberativa, reconocemos una nueva dimensión en la autoconcepción del agente. Pues la moralidad requiere que una persona tenga la capacidad de obligarse a sí misma, de llegar a un acuerdo con sus congéneres, con la seguridad que les proporciona la conciencia de que puede cumplir y cumplirá su parte en el acuerdo, sin tomar en cuenta muchas de aquellas consideraciones que normal y justificadamente entrarían en sus futuras deliberaciones. Y esto es más que la capacidad de colocar en un único todo coherente sus deseos e intereses con sus propias creencias. A pesar de que esta última capacidad unificadora tiene que extender su atención al pasado y al futuro, la unificación que ella logra puede ser, ella misma, restringida a aquel presente extendido en el cual la persona juzga y decide. Pero, al obligarse uno mismo a una acción futura acorde con los acuerdos de uno mismo, uno tiene, al menos, que fijar un subconjunto de los propios deseos y creencias que serán sostenidos en ese futuro. El sí mismo que presta el acuerdo y el sí mismo que cumple debe ser uno. «Uno mismo tiene primeramente que haberse vuelto *calculable, regular, necesario*, también en su propia imagen de sí mismo, para ser capaz de ser el garante de *su propio futuro*, que es lo que hace quien promete»<sup>17</sup>.

Al desarrollar «*el derecho a formular promesas*»<sup>18</sup>, los seres humanos hemos encontrado un baluarte contractualista en contra de la muerte de la moralidad.

(Trad. de Silvia Mendlewicz y Albert Calsamiglia)

---

<sup>17</sup> Nietzsche, *On the Genealogy of Morals*, trad. de Walter Kaufmann y R. J. Hollingdale, Random House, Nueva York, 1967, Segundo Ensayo, sección 1, pág. 58.

<sup>18</sup> *Ibidem*, pág. 57.

