

CONTRACTUALISMO Y UTILITARISMO*¹

Thomas M. Scanlon

En este artículo de 1982, Thomas Scanlon esboza una explicación filosófica de los deberes morales recíprocos que alcanza su exposición completa, con algunos cambios, en 1998 con el libro *What We Owe To Each Other*. Tal explicación, presentada como una forma de “contractualismo”, caracteriza el razonamiento y la motivación moral como la voluntad de justificar ante otro las propias acciones según principios que nadie podría rechazar razonablemente para re-

THOMAS M. SCANLON. (Indianapolis, 1940-) Profesor Alford de religión natural, filosofía moral y política civil en el Departamento de Filosofía de la Universidad de Harvard. Primero se interesó en lógica-matemática (fue discípulo de Paul Benacerraf), pero pronto se dedicó a la filosofía moral y filosofía política. Ha desarrollado una versión del contractualismo similar a las de Rawls, Kant y Rousseau. Tiene además artículos sobre libertad de expresión, derechos civiles y el concepto de bienestar, algunos de los cuales están recogidos en su último libro: *The Difficulty of Tolerance* (2003).

* T. M. Scanlon, “Contractualism and Utilitarianism”, publicado en el volumen colectivo *Utilitarianism and Beyond*, editado por Bernard Williams y Amartya Sen (Maison de las Sciences de l’Homme-Cambridge University Press, 1982). El artículo fue reeditado en la colección de ensayos del propio Scanlon: *The Difficulty of Tolerance* (Cambridge University Press, 2003). Su publicación en castellano en esta edición cuenta con la debida autorización. La traducción a partir del texto publicado en *Utilitarianism and Beyond*, estuvo a cargo de Pablo Solari, con la colaboración de Nicolás Goic.

¹ Estoy muy en deuda con Derek Parfit por la paciente crítica y la tremendamente útil discusión de muchas versiones previas de este documento. Gracias se adeudan también a las muchas audiencias que han escuchado partes de aquellas versiones entregadas como conferencias y respondido amablemente con útiles comentarios. En particular, estoy en deuda con Marshall Cohen, Ronald Dworkin, Owen Fiss y Thomas Nagel por la valiosa crítica.

gular la conducta colectiva. De esta explicación se derivan, potencialmente, implicaciones para el discurso normativo de primer orden. El autor ofrece asimismo una reflexión meta-teórica que arroja ciertos criterios básicos para construir y evaluar teorías morales. Siguiendo esos criterios, el contractualismo aquí presentado sería preferible tanto a la concepción utilitarista basada en la agregación del bienestar de una pluralidad de individuos, como a las estrategias de reconstrucción de la justificación moral en términos de la teoría de la elección racional.

El utilitarismo ocupa un lugar central en la filosofía moral de nuestro tiempo. No es la doctrina que sostiene la mayoría; seguramente, muy pocos se declararían “utilitaristas de la acción”. Sin embargo, para muchos es una posición a la que se ven arrastrados cuando intentan responder teóricamente por sus opiniones morales. Dentro de la filosofía moral, representa una posición que hay que combatir si se quiere evitar adoptarla. Esto es así a pesar de que las implicaciones del utilitarismo de la acción difieren ostensiblemente de convicciones morales firmemente arraigadas, mientras que el utilitarismo de la regla (la más común entre las formulaciones alternativas) le parece a la mayoría de la gente una inestable componenda.

Pienso que el gran atractivo del utilitarismo se debe a consideraciones filosóficas más o menos sofisticadas, que nos empujan en una dirección muy distinta de nuestras convicciones morales de primer orden. En particular, mucho de su atractivo deriva de presuntas dificultades en los fundamentos de las visiones rivales. Lo que una exitosa alternativa al utilitarismo debe hacer es, ante todo, agotar esta fuente de energía ofreciendo una explicación clara de los fundamentos del razonamiento moral no-utilitarista. En lo que sigue, describiré el problema en detalle, planteando las preguntas que debe responder una relación filosófica de los fundamentos de la moralidad. Luego, esbozaré una versión del contractualismo que, según argüiré, ofrece un mejor grupo de respuestas que las suministradas por las versiones más llanas del utilitarismo. Finalmente, explicaré por qué el contractualismo, como lo entiendo yo, no reconduce en sus resultados normativos a alguna fórmula utilitarista.

El contractualismo ya fue propuesto anteriormente como alternativa al utilitarismo por John Rawls en *Una Teoría de la Justicia* (1971). A pesar de la amplia discusión que este libro ha recibido, creo que se ha subestimado el atractivo del contractualismo como doctrina fundacional. En particular, no se ha valorado de modo suficiente la oferta que hace el contractualismo para una explicación plausible de la motivación moral. La versión del con-

tractualismo que presentaré difiere de la de Rawls en una cantidad de aspectos. Específicamente, no utiliza la noción de una elección bajo un velo de ignorancia, o hace un uso diferente y más limitado de ella. Una consecuencia de esta diferencia consiste en destacar de manera clara el contraste entre contractualismo y utilitarismo.

I

Un área de estudio tal como la filosofía moral existe por la misma razón que existe un área tal como la filosofía de las matemáticas. En los juicios morales así como en los juicios matemáticos, tendemos a otorgar cierto grado de confianza e importancia a un conjunto de creencias putativamente objetivas. Pero, al reflexionar, no es para nada obvio a qué pueden referirse esos juicios, o en virtud de qué puede decirse que algunos son correctos o defendibles y otros no. La cuestión sobre el objeto de estos juicios, o sobre los fundamentos de su verdad, es la primera cuestión filosófica tanto en la moralidad como en las matemáticas. Segundo, en ambas áreas parece ser posible descubrir la verdad simplemente al pensar o razonar al respecto. La experiencia y la observación pueden ayudar, pero la observación, en el sentido normal de la palabra, no es el medio corriente de descubrimiento en ambos ámbitos. Así, dada cualquier respuesta positiva a la primera cuestión —cualquier especificación del objeto o fundamento de verdad en matemáticas o moralidad—, necesitamos una epistemología compatible que explique cómo es posible descubrir los hechos sobre estos objetos mediante los medios que, aparentemente, usamos para ello.

Dada esta similitud en las cuestiones que dan origen a la filosofía moral y a la filosofía de las matemáticas, no es sorprendente que las respuestas dadas comúnmente caigan en tipos generales semejantes. Si entrevistáramos estudiantes de un curso novato de matemáticas, muchos de ellos se declararían a favor de algún tipo de convencionalismo. Sostendrían que las matemáticas proceden de definiciones y principios que son arbitrarios o justificados instrumentalmente, y que el razonamiento matemático consiste en percibir lo que se sigue de estas definiciones y principios. Otros, unos pocos, serán quizá realistas o platónicos, para quienes las verdades matemáticas son un tipo especial de hecho no-empírico que podemos percibir a través de cierta forma de intuición. Otros serán naturalistas que sostienen que la matemática, entendida apropiadamente, es sólo la ciencia empírica más abstracta. Finalmente, están aquellos (aunque probablemente no en un curso novato promedio) que sostienen que no hay hechos matemáticos en el mundo “afuera de nosotros”, pero que las verdades matemáti-

cas son verdades objetivas sobre las construcciones mentales de que somos capaces. Kant sostenía que la matemática pura era un ámbito de verdades objetivas dependientes de la mente, y el intuicionismo matemático de Brouwer es otra teoría de este tipo (con la importante diferencia de ofrecer fundamentos, más que para la verdad en sentido clásico de los juicios matemáticos, para garantizar su aseverabilidad). Todas estas posiciones tienen sus correlatos naturales en la filosofía moral. El intuicionismo del tipo expuesto por W. D. Ross es quizá el análogo más próximo al platonismo matemático, y la teoría de Kant es la versión más familiar de la tesis de que la moralidad es una esfera de verdades objetivas dependientes de la mente.

Todas las visiones que he mencionado (con algunas reservas en el caso del convencionalismo) dan respuestas positivas (esto es, no-escépticas) a la primera pregunta filosófica sobre las matemáticas. Cada una identifica algún fundamento de verdad objetivo, o al menos intersubjetivo, para los juicios matemáticos. El escepticismo rotundo y las versiones subjetivas de la dependencia mental (análogos del emotivismo o prescriptivismo) son menos atrayentes como filosofías matemáticas que como filosofías morales. En parte, esto es así por el mayor grado de acuerdo intersubjetivo que existe sobre los juicios matemáticos. Pero también debido a cuestiones ulteriores que la reflexión filosófica, en uno y otro campo, debe enfrentar.

Ni las matemáticas ni la moralidad pueden entenderse como describiendo un ámbito de hechos que existen con independencia del resto de la realidad. Se supone que cada una de ellas está conectada con otras cosas. Los juicios matemáticos dan lugar a predicciones sobre aquellos ámbitos a los que las matemáticas se aplican. Esta conexión es algo que una relación filosófica de la verdad matemática tiene que explicar, pero el que podamos observar y aprender de la exactitud de tales predicciones también apoya nuestra creencia en la verdad matemática objetiva. En el caso de la moralidad, la conexión principal es, o se supone generalmente que es, con la voluntad. Dado un candidato cualquiera para el rol de objeto de la moralidad, debemos explicar por qué alguien habría de interesarse en él, y la necesidad de responder a esta pregunta por la motivación otorga un fuerte apoyo a las visiones subjetivistas.

Pero, ¿qué debe decir sobre la motivación moral una teoría filosófica adecuada de la moralidad? Pienso que no necesita mostrar que la verdad moral da, a la persona que la conoce, una razón para actuar que apele a sus deseos presentes o al progreso de sus intereses. Hallo del todo comprensible que la exigencia moral se aplique correctamente a una persona aunque ella no tenga ninguna razón de ese tipo para cumplir con esa exigencia. Es una cuestión disputada (que habré de dejar de lado) si las exigencias mora-

les dan, a quienes se aplican, un tercer tipo de razones para cumplir con ellas. Pienso que lo que una filosofía moral adecuada debe hacer, es aclararnos la naturaleza de las razones que la moralidad provee de por sí, o al menos, las que provee para los que se preocupan de ella. Una teoría filosófica de la moralidad tiene que ofrecer una explicación de estas razones que sea, por un lado, compatible con su explicación de la verdad y el razonamiento morales, y por otro lado, apoyada por un análisis plausible de la experiencia moral. Una filosofía moral satisfactoria no dejará la preocupación por la moralidad como una simple preferencia especial, como un fetiche o gusto idiosincrásico de cierta gente. Tiene que hacer comprensible por qué la gente puede tomar en serio las razones morales, y por qué a quienes se ven compelidos por ellas les parece que son razones especialmente rigurosas e insoslayables.

También hay una cuestión ulterior sobre si ser susceptible a tales razones es compatible con el bien de una persona o si es, como argumenta Nietzsche, un desastre psicológico para ella. Al defender la moralidad, uno tiene que mostrar que no es desastrosa en este sentido, pero no trataré aquí esta segunda cuestión motivacional. La menciono sólo para distinguirla de la primera cuestión, que es mi presente preocupación.

La tarea de dar una explicación filosófica de los contenidos de la moralidad difiere tanto del análisis semántico de los términos morales como del intento de encontrar la formulación más coherente de nuestras convicciones morales de primer orden. Una organización máximamente coherente de nuestras convicciones morales de primer orden podría proveernos de un valioso tipo de explicación: aclararía cómo varios preceptos, nociones y juicios morales aparentemente dispares están relacionados entre sí, indicando así hasta qué grado los conflictos entre ellos son cuestiones fundamentales y hasta qué grado pueden ser resueltos o explicados convincentemente. Pero la investigación filosófica sobre el objeto de la moralidad adopta una mirada más externa. Busca explicar qué tipo de verdades son las verdades morales describiéndolas en relación con otras cosas en el mundo y con nuestras preocupaciones particulares. Una relación de cómo podemos saber la verdad sobre la moralidad tiene que basarse en una explicación externa de la clase de cosas que son las verdades morales, más que en una lista de verdades morales particulares, aunque sea una lista máximamente coherente. Asimismo, esto parece válido también para las explicaciones acerca de cómo las convicciones morales pueden darle a uno razones para actuar².

² Aunque aquí son más importantes los lazos entre la naturaleza de la moralidad y su contenido. No es claro que una relación de la naturaleza de la moralidad que deja su contenido *enteramente* abierto podría ser la base de una relación plausible de la motivación moral.

La coherencia entre nuestras convicciones morales de primer orden —lo que Rawls ha llamado equilibrio reflexivo en sentido estrecho³— parece insatisfactoria⁴ como explicación de la verdad moral o como relación de las bases de justificación en la ética sólo porque, tomada en sí misma, una relación máximamente coherente de nuestras convicciones morales no necesita proveernos de lo que yo he llamado una explicación filosófica del objeto de la moralidad. Sin importar lo internamente coherentes que puedan volverse nuestras convicciones morales, siempre será posible sospechar que no tienen ninguna conexión relevante con algo. Ellas pueden ser un mero conjunto de reacciones socialmente inculcadas, quizá consistentes entre sí, pero no juicios de un género tal que pueda decirse propiamente que son correctos o incorrectos. Una teoría filosófica de la naturaleza de la moralidad puede contribuir a acrecentar la confianza que tenemos en convicciones morales de primer orden, principalmente, disipando estas dudas sobre el objeto de estudio. En tanto que incluye una epistemología moral, una teoría filosófica nos puede guiar hasta nuevas formas de argumentación moral, pero no necesita hacerlo. El tipo de argumentación moral con el que estamos más o menos familiarizados, puede seguir siendo la única forma de justificación en ética. Pero sea que conduzca o no a revisar nuestros modos de justificación, una buena teoría filosófica debería aclarar más nuestra comprensión de cuáles son las mejores formas de argumentación moral, y qué tipo de verdad es esa que alcanzamos mediante estas formas. (En gran medida puede decirse lo mismo, creo yo, de la contribución que hace la filosofía de las matemáticas a nuestra confianza en juicios matemáticos particulares y en particulares formas del razonar matemático.)

Como cualquier tesis sobre moralidad, una relación filosófica del objeto de la moralidad tiene que tener alguna conexión con el significado de los términos morales: debe ser plausible alegar que el objeto descrito es en realidad aquello a lo que estos términos refieren, al menos en muchos de sus usos normales. Pero el significado corriente de los términos morales es el producto de muchas convicciones morales diferentes, sostenidas por hablantes pasados y presentes, y este significado es compatible con una va-

³ Véase Rawls, J.: "The Independence of Moral Theory" (1974-1975); Daniels, N.: "Wide Reflective Equilibrium and Theory of Acceptance in Ethics". Cuán próximos estarán el proceso de lo que yo llamo explicación filosófica y la búsqueda de "equilibrio reflexivo amplio", como es entendido por Rawls y por Daniels, es una cuestión que no puedo discutir aquí.

⁴ Para expresiones de esta insatisfacción, véase Singer, P.: "Sidgwick and Reflexive Equilibrium", 1974, y Brandt, R.: *A Theory of the Good and The Right*, 1979.

riedad de visiones morales y con una variedad de visiones de la naturaleza de la moralidad. Después de todo, los términos morales son usados para expresar muchas visiones de este tipo diferentes, y la gente que expresa estas visiones no está usando incorrectamente los términos morales, aunque pueda estar errado lo que alguien diga. Igual que un juicio moral de primer orden, una caracterización filosófica del objeto de la moralidad es ya un planteamiento sustantivo sobre la moralidad, aunque un género diferente de planteamiento.

En tanto que una caracterización filosófica de la moralidad hace un género de afirmación que difiere de un juicio moral de primer orden, no implica esto que una teoría filosófica de la moralidad haya de ser neutral entre doctrinas normativas rivales. La adopción de una tesis filosófica sobre la naturaleza de la moralidad tendrá, casi siempre, algún efecto sobre la plausibilidad de afirmaciones morales particulares, pero las teorías filosóficas de la moralidad varían ampliamente en la extensión e inmediatez de sus implicaciones normativas. En un extremo está el intuicionismo, entendido como la tesis filosófica según la cual la moralidad se ocupa de ciertas propiedades no-naturales. Ross⁵ considera que la rectitud, por ejemplo, es la propiedad de “adecuación” o “pertinencia moral”. El intuicionismo sostiene que podemos identificar las instanciaciones de estas propiedades, y que podemos reconocer como auto-evidentes ciertas verdades generales sobre ellas, pero que no podemos analizar ulteriormente estas propiedades o explicarlas en términos de otras nociones. Así entendido, el intuicionismo es, en principio, compatible con una gran variedad de posiciones normativas. Según cuales sean las verdades generales sobre esa propiedad llamada rectitud moral que se tomen como auto-evidentes, alguien podría ser, por ejemplo, un intuicionista utilitarista o un intuicionista que cree en los derechos morales.

El otro extremo está representado por el utilitarismo filosófico. El término “utilitarismo” es usado, generalmente, para referirse a una familia de doctrinas normativas específicas —doctrinas que pueden ser defendidas sobre la base de un número de tesis filosóficas diferentes sobre la naturaleza de la moralidad. En este sentido del término, uno podría ser utilitarista, por ejemplo, sobre la base de razones intuicionistas o contractualistas. Pero lo que yo llamo “utilitarismo filosófico” es una tesis filosófica particular sobre el objeto de la moralidad, a saber, la tesis de que los únicos hechos

⁵ Ross, W. D.: *The Foundations of Ethics*, 1939, pp. 52-54, 315. [Traducción al castellano: *Fundamentos de Ética*, Eudeba, 1972.]

morales fundamentales son hechos sobre el bienestar individual⁶. Creo que esta tesis resulta muy plausible para mucha gente, y que mientras algunos son utilitaristas por otras razones, es el atractivo del utilitarismo filosófico lo que responde por la extendida influencia de los principios utilitaristas.

A la gente le parece evidente que existe algo así como que la situación de los individuos puede mejorar o empeorar. Tales hechos tienen una fuerza motivacional obvia; es del todo comprensible que la gente se vea motivada por ellos de modo muy semejante a cómo, se supone, está motivada por consideraciones morales. Más aún, estos hechos son claramente relevantes para la moralidad tal como la entendemos hoy. Las afirmaciones sobre el bienestar individual son un tipo válido de punto de partida para la argumentación moral. Pero mucha gente halla muy difícil ver cómo podría siquiera haber otros puntos de vista, puntos de vista independientes. Los requerimientos morales sustantivos, independientes del bienestar individual, les parecen una forma objetable de intuicionismo, donde estos requerimientos representarían “hechos morales” difíciles de explicar. No hay problema en reconocer como un hecho el que cierto acto es, por ejemplo, una instancia de mentir o romper una promesa. Y un utilitarista puede reconocer que tales hechos tienen, a menudo, importancia moral (derivada): son moralmente significativos por sus consecuencias en relación con el bienestar individual. Los problemas, y el cargo de “intuicionismo”, surgen cuando se afirma que tales actos son incorrectos en un sentido irreducible al hecho de disminuir el bienestar individual. ¿Cómo pudo esta propiedad independiente, ser moralmente incorrecto, ser entendida de modo que se le diera el tipo de importancia y fuerza motivacional que se concede a las consideraciones morales? Si uno acepta la idea según la cual no hay propiedades morales que tengan este tipo de importancia intrínseca, entonces, el utilitarismo puede parecer la única relación sostenible de la moralidad. Y una vez que el utilitarismo filosófico es aceptado, parece imponérsenos una forma de utilitarismo normativo como la correcta teoría moral de primer orden. El utilitarismo tiene así, para muchos, algo de la importancia que tienen el formalismo de Hilbert y el intuicionismo de Brouwer para sus respectivos partidarios. Es una visión que parece imponérsenos por la necesidad de dar una respuesta filosóficamente defendible sobre el objeto. Pero todo esto nos deja ante una dura elección: o podemos abandonar muchas de nuestras convicciones morales de primer orden o podemos tratar de salvarlas, ya sea

⁶ Para los propósitos de esta discusión dejo abierta la importante pregunta por los individuos que han de contar y cómo se ha de entender “bienestar”. El utilitarismo filosófico conservará el atractivo que me ocupa bajo muchas respuestas diferentes a estas cuestiones.

obteniéndolas como verdades derivadas ya sea desmontándolas como ficciones útiles e inofensivas.

Podría parecer que el atractivo del utilitarismo filosófico, como yo lo he descrito, es espurio, tan pronto como esta teoría tiene que ser equivalente a una forma de intuicionismo (diferente de otras sólo en que implica apelar una única vez a la intuición), o al tipo de naturalismo definicional ya refutado hace mucho por Moore y otros. Pero no pienso que se pueda desechar la doctrina tan fácilmente. El utilitarismo filosófico es una tesis filosófica sobre la naturaleza de la moralidad. Como tal, está a la par con el intuicionismo o con la forma de contractualismo que después defenderé en este documento. Ninguna de estas tesis necesita que su verdad sea un asunto de definición; si una de ellas es verdadera, de ahí no se sigue que una persona que la niegue esté usando mal las palabras, “correcto”, “incorrecto”, y “debe”. Tampoco todas estas tesis son formas de intuicionismo, si el intuicionismo es entendido como la visión de que los hechos morales tienen que ver con propiedades especiales, no-naturales, que podemos aprehender por captación intuitiva pero que no necesitan ni admiten ningún análisis ulterior. Tanto el contractualismo como el utilitarismo filosófico son, de modo específico, incompatibles con esta afirmación. El contractualismo y el utilitarismo filosófico, como otras tesis filosóficas sobre la naturaleza de la moralidad (incluyendo, diría yo, el mismo intuicionismo), han de ser apreciadas de acuerdo a su éxito en ofrecer una explicación de la creencia, argumentación y motivación morales, tal que sea compatible con nuestras convicciones generales sobre el mundo: nuestras creencias sobre qué tipos de cosas hay en el mundo, sobre los tipos de observación y razonamiento de que somos capaces, y sobre los tipos de razones que tenemos para la acción. Un juicio sobre qué relación de la naturaleza de la moralidad (o de las matemáticas) es la más plausible, en este sentido general, es justamente eso: un juicio de plausibilidad global. No se lo describe de modo provechoso como una captación de conceptos, o como una captación intuitiva especial de otro tipo.

Si el utilitarismo filosófico es aceptado, entonces alguna forma de utilitarismo parece imponérsenos como una doctrina normativa, pero se necesitan argumentos ulteriores para determinar qué forma deberíamos aceptar. Si todo lo que importa moralmente es el bienestar de los individuos, ninguno de los cuales es tomado más en cuenta que los demás, y si lo que importa en el caso de cada individuo es sólo el grado en que es afectado su bienestar, entonces, parecería seguirse que la base para la valoración moral

es llegar a maximizar la *suma*⁷ de bienestar individual. Es una cuestión ulterior si acaso debemos usar este rasero en la crítica de acciones individuales, en la selección de reglas o políticas, o en la inculcación de hábitos y disposiciones para actuar. Igual cosa ocurre con la cuestión sobre cómo se ha de entender el “bienestar” mismo. Luego, la hipótesis —de que mucho del atractivo del utilitarismo como doctrina normativa deriva del encanto del utilitarismo filosófico— explica cómo puede haber gente convencida de la validez de alguna forma de utilitarismo y que, al mismo tiempo, tenga muy poca certeza de qué forma sea ésta: si se trata acaso de utilitarismo “directo” o “de la acción”, o de alguna forma de “regla” indirecta o utilitarismo “del motivo”. Lo que estas visiones tienen en común, a pesar de sus diferentes consecuencias normativas, es que identifican la misma clase de hechos morales fundamentales.

II

Entonces, si lo que he dicho sobre el atractivo del utilitarismo es correcto, lo que tiene que hacer una teoría rival es proveer una alternativa al utilitarismo filosófico en tanto concepción del objeto de la moralidad. Esto es lo que busca hacer la teoría que he de llamar contractualismo. Pero aun si ella tiene éxito en eso, y se la juzga superior al utilitarismo filosófico como relación de la naturaleza de la moralidad, no habrá sido refutado el utilitarismo normativo. Persistirá la posibilidad de sentar el utilitarismo normativo sobre otras bases, por ejemplo, como el resultado normativo del mismo contractualismo. Pero habrá sido desplazado un argumento directo y, a mi juicio, influyente a favor del utilitarismo normativo.

Para dar un ejemplo de lo que quiero decir con contractualismo, una explicación contractualista de la naturaleza de lo moralmente incorrecto tendría que enunciarse de la siguiente manera:

- Un acto es incorrecto si su realización, dadas las circunstancias, fuera desaprobada en algún sistema de reglas generales de conducta que nadie podría rechazar razonablemente como base para un acuerdo general informado y no-forzado.

Esto se propone como una caracterización del tipo de propiedad que es lo moralmente incorrecto. Como el utilitarismo filosófico, tendrá conse-

⁷ Es a través de un argumento bien distinto que se llega de modo más plausible al “utilitarismo del promedio” [*average utilitarianism*], un argumento más familiar al contractualismo. Discuto un argumento tal en la sección IV, más abajo.

cuencias normativas, pero no es mi propósito actual explorarlas en detalle. Lo aquí formulado es, en tanto relación contractualista de una noción moral, sólo una aproximación que podría necesitar considerables modificaciones. Aquí puedo ofrecer, a modo de aclaración, unos pocos comentarios.

La idea de “acuerdo informado” pretende excluir el acuerdo basado en superstición o en opiniones falsas sobre las consecuencias de las acciones, incluso si estas creencias son tales que, para la persona en cuestión, sería razonable tenerlas. Por otra parte, la fuerza del calificativo “razonable” se propone excluir rechazos que no serían razonables *dada* la aspiración de encontrar principios que podrían ser la base del acuerdo general informado y no-forzado. Dada esta aspiración, no sería razonable, por ejemplo, rechazar un principio porque te impone una carga si todo principio alternativo impusiese a otros cargas mucho mayores. Más abajo tendré más que decir sobre los fundamentos para el rechazo.

El requisito de que el acuerdo hipotético, que es el objeto de argumentación moral, no sea forzado, pretende no sólo descartar la coacción, sino también excluir el verse forzado a aceptar un acuerdo por encontrarse en una débil posición de regateo, por ejemplo porque otros son capaces de aguantar más tiempo y, por ello, de insistir en mejores términos. La argumentación moral hace abstracción de tales consideraciones. La única presión relevante para el acuerdo viene del deseo de encontrar, y estar de acuerdo en, principios que nadie que tenga este deseo podría rechazar razonablemente. Según el contractualismo, la argumentación moral tiene que ver con la posibilidad de acuerdo entre personas impulsadas todas por este deseo, e impulsadas en el mismo grado por él. Pero esta suposición contrafáctica caracteriza únicamente al acuerdo que interesa a la moralidad, no al mundo al cual se aplican los principios morales. Aquellos que están preocupados con la moralidad buscan, para aplicar a su mundo imperfecto, principios que no podrían rechazar razonablemente, y que otros en este mundo, que en este momento no están impulsados por el deseo de acuerdo, no podrían rechazar razonablemente de llegar a estar así impulsados⁸.

Es por las siguientes razones que la relación contractualista de lo moralmente incorrecto refiere más a principios “que nadie podría rechazar razonablemente” que a principios “que todos podrían aceptar razonablemente”⁹. Considere un principio bajo el que alguna gente sufrirá severas penurias, y suponga que estas penurias son evitables. Esto es, hay princi-

⁸ Estoy en deuda con Gilbert Harman por comentarios que me han ayudado a clarificar mi enunciación del contractualismo.

⁹ Un punto que debo a Derek Parfit.

prios alternativos bajo los cuales nadie tendría que soportar cargas comparables. Podría ocurrir, sin embargo, que la gente sobre la que caen estas penurias es particularmente sacrificada, y está dispuesta a aceptar estas cargas por lo que ellos ven como el bien más grande de todos. No diríamos, pienso yo, que hacer esto sería algo no-razonable de su parte. Por otra parte, puede que no sea no-razonable de su parte rehusarse a estas cargas y, por consiguiente, no no-razonable para alguien rechazar un principio que le requiriera soportarlas. Si este rechazo fuera razonable, entonces el principio que impone estas cargas queda puesto en cuestión, a pesar de que cierta gente particularmente sacrificada pueda aceptarlo (razonablemente). Luego, aquello sobre lo que gira la argumentación moral es la razonabilidad de rechazar un principio, más que en la razonabilidad de aceptarlo.

Parece probable que muchos conjuntos no-equivalentes de principios van a pasar la prueba de no-rechazabilidad. Esto lo sugiere, por ejemplo, el que haya muchas maneras diferentes de definir deberes importantes, de las cuales ninguna es más o menos rechazable que las otras. Hay, por ejemplo, muchos sistemas diferentes para lograr-acuerdos, y hay muchos modos diferentes de asignar la responsabilidad de ocuparse de otros. No se sigue de ello, sin embargo, que cualquier acción permitida por (al menos) uno de estos conjuntos de principios no pueda ser moralmente incorrecta según el contractualismo. Si es importante para nosotros tener *algún* deber de un tipo determinado (algún deber de fidelidad a los acuerdos, o algún deber de ayuda mutua) del que hay muchas formas moralmente aceptables, entonces una de estas formas necesita quedar establecida por convención. En un ordenamiento en que una de estas formas *está* establecida convencionalmente, los actos que ella no permita serán incorrectos en el sentido dado en la definición. Pues, dada la necesidad de tales convenciones, no podría ser objeto de acuerdo general un conjunto de principios que permita a alguien desatender las definiciones de deberes importantes establecidas convencionalmente (y moralmente aceptables). Tal dependencia de la convención introduce un grado de relativismo cultural en la moralidad contractualista. Por añadidura, lo que una persona puede rechazar razonablemente dependerá de las aspiraciones y condiciones que sean importantes en su vida, y éstas también van a depender de la sociedad en que vive. La definición dada más arriba permite variaciones de ambos tipos, al hacer que lo incorrecto de una acción dependa de las circunstancias en las que se lleva a cabo.

El enunciado parcial del contractualismo que he dado, tiene el carácter abstracto apropiado para una explicación del objeto de la moralidad. A

primera vista, no involucra afirmaciones específicas sobre los principios en que se podría llegar a consentir, o aun si hay un único conjunto de principios que podrían ser la base de acuerdo. Una manera, aunque no la única, para que el contractualista llegue a afirmaciones morales sustantivas, sería dar una definición técnica de la noción relevante de acuerdo, por ejemplo, especificando las condiciones bajo las cuales se ha de alcanzar acuerdo, las partes de este acuerdo, y el criterio de razonabilidad que ha de ser empleado. Diferentes filósofos contractualistas han hecho esto de maneras diferentes. Lo que debe afirmarse de una definición tal es que (dadas las circunstancias en que se aplicará) lo que describe es, en realidad, el tipo de acuerdo no-forzado y razonable al que aspira la argumentación moral. Pero el contractualismo también puede entenderse como una descripción informal del objeto de la moralidad, sobre cuya base las formas ordinarias de razonamiento moral pueden ser comprendidas y apreciadas sin seguir la vía de una noción técnica de acuerdo.

¿Quién ha de estar incluido en el acuerdo general al que se refiere el contractualismo? El alcance de la moralidad es una difícil cuestión de moralidad sustantiva, pero una teoría filosófica de la naturaleza de la moralidad debería proveer alguna base para responder esa pregunta. Lo que una teoría adecuada debería hacer es proveer un marco en el que puedan desarrollarse aquellos argumentos, presuntamente relevantes, a favor o en contra de las diferentes interpretaciones de los límites de moralidad. A menudo se piensa que el contractualismo no puede proveer ninguna tesis plausible como respuesta a este asunto. Los críticos: o acusan al contractualismo de no ofrecer respuesta alguna, porque tiene que comenzar dando por sentado algún grupo de partes contratantes; o lo acusan de sugerir una respuesta muy restrictiva, por cuanto un contrato requiere que cada parte sea capaz de hacer y respetar acuerdos, y de ofrecer al resto algún beneficio a cambio de su cooperación. Ninguna de estas objeciones se aplica a la versión del contractualismo que estoy defendiendo. Ella parece implicar la siguiente especificación general sobre alcance de la moralidad: la moralidad se aplica a un ser si la noción de justificación ante tal clase de ser tiene sentido. ¿Cuáles son las condiciones para que esto sea el caso? Aquí sólo puedo sugerir algunas condiciones necesarias. La primera es que el ser en cuestión tenga un bien, es decir, que en un sentido claro se pueda decir que las cosas van mejor o peor para ese ser. Esto dota parcialmente de sentido a la idea de lo que sería razonable aceptar, a nombre de ese ser, por parte de un albacea [*trustee*]. Para un albacea sería razonable aceptar, al menos, aquellas cosas que son buenas, o no malas, para el ser en cuestión. Usando esta idea de albaceazgo [*trusteeship*], podemos extender la noción de aceptación

hasta aplicarla a seres que son incapaces de, literalmente, consentir algo. Pero esta noción mínima de albaceazgo es demasiado débil como para proveer una base para la moralidad, según el contractualismo. La moralidad contractualista descansa en nociones sobre lo que sería razonable aceptar o rechazar, nociones que son esencialmente comparativas. Si acaso rechazar cierto principio no sería razonable para mí, dada la aspiración de encontrar principios que nadie con esta aspiración rechazaría razonablemente, no depende sólo de cuánto podrían herirme en términos absolutos las acciones permitidas por tal principio, sino también de cómo esa pérdida potencial se compara con otras pérdidas potenciales para otros que se encuentren bajo este mismo principio y principios alternativos a él. Así, para que un ser entre en relaciones morales con nosotros, no basta que tenga un bien, también es necesario que su bien sea similar al nuestro, de modo que provea una base para un sistema de comparabilidad. Sólo sobre la base de un sistema tal podemos dar el sentido apropiado a la noción de lo que un albacea podría rechazar razonablemente a nombre de un ser.

Pero el alcance de un posible albaceazgo es más amplio que el de moralidad. Uno podría actuar como albacea de una plantación de tomates, de un bosque o de cualquier otra colonia, y tales entidades no están incluidas en la moralidad. Esto quizás puede ser explicado apelando al requisito de compatibilidad: si bien estas entidades tienen un bien, no es compatible con el nuestro de modo que provea una base para la argumentación moral. Pero, además, en estos casos no hay asidero suficiente para sostener la noción de justificación ante un ser. Otro requisito mínimo para esta noción es que el ser constituya un punto de vista; es decir, que haya algo así como lo que es ser como ese ser, algo así como lo que le parece a él que es el mundo. Sin este requisito, no estamos en una relación con tal ser que haga apropiada siquiera la hipotética justificación *ante él*.

Sobre la base de lo dicho hasta ahora, el contractualismo puede explicar por qué a muchos les debió haber parecido que la capacidad de sentir dolor cuenta a favor del rango moral: un ser que tiene esta capacidad parece también satisfacer las tres condiciones que recién he mencionado como necesarias para que la idea de justificar ante él tenga sentido. Si un ser puede sentir dolor, entonces constituye un centro de conciencia al que puede dirigirse la justificación. Sentir dolor es un modo claro en que la situación puede verse afectada para peor; y que su dolor tenga alivio, un modo claro en que puede verse afectado para mejor; y estos modos claros son formas de prosperidad y desgracia que parecen directamente comparables a las nuestras.

No es claro que las tres condiciones que he enumerado como necesarias sean también suficientes para que tenga sentido la idea de justificación ante un ser. Si acaso lo son o no lo son, y qué más se requeriría, son dificultades y cuestiones disputadas. Algunos restringirían la esfera moral a quienes las justificaciones podrían ser, en principio, comunicadas; o a quienes podrían consentir realmente algo; o a quienes tienen la capacidad de entender argumentaciones morales. El contractualismo, como lo he planteado yo, no resuelve inmediatamente estos problemas. Todo lo que afirmo es que provee una base para argumentar sobre ellos, base que es, al menos, tan plausible como la ofrecida por relaciones rivales acerca de la naturaleza de la moralidad. Las restricciones al alcance de la moralidad que he propuesto aquí, se han de entender, naturalmente, como afirmaciones debatibles en torno a las condiciones bajo las cuales tiene sentido la relevante noción de justificación. Y también, sobre esta base, pueden ser plausiblemente entendidos los argumentos que generalmente se ofrecen a favor o en contra de estas afirmaciones.

Algunas otras restricciones posibles al alcance de la moralidad son rechazables de modo más evidente. La moralidad podría acotarse a quien tiene la capacidad de respetar sus restricciones, o a quienes son capaces de otorgarles un beneficio recíproco a otros participantes. Pero es extremadamente poco plausible suponer que los seres excluidos por estos requerimientos están totalmente fuera de la protección de la moralidad. El contractualismo, como lo he formulado yo¹⁰, puede explicar por qué es esto así: por sí sola, la ausencia de estas capacidades no hace nada para socavar la posibilidad de justificación ante un ser. Pero lo que puede hacer, en algunos casos, es alterar las justificaciones que son relevantes. Sugiero que las capacidades de control deliberativo y beneficio recíproco tienen importancia, cualquiera sea el grado de ella, no como condiciones cuya ausencia suspende el marco moral en conjunto, sino como condiciones que alteran los deberes que los seres tienen y los deberes que otros tienen para con ellos.

¹⁰ En esta visión, en contraste con algunas otras en las que se emplea la noción de un contrato, lo que es fundamental para la moralidad es el deseo de acuerdo razonable, no el afán de ventaja mutua. Véase la sección V más abajo. Debería estar claro que esta versión del contractualismo puede responder por la situación moral de personas futuras que, como resultado de lo que hacemos ahora, se verán afectadas para mejor o peor. Es menos claro cómo puede tratar con el problema presentado por personas futuras que no habrían nacido a no ser por acciones nuestras que, a la vez, empeoraron las condiciones en que viven. ¿Tienen tales personas razón para rechazar principios que autorizan la realización de estas acciones? Este difícil problema, que no puedo explorar aquí, es planteado por Derek Parfit en "On Doing the Best for Our Children", 1976, pp. 100-115.

III

Hasta ahora he dicho poco sobre el contenido normativo del contractualismo. Por todo lo que he dicho, el utilitarismo del acto podría resultar ser un teorema del contractualismo. No pienso que éste sea el caso, pero mi tesis principal es que el contractualismo todavía tiene un contenido distintivo como tesis filosófica acerca de la naturaleza de la moralidad, cualesquiera que sean sus implicaciones normativas. Este contenido —la diferencia, por ejemplo, entre ser un utilitarista porque la fórmula utilitarista es la base de acuerdo general y serlo por otras razones— se muestra claramente en la respuesta que un contractualista da a la primera pregunta motivacional.

El utilitarismo filosófico es una visión plausible, en parte, porque los hechos que identifica como fundamentales para la moralidad —hechos sobre el bienestar individual— tienen una fuerza motivacional obvia. Los hechos morales pueden motivarnos, según esta visión, por nuestra identificación compasiva con el bien de otros. Pero si nos movemos del utilitarismo filosófico a una fórmula utilitarista específica como el criterio de la acción correcta, la forma de motivación a que apela el utilitarismo se vuelve más abstracta. Si el utilitarismo clásico es la doctrina normativa correcta, entonces la fuente natural de la motivación moral será una tendencia a ser movilizado por cambios en el bienestar agregado, sin importar la composición concreta de estos cambios. Debemos ser movilizados en la misma medida por ganancias agregadas de la misma magnitud, ya sea que se las obtenga aliviando el agudo sufrimiento de unos pocos o al entregar beneficios diminutos a un vasto número de personas, quizás a expensas de una moderada incomodidad para unos pocos. Esto es muy diferente del tipo cotidiano de compasión por personas individuales, pero un utilitarista podría argüir que este deseo más abstracto es aquello en lo que la natural compasión se convierte si se la corrige mediante la reflexión racional. Este deseo tiene el mismo contenido que la compasión —una preocupación por el bien de otros— pero ni es parcial ni selectiva en la elección de sus objetos.

Dejando de lado el problema de la plausibilidad psicológica de esta compasión empatada, ¿es una buena candidata para el rol de motivación moral? Ciertamente, el tipo usual de compasión es uno de los varios motivos que alguna vez pueden impulsar a alguien a hacer lo correcto. Puede ser el motivo dominante, por ejemplo, cuando corro en ayuda de un niño que sufre. Sin embargo, algo más está operando cuando siento que el artículo de Peter Singer¹¹ sobre la hambruna me convence, y me veo atónito al recono-

¹¹ Singer, Peter: "Famine, Affluence and Morality", 1972, pp. 229-243.

cer lo que parece una exigencia moral evidente. Pensando cuánto bien podría hacer yo a los habitantes de países azotados por la sequía, se suma una consideración ulterior, y aparentemente distinta, que me sobrecoge: sería incorrecto que yo no acudiera en su ayuda, pudiendo hacerlo a tan bajo costo para mí. Un utilitarista podría responder que su explicación de la motivación moral no puede ser defectuosa por no captar este aspecto de la experiencia moral, puesto que se trata sólo de un reflejo de nuestra crianza moral no-utilitarista. Más aun, tal aspecto de la experiencia moral tiene que carecer de fundamento. Pues ¿qué tipo de hecho podría ser este presunto hecho ulterior de incorrección moral?, y ¿cómo podría darnos una razón ulterior y especial para actuar? Luego, el problema para el contractualismo es si puede proveer una respuesta satisfactoria para este desafío.

Según el contractualismo, la fuente de motivación directamente gatillada por creer que una acción es incorrecta es el deseo de ser capaz de justificar las propias acciones frente a otros con fundamentos que ellos no podrían rechazar razonablemente¹². Encuentro que ésta es una relación de la motivación moral extremadamente plausible —al menos, una mejor relación de mi experiencia moral que la alternativa ofrecida por el utilitarismo natural— y me parece que constituye un punto fuerte para la visión contractualista. Todos querríamos encontrarnos en un acuerdo efectivo con las personas que nos rodean, pero el deseo que el contractualismo identifica como básico para la moralidad no lleva simplemente a conformarse con cualesquiera criterios aceptados por los demás. El deseo de poder justificar las propias acciones frente a otros sobre fundamentos que ellos no podrían rechazar razonablemente tendrá satisfacción cuando sepamos que hay justificación adecuada para nuestras acciones aunque otros se nieguen de hecho a aceptarla (quizás porque no tienen interés en encontrar principios que ni nosotros ni otros podrían rechazar razonablemente). De modo similar, una persona impulsada por este deseo no se dará por satisfecha con que otros acepten una justificación por sus acciones si ella considera que esta justificación es espuria.

Un modo informal de comprobar si usted considera como suficiente una justificación, es saber si acaso aceptaría esa justificación si estuviera en la posición de otra persona. Esta conexión, entre la idea de “cambiar lugares” y la motivación que subyace a la moralidad, explica la recurrencia de razonamientos estilo “regla de oro” en diferentes sistemas de moralidad y en la doctrina de varias religiones. Pero el experimento mental de cambiar

¹² Razonablemente, es decir, dado el deseo de encontrar principios que otras personas similarmente motivadas no podrían rechazar razonablemente.

lugares es sólo una guía preliminar; la cuestión fundamental es lo que sería no-razonable rechazar como una base de acuerdo general informado y no-forzado. Como observó Kant¹³, nuestros diferentes puntos de vista individuales, tomados tal como son, serían simplemente irreconciliables. La “armonía entre juicios” requiere la construcción de una forma genuinamente interpersonal de justificación que es, no obstante, algo en lo que cada individuo, unilateralmente, podría consentir. Desde este punto de vista interpersonal, contarán como prejuicio tanto una cierta medida del modo en que se ven las cosas desde el punto de vista de otra persona, como una cierta medida del modo en que se ven desde el mío propio.

No estoy afirmando que sea universal, o siquiera “natural”, el deseo de poder justificar las propias acciones frente a otros con fundamentos que no podrían rechazar razonablemente. La “educación moral” me parece entendida plausiblemente, si se la toma como un proceso de cultivar este deseo y de darle forma, en gran parte al aprender cuáles justificaciones los demás están dispuestos *de facto* a aceptar, al descubrir cuáles considera usted que son aceptables al confrontarlas desde una variedad de perspectivas, y al valorar la aceptación o rechazo, tanto suyos como ajenos, de esas justificaciones a la luz de una mayor experiencia.

De hecho, me parece que, para la mayoría, es bastante fuerte el deseo de poder justificar las propias acciones (e instituciones) con fundamentos que uno considera aceptables. La gente está dispuesta a llegar muy lejos, haciendo tremendos sacrificios, para evitar admitir que sus acciones e instituciones son injustificables. La notable insuficiencia de la motivación moral como una manera de conseguir que la gente haga lo correcto, no se debe a la simple debilidad del motivo de fondo, sino más bien a que este móvil es fácilmente distorsionado por el interés propio y el autoengaño.

Podría objetarse razonablemente aquí que la fuente de motivación que he descrito no está amarrada exclusivamente a la noción contractualista de la verdad moral. La relación de la motivación moral que ofrecí se refiere a la idea de una justificación que sería no-razonable rechazar, y esta idea es potencialmente más amplia que la noción contractualista de acuerdo. Sea M una relación no-contractualista de la verdad moral. Según M, podríamos suponer, lo incorrecto de una acción es simplemente una característica moral de esa acción, en virtud de la cual esa acción no debe ser realizada. Según M, un acto tiene esta característica independiente de la tendencia de personas informadas de llegar a acuerdo al respecto. Sin embargo, puesto

¹³ Kant, Immanuel: *The Moral Law*, traducción de H. J. Paton, 1948, sección 2, nota 14. [En castellano: *Fundamentación para la Metafísica de las Costumbres*.]

que personas informadas están presumiblemente en una posición de reconocer lo incorrecto de un tipo de acción, pareciera seguirse que si una acción es incorrecta, luego tales personas estarían de acuerdo en que no se ha de realizar. Similarmente, si un acto no es moralmente incorrecto, y hay justificaciones morales adecuadas para realizarlo, entonces habrá presumiblemente para éste una justificación moral que no podría rechazar razonablemente una persona informada. Así, incluso si M, o el no-contractualismo, es la relación correcta de la verdad moral, todavía podría servir como una base para la motivación moral el deseo de poder justificar mis acciones frente a otros sobre fundamentos que no podrían rechazar razonablemente.

Lo que esto muestra es que el atractivo del contractualismo, igual que el del utilitarismo, descansa parcialmente en un escepticismo calificado. Una teoría no-contractualista de la moralidad puede hacer uso de la fuente de motivación a que apela el contractualismo. Pero un argumento moral gatillará esta fuente de motivación sólo en virtud de ser una buena justificación para actuar de un cierto modo, una justificación que les sería no-razonable a otros no aceptar. Así, una teoría no-contractualista debe afirmar que existen propiedades morales que tienen fuerza justificativa independiente de su reconocimiento en un acuerdo ideal cualquiera. Estas propiedades representarían lo que John Mackie ha llamado instancias de intrínsecos “debe-hacerse” y “no-debe-hacerse”¹⁴. Parte del atractivo del contractualismo descansa en considerar, como ya lo plantea Mackie, lo problemática que es la existencia de tales propiedades “en el mundo”. Por contraste, el contractualismo busca explicar el carácter justificativo de las propiedades morales, así como su fuerza motivacional, en términos de la noción de acuerdo razonable. En algunos casos las propiedades morales mismas han de ser entendidas en términos de esta noción. Esto es así, por ejemplo, en el caso arriba considerado de la propiedad de lo moralmente incorrecto. Pero también hay propiedades de hacer-lo-correcto y hacer-lo-incorrecto que son, ellas mismas, independientes de la noción contractualista de acuerdo. Considero que la propiedad de ser un acto de asesinato por el puro placer de matar es un tal tipo de propiedad de hacer-lo-incorrecto. Son tales propiedades de hacer-lo-incorrecto porque sería razonable rechazar cualquier conjunto de principios que permitió los actos que ellas caracterizan. Así, mientras haya “en el mundo” propiedades moralmente relevantes que son independientes de la noción contractualista de acuerdo, éstas no constituyen instancias de “deber-hacerse” y “no-deber-hacerse” intrínsecos: su re-

¹⁴ Mackie, John: *Ethics, Inventing Right and Wrong*, 1977, p. 42. [Traducción al castellano: *Ética: la Invención de lo Bueno y lo Malo*, Gedisa, 2000.]

levancia moral —su fuerza en las justificaciones así como su vínculo con la motivación— se ha de explicar con fundamentos contractualistas.

En particular, el contractualismo puede responder por la aparente importancia moral de los hechos sobre el bienestar individual, hechos que el utilitarismo tiene por fundamentales. El bienestar individual será moralmente significativo, según el contractualismo, no porque sea intrínsecamente valioso ni porque promoverlo sea patentemente una característica de hacerlo-correcto, sino simplemente porque un individuo podría rechazar razonablemente una forma de argumentación que no tome en cuenta su bienestar. Sin embargo, plantear que el bienestar tiene importancia moral es sólo una primera aproximación, pues hay que resolver dificultades sobre cómo entender exactamente el “bienestar” y sobre el modo en que se nos pide que consideremos el bienestar de los demás al decidir qué hacer. Por ejemplo, no se sigue de esta explicación que un deseo dado tendrá siempre y en todo lugar el mismo peso al determinar el carácter correcto de una acción que promovería su satisfacción, un peso proporcional a su fuerza o “intensidad”. Lo que especifica la fuerza de hacer-lo-correcto de los deseos de una persona es, lo que podríamos llamar, una concepción de los intereses moralmente legítimos de una persona. Una concepción tal es un producto de la argumentación moral: no es algo dado simplemente por la idea de lo que es racional desear para un individuo, como sí podría serlo la noción de bienestar individual. No todo lo que deseo racionalmente será algo tal que otros tomen en cuenta al decidir cursos de acción porque se ven en la necesidad de reconocer que yo tengo en ello un interés legítimo. El campo de cosas que podrían ser objeto de mis deseos racionales es, en verdad, muy amplio, y el campo de pretensiones de legitimidad que otros no podrían negarse razonablemente a reconocer será, con toda seguridad, más restringido que aquél. Habrá una tendencia de los intereses a coincidir con el deseo racional —una tendencia de aquellas condiciones que hacen racional desear algo, a establecer también en ello un interés legítimo— pero esto no ocurrirá siempre.

Luego, un efecto del contractualismo es acabar con la distinción tajante —a la que apelan los argumentos en favor del utilitarismo— entre el rango concedido al bienestar individual y el concedido a otras nociones morales. Se requiere un marco de argumentación moral para definir nuestros intereses legítimos y para responder por su fuerza moral. Este mismo marco contractualista puede también responder por la fuerza de otras nociones morales tales como los derechos, la responsabilidad individual y la imparcialidad procedimental.

IV

Parece improbable que el utilitarismo del acto vaya a ser un teorema de la versión del contractualismo que he descrito. La importancia moral positiva de los intereses individuales es un reflejo directo del requisito contractualista de defender las acciones ante cada persona sobre fundamentos que ella no pueda rechazar razonablemente. Pero, desde aquí, hay un largo trecho hasta concluir que cada individuo tiene que aceptar deliberar siempre desde la perspectiva del máximo beneficio agregado, y aceptar justificaciones que apelen sólo a esta consideración. Es muy posible que según el contractualismo algunas cuestiones morales puedan resolverse apropiadamente apelando al máximo bienestar agregado, aunque éste no es ni el único ni el último criterio de justificación.

Lo que parece menos improbable es que el contractualismo resulte coincidiendo con alguna forma de utilitarismo “de dos pisos”. No puedo examinar a fondo esa posibilidad aquí. En verdad, el contractualismo comparte con estas teorías el importante rasgo de considerar que la defensa de acciones individuales tiene que proceder vía una defensa de principios que autorizarían esos actos. Pero el contractualismo difiere, en un sentido importante, de *algunas formas* de utilitarismo de dos pisos. El rol de los principios en el utilitarismo es fundamental; no entran meramente como mecanismos para promover actos que son correctos según algún otro criterio. Dado que no establece dos formas de razonamiento moral potencialmente conflictivas, el contractualismo evita la inestabilidad que suele atormentar al utilitarismo de la regla.

La pregunta fundamental aquí es, sin embargo, si los principios a que conduce el contractualismo tienen que ser tales que su adopción general (sea ideal o bajo condiciones algo más realistas) promovería el máximo bienestar agregado. A muchos les ha parecido que así tiene que ser. Para indicar por qué no estoy de acuerdo con ellos, voy a reflexionar sobre uno de los argumentos más conocidos en favor de esa conclusión, y explicaré por qué creo que fracasa. Con esto también tendré ocasión de examinar la relación entre la versión de contractualismo que aquí defiende y la versión desarrollada por Rawls.

El argumento a considerar, que nos es familiar por los escritos de Harsanyi¹⁵ y otros, procede vía una interpretación de la noción contrac-

¹⁵ Véase Harsanyi, John C.: “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility”, 1955, pp. 434. Aquí él discute un argumento que presentó antes en “Cardinal Utility in Welfare Economics and in Theory of Risk-Taking”, 1953.

tualista de aceptación y conduce al principio de máxima utilidad promedio. Para pensar en un principio como candidato para acuerdo unánime, debo pensarlo no sólo en cuanto aceptable para *mí* (quizá en virtud de mi situación particular, de mis gustos, etc.) sino en cuanto aceptable¹⁶ para otros a su vez. Para ser relevante, mi juicio sobre la aceptabilidad tiene que ser imparcial. ¿Qué significa esto? Podría decirse que considerar imparcialmente que un principio es aceptable es considerar que usted tendría razón para aceptarlo sin importar quién sea. Es decir, y aquí viene la interpretación, considerar si se trata de un principio que sería racional aceptar si usted no sabe la situación personal en que se encuentra, y piensa que tiene probabilidades iguales de estar en cualquiera de esas situaciones. (“Estar en la situación de una persona” es aquí estar en sus circunstancias objetivas y evaluarlas desde la perspectiva de sus gustos y preferencias.) Pero se afirma que el principio que sería racional preferir dadas estas circunstancias —el que ofrecería al que elige la mayor utilidad esperada— sería aquel principio bajo el cual la utilidad promedio de las partes afectadas sería la más alta.

Se podría cuestionar este argumento en varios puntos, pero lo que me concierne ahora es la interpretación de la imparcialidad. El argumento puede ser desmontado en tres etapas. La primera de estas etapas es la idea según la cual los principios morales tienen que ser imparcialmente aceptables. La segunda etapa es la idea de escoger principios ignorando la propia posición (incluyendo gustos, preferencias, etc.). La tercera etapa es la idea de elección racional asumiendo que uno tiene las mismas posibilidades de ocupar cualquier posición. Por el momento, permítaseme dejar de lado el tránsito de la segunda a la tercera etapa y concentrarme en el primer paso: el de la primera a la segunda etapa. Hay una manera de hacer este paso que es, pienso yo, del todo válida, pero no rinde la conclusión que pide el argumento. Si yo creo que un determinado principio P no podría ser rechazado razonablemente como base para un acuerdo general, informado y no-forzado, entonces tengo que creer no sólo que sería razonable para mí aceptarlo, sino razonable para otros también, mientras estemos todos buscando una base para acuerdo general. Consecuentemente, tengo que creer que poseería una razón para aceptar P sin importar la posición social que haya de ocupar (aunque, por razones mencionadas anteriormente, podría no creer

¹⁶ Al discutir los planteamientos de Harsanyi y Rawls, en general me atenderé a ellos al hablar de la aceptabilidad de los principios, antes que de su no-rechazabilidad. La diferencia entre estos dos modos de hablar, apuntada más arriba, es importante sólo en la versión del contractualismo que presento yo ahora: en consonancia, hablaré de rechazabilidad sólo cuando se trate de contrastar mi propia versión con la de ellos.

que *estaría* de acuerdo con P si estuviera yo en algunas de estas posiciones). Ahora podría pensarse que es imposible asignar sentido alguno a la noción de elegir o consentir un principio ignorando la propia posición social, especialmente cuando esto incluye ignorar los propios gustos, preferencias, etc. Pero existe, al menos, un sentido mínimo que podría asignarse a esta noción. Si fuera razonable para todos elegir o consentir P, entonces que yo sepa que tengo razones para ello no depende de conocer mi posición particular, mis gustos, mis preferencias, etc. Así, mientras tenga algún sentido hablar de elegir o consentir algo en ausencia de este conocimiento, se podría decir que tengo razones para elegir o consentir aquellas cosas que todos tienen razones para elegir o consentir (asumiendo, siempre, la aspiración de encontrar principios que todos puedan consentir). Y, de hecho, este mismo razonamiento nos puede llevar hasta una versión de la tercera etapa. Porque si considero que P es un principio tal que todos y cada uno tienen razón para estar de acuerdo con él, luego, se podría decir que yo tendría razón para estar de acuerdo con él si pensara que tengo la misma probabilidad de ser cualquiera o, de hecho, si asignara cualquier otro conjunto de probabilidades para ser una u otra de las personas en cuestión.

Pero está claro que con el argumento original se esperaba alcanzar una conclusión diferente. Esa conclusión versaba sobre lo que sería racional que una persona que protege sus propios intereses eligiera o consintiera. La conclusión que hemos alcanzado apela a una noción diferente: la idea de lo que no sería razonable que las personas rechacen, dado que buscan una base para acuerdo general. La dirección de la explicación es bien diferente en los dos argumentos. El argumento original buscaba explicar la noción de aceptabilidad imparcial de un principio ético, apelando a la noción —que aparece como siendo más clara— de elección racional que protege los propios intereses bajo condiciones especiales. Mi argumento revisado explica cómo sería posible asignar *un* sentido a la idea de elección o consentimiento bajo condiciones de ignorancia de la propia posición, dada cierta idea de lo que no sería razonable que alguien rechace como una base para consentimiento general. Esto insinúa un problema para mi versión de contractualismo: puede alegarse que ella es incapaz de explicar la noción central sobre la que descansa. A esto yo respondería que mi versión de contractualismo no busca explicar esta noción. Sólo trata de describirla claramente y de mostrar cómo pueden ser entendidos otros rasgos de la moralidad en términos de tal noción. En especial, no trata de explicar esta noción reduciéndola a la idea de lo que maximizaría el interés propio de una persona que

elige desde una posición de ignorancia o que asume tener las mismas probabilidades de ser cualquiera.

La plausibilidad inicial que tiene el paso de la primera etapa a la segunda etapa del argumento original descansa en una sutil transición desde una de esas nociones a la otra. Para creer que un principio es moralmente correcto hay que creer que es uno tal que todos razonablemente acordarían y nadie razonablemente lo rechazaría. Pero mi convicción en que determinado principio cumple con tales condiciones puede, a menudo, encontrarse distorsionada por una tendencia a tomar más en serio sus ventajas para mí que los posibles costos para otros. Por esta razón, la idea de “ponerme en el lugar de otro” es un útil recurso correctivo. Se puede decir lo mismo del experimento mental que es preguntar por aquello en lo que podría estar de acuerdo ignorando mi verdadera posición. Pero ambos experimentos mentales son mecanismos para considerar con mayor exactitud la pregunta por aquello que *todos y cada uno* podrían acordar o lo que ninguno podría rechazar razonablemente. En otras palabras, involucran el patrón de razonamiento exhibido en mi forma revisada del argumento en tres etapas, y no el patrón del argumento como se dio originalmente. Es una pregunta muy diferente aquella por lo que maximizaría las expectativas de una única persona que, buscando proteger sus propios intereses, elige ignorando su verdadera posición. Esto puede verse al considerar la posibilidad de que la distribución con la más alta utilidad promedio, llámesela A, pueda implicar niveles extremadamente bajos de utilidad para algunas personas, niveles mucho más bajos que el mínimo que cualquiera disfrutaría bajo una distribución más equitativa.

Suponga que A es un principio que sería racional elegir para alguien que protege su propio interés, con la misma probabilidad de estar en la posición de cualquiera. ¿Se sigue que nadie podría rechazar A razonablemente? Parece evidente que la respuesta es no¹⁷. Suponga que la situación de aquellos cuya situación empeoraría bajo el principio A, llámeselos Perdedores, es extremadamente mala, y que hay alternativa a A, llámesele E, bajo la cual nadie podría estar en una situación ni por asomo tan mala como la de los Perdedores. *Prima facie*, los Perdedores parecen tener un fundamento razonable para quejarse de A. Su objeción puede ser rebatida, apelando a los sacrificios que se les impondrían a algunos otros individuos al elegirse E antes que A. Pero que A proporcione la más alta utilidad promedio no

¹⁷ La discusión que sigue a continuación tiene mucho en común con el contraste trazado por Thomas Nagel (*Mortal Questions*, 1979, capítulo 8, “Equality”) entre principios de mayoría y principios de unanimidad. Estoy en deuda con la discusión que hace Nagel al respecto.

resuelve la cuestión, puesto que podría deberse a que mucha gente está ligeramente mejor bajo A que bajo E.

Cuando consideramos un principio desde el contractualismo, nuestra atención se dirige, natural y primariamente, hacia quienes les iría peor bajo tal principio. Esto se debe a que si alguien tiene bases razonables para objetar el principio, es *probable* que sean ellos. Sin embargo, de esto no se sigue que el contractualismo nos exige siempre seleccionar aquel principio bajo el cual las expectativas del que está peor sean las más altas. Lo razonable de la objeción de los Perdedores no se establece simplemente porque les va peor bajo A y a nadie le iría así de mal bajo E. La fuerza de su queja depende también de que su posición bajo A es, en términos absolutos, muy mala, y sería significativamente mejor bajo E. Esta queja tiene que ser sopesada con la queja de los individuos a quienes les iría peor bajo E. Hay que preguntar entonces, ¿es no-razonable que alguien rechace tolerar la situación de los Perdedores bajo A para que algún otro tenga la posibilidad de disfrutar los beneficios a los que tendría que renunciar bajo E? Su caso se debilita en la medida en que mejora la situación en que se supone a los perdedores bajo A, o en la medida en que su ganancia bajo E disminuye en relación con los sacrificios que requiere producirla.

Vale la pena destacar un rasgo de la argumentación contractualista, tal como la he presentado hasta ahora: es no-agregativa. Lo que se compara son las ganancias, pérdidas y niveles de bienestar de los individuos. Cuán en cuenta pueden ser tomadas las consideraciones agregativas en la argumentación contractualista, ésta es una cuestión ulterior demasiado extensa como para tratarla ahora.

He estado criticando un argumento a favor del Utilitarismo del Promedio que se asocia, generalmente, con el nombre de Harsanyi. Mi objeción a este argumento (dejando de lado las últimas acotaciones sobre el *maximin*) tiene una semejanza obvia con las objeciones elevadas por Rawls¹⁸. Pero las objeciones que he elevado yo se aplican igualmente contra algunos rasgos de la propia argumentación de Rawls. Rawls acepta el primer paso del argumento que he descrito. Es decir, opina que los principios de justicia correctos son aquellos que aceptarían personas racionales preocupadas en promover sus intereses bajo las condiciones que define su Posición Original, en la que ignoran los talentos particulares que poseen, su concepción del bien y la posición social (o generación) en que han nacido. Rawls recha-

¹⁸ Por ejemplo, el argumento intuitivo contra el utilitarismo en la página 14 de *A Theory of Justice* y su repetida advertencia de la imposibilidad de esperar que algunas personas acepten situaciones inferiores de vida para promover las expectativas superiores de otros.

za el argumento en el segundo paso, es decir, al pretender que lo racional es que las personas en tal situación elijan aquellos principios que ofrecen la máxima utilidad esperada, asumiendo que tienen la misma probabilidad de estar en la situación de cualquiera dentro de tal sociedad. No obstante, opino que ya se ha cometido un error al dar el primer paso.

Esto puede llevarse a la luz si consideramos una ambigüedad presente en la idea de una aceptación por parte de personas “preocupadas por promover sus intereses”. Según una lectura, éste es un ingrediente esencial para una argumentación contractualista; para otra lectura, es algo evitable y, así lo pienso yo, un error. Para la primera de las lecturas, los intereses en cuestión son meramente aquellos de los miembros de la sociedad a quienes se han de aplicar los principios de justicia (y quienes, en último término, tienen que aceptar los principios). Lo que da sustancia a las cuestiones de justicia es que las partes tienen intereses que buscan promover y que pueden entrar en conflicto. Para la segunda de las lecturas, la preocupación “por promover sus intereses” que está en cuestión es una preocupación de las partes en la Posición Original de Rawls, y es esta preocupación la que determina, en primera instancia¹⁹, qué principios de justicia van adoptar. Se ha de alcanzar acuerdo unánime entre estas partes, cada una motivada para favorecer su propia situación tanto como sea capaz, al privarles de cualquier información que pudiera darles razones para hacer una elección distinta a la de cualquier otro. Bajo el velo de la ignorancia, lo que ofrece las mejores perspectivas para uno ofrecerá las mejores perspectivas para todos, dado que nadie puede decir lo que le beneficiaría de modo particular. Luego, Rawls dice que la elección de los principios puede llevarse a cabo desde el punto de vista de un único individuo racional bajo el velo de la ignorancia.

Cualquiera sean las reglas de elección racional que digamos que usa este individuo único preocupado por promover sus propios intereses lo mejor que pueda, debería despertar nuestras sospechas esta reducción del problema al caso de la elección de una única persona que protege sus intereses. Como ya indiqué al criticar a Harsanyi, es importante preguntar si se supone que este individuo único acepta un principio porque considera que se trata de uno que no podría rechazar razonablemente cualquiera sea la posición que termine ocupando, o si, por el contrario, se supone que es aceptable para una persona en cualquier posición social porque sería la elección racional para una persona individual que protege sus intereses

¹⁹ Aunque luego tienen que asegurarse de ver que los principios que han escogido serán estables, no generan tensiones de compromiso intolerables, etc. Como sostengo más abajo, estas consideraciones ulteriores pueden interpretarse en un sentido que acerca la teoría de Rawls a la versión del contractualismo aquí presentada.

bajo un velo de ignorancia. Más arriba he alegado que el argumento a favor del utilitarismo del promedio implica una transición encubierta desde el primer patrón de razonamiento al segundo. También el argumento de Rawls parece tener esta segunda forma; hace descansar la defensa de sus dos principios de justicia, al menos al comienzo, en proposiciones sobre lo que sería racional que una persona preocupada por promover sus propios intereses eligiera bajo un velo de ignorancia. Sin embargo, yo diría que la plausibilidad de los argumentos de Rawls en favor de sus dos principios (por encima del principio de la utilidad promedio) queda intacta, y en ciertos casos se ve realizada, cuando estos principios son interpretados como instancias de la primera forma de argumentación contractualista.

Algunos de estos argumentos tienen un carácter moral informal. Ya mencioné su observación sobre lo inaceptable que es imponer expectativas inferiores a algunos para promover las expectativas superiores de otros. Específicamente, Rawls dice que las partes de la Posición Original que están preocupadas “de elegir principios cuyas consecuencias en sus vidas están dispuestos a asumir sin importar la generación a la que finalmente pertenecen”²⁰ o, presumiblemente, sin importar la posición social que resulte que tienen. Éste es un claro enunciado enmarcado en la primera forma de argumentación contractualista. Algo más tarde y en favor de los dos principios, Rawls observa que son “aquellos que elegiría una persona para el diseño de una sociedad en la que su enemigo le asigna a esa persona su lugar”²¹. Rawls descartaría después esta observación, diciendo que las partes “no deberían razonar partiendo de premisas falsas”²², pero vale la pena preguntarse por qué pudo haber parecido que plantearla era algo plausible. Asumo que la razón es la siguiente. En la primera forma de argumentación contractualista, que tiene por objeto encontrar principios que sean aceptables para cada persona, la asignación de mi lugar en la sociedad por un oponente malintencionado es un experimento mental con un rol heurístico semejante al del velo de ignorancia: es una vía para comprobar si uno realmente considera que un principio es aceptable desde todos los puntos de vista o si, por el contrario, uno no está tomando realmente en serio su efecto sobre personas que ocupan posiciones sociales diferentes a la propia.

Pero todas estas son observaciones informales, y es justo suponer que el argumento de Rawls (igual que el argumento en pro de la utilidad promedio) pretende desplazarse desde la idea, informal y contractualista, de

²⁰ Rawls: *A Theory of Justice*, 1971, p. 137.

²¹ Rawls: *A Theory of Justice*, 1971, p. 152.

²² Rawls: *A Theory of Justice*, 1971, p. 153.

principios “aceptables para todos”, a la idea de elección racional bajo un velo de ignorancia. Rawls espera que esta última sea una idea más precisa y más capaz de proveer resultados definitivos. Permítaseme entonces volver sobre su argumento formal para que las partes en la Posición Original elijan el Principio de la Diferencia. Rawls cita tres rasgos de la decisión que encaran las partes en la Posición Original, rasgos que según él hacen racional que las partes usen la regla *maximin* y que, por ello, elijan el Principio de la Diferencia como uno de los principios de justicia. Estos rasgos son: (1) la ausencia de cualquier base objetiva para estimar las probabilidades, (2) que algunos principios podrían tener consecuencias que “ellos difícilmente aceptarían”, mientras que (3) es posible que ellos (siguiendo el *maximin*) se aseguren un prospecto mínimo, tal que las mejoras por sobre este mínimo importan muy poco²³. El primero de estos rasgos es ligeramente enigmático, y lo dejo de lado. Sin embargo, parece claro que las otras consideraciones mencionadas tienen, al menos, tanta fuerza en una argumentación contractualista (sobre lo que todos podrían acordar razonablemente) como la tienen al determinar la elección racional de una única persona preocupada por promover sus intereses. Expresan la fuerza que tienen las objeciones de los “perdedores” respecto a un esquema que maximice la utilidad promedio a sus expensas, en comparación con las réplicas que otros podrían hacer ante un arreglo más igualitario.

Adicionalmente a este argumento sobre la elección racional, Rawls invoca otras consideraciones que figuran entre “las razones más importantes a favor de los dos principios”, consideraciones que, según él, hacen un uso extendido del concepto de contrato²⁴. Las partes en la Posición Original, dice Rawls, pueden acordar principios de justicia sólo si piensan que este acuerdo es tal que serán realmente capaces de vivir bajo él. Plantea que pensar esto es más plausible respecto de los dos principios que del principio de la utilidad promedio, bajo el cual los sacrificios demandados (“las tensiones del compromiso”) serían muchos más elevados. Un segundo planteamiento en este contexto es que los dos principios de justicia tienen una estabilidad psicológica mayor que el principio de la utilidad promedio. Rawls dice que es más plausible creer que en una sociedad formada por personas satisfechas, ellas seguirán aceptando los dos principios y continuarán estando motivadas a actuar de acuerdo con ellos. Por otro lado, la aceptación continua del principio de la utilidad promedio requeriría un grado excepcional de identificación con el bien de la totalidad social por parte de quienes se demandaran los sacrificios.

²³ Rawls: *A Theory of Justice*, 1971, p. 154.

²⁴ Rawls: *A Theory of Justice*, 1971, sec. 29, pp. 175ss.

Estas observaciones pueden comprenderse como planteamientos sobre la “estabilidad” (en un sentido sumamente práctico) de una sociedad fundada sobre los dos principios de justicia postulados por Rawls. Pero también pueden ser vistos como un intento de mostrar que un principio al que se ha llegado por el camino de la segunda forma de razonamiento contractualista, también cumplirá con los requerimientos de la primera forma de razonamiento, es decir, que se trata de un principio que nadie podría rechazar razonablemente. La pregunta que surge “¿es la aceptación de este principio un acuerdo que realmente podrías soportar por el resto de tu vida?”, al igual que la idea de tu peor enemigo asignándote un lugar en la sociedad, es un experimento mental que permite usar nuestras propias reacciones para comprobar nuestras propias afirmaciones de que nadie podría rechazar razonablemente ciertos principios. También pueden ser invocados con este fin los principios generales de la psicología humana.

El argumento final de Rawls es que la adopción de sus dos principios proporciona un soporte público al autorrespeto de cada miembro de la sociedad, y “provee una interpretación más fuerte y más característica de la idea kantiana”²⁵ del deber de tratar a las personas como fines, y no meramente como medios para el mayor bien colectivo. Pero cualquiera que sea la diferencia que aquí pueda existir entre los dos principios de Rawls y el principio de la utilidad promedio, existe un contraste igual de marcado entre los dos patrones de razonamiento contractualista que distinguimos más arriba. La conexión con el autorrespeto, y con la fórmula kantiana, resiste cuando requerimos que los principios de justicia sean tales que ningún miembro de la sociedad pueda rechazarlos razonablemente. Esta conexión se debilita cuando nos volvemos hacia la idea de una elección que promueva los intereses de un único individuo racional para quien las variadas vidas individuales en una sociedad son sólo posibilidades diferentes e innumerables. Y esto es así sin importar qué reglas de decisión decimos que usa quien escoge racionalmente. El argumento basado en el *maximin* parece preservar esta conexión porque reproduce bajo la forma de un planteamiento sobre la elección racional lo que es (en términos ligeramente diferentes) un atractivo argumento moral.

La “situación de la elección” que es fundamental para el contractualismo (como yo lo describo) resulta al comenzar con individuos “mutuamente desinteresados” que gozan de un completo conocimiento de su condición y al suponer, adicionalmente, que cada una de las partes tiene un deseo (y no, como se insinúa a veces, benevolencia) de encontrar princi-

²⁵ Rawls: *A Theory of Justice*, 1971, p. 183.

pios que ninguno podría rechazar razonablemente tan pronto como participa también de este deseo. En muchas ocasiones, Rawls considera de pasada una idea semejante²⁶. Es rechazada en favor de su propia idea de elección mutuamente desinteresada bajo un velo de ignorancia basándose en que sólo este último nos permite alcanzar resultados definitivos: “si requerimos unanimidad al escoger principios incluso disponiendo de toda la información, sólo podremos decidir unos pocos casos más bien obvios”²⁷. Pienso que esta presunta ventaja es cuestionable. Quizá se deba a que mis expectativas son más modestas que las de Rawls en lo que respecta a la argumentación moral. Sin embargo, y como ya he alegado, casi todos los argumentos del propio Rawls tienen, al menos, la misma fuerza si son interpretados como argumentos en el marco del contractualismo que yo he venido proponiendo. Una posible excepción es el argumento basado en el *maximin*. Si consideráramos que el Principio de la Diferencia se aplica de modo generalizado a decisiones en política pública, entonces las implicaciones de la segunda forma de razonamiento contractualista (de la que derivamos tal principio) tendrían mucho más alcance que la forma más laxa de argumentación que he empleado yo y que procede comparando pérdidas. Pero estas implicaciones más vastas del principio no son siempre plausibles, y no creo que Rawls pretenda que sea aplicado tan ampliamente. Su intención es que el Principio de la Diferencia debería ser aplicado sólo a las inequidades más importantes que generan las instituciones básicas de una sociedad, y esta restricción es un reflejo de las particulares condiciones bajo las cuales Rawls dice que el *maximin* es la base apropiada para la elección racional: algunas elecciones tienen resultados que uno difícilmente acepta, mientras que importan muy poco las ganancias por sobre ese mínimo que uno puede asegurarse, etc. Luego, de aquí se sigue que aplicando el Principio de la Diferencia —identificando los límites de su aplicabilidad— debemos recurrir a la comparación informal de pérdidas que es central en la forma de contractualismo que yo he descrito.

V

He descrito esta versión del contractualismo sólo en sus líneas generales. Es necesario decir mucho más para clarificar sus nociones y elaborar sus implicaciones normativas. Espero haber dicho lo suficiente para indi-

²⁶ Por ejemplo, Rawls: *A Theory of Justice*, 1971, pp. 141, 148, aunque estos pasajes pueden no distinguir claramente entre esta alternativa y una que supone benevolencia.

²⁷ Rawls: *A Theory of Justice*, 1971, p. 141.

car su atractivo como teoría filosófica de moralidad y como una relación de la motivación moral. He presentado el contractualismo como una alternativa al utilitarismo, pero el rasgo característico de la doctrina puede ser llevado a la luz contrastándolo con una visión algo diferente.

A veces escucho decir²⁸ que la moralidad es un mecanismo para nuestra protección mutua. De acuerdo al contractualismo, esta visión es parcialmente verdadera, pero incompleta en un sentido fundamental. Nuestra preocupación por proteger nuestros intereses centrales tendrá un efecto importante en aquello que podríamos consentir razonablemente. Luego, esto tendrá un efecto importante en el contenido de la moralidad si el contractualismo es correcto. Dependiendo del grado en que observemos esta moralidad, estos intereses se verán beneficiados por ella. Si no tenemos ningún deseo de poder justificar nuestras acciones ante otros con fundamentos que podrían aceptar razonablemente, la esperanza de obtener esta protección nos daría una razón para tratar de inculcar este deseo en otros, quizá hipnotizando o condicionando a las masas, aunque esto implique que también lo adquirimos nosotros mismos. Pero dado que nosotros ya tenemos este deseo, nuestra preocupación por la moralidad es menos instrumental.

El contraste puede plantearse como sigue. Según un enfoque, la preocupación por la protección es algo fundamental y el acuerdo general deviene relevante en tanto es un medio o una condición necesaria para asegurar esta protección. Según el otro enfoque, contractualista, el deseo de protección es un factor importante que determina el contenido de la moralidad porque determina aquello que puede acordarse razonablemente. Pero la idea de acuerdo general no surge como un medio para asegurarse protección. Es, en un sentido más fundamental, el objeto de la moralidad.

REFERENCIAS BIBLIOGRÁFICAS

- Brandt, R.: *A Theory of the Good and The Right*. Oxford University Press, 1979.
- Daniels, N.: "Wide Reflective Equilibrium and Theory of Acceptance in Ethics". En *Journal of Philosophy*, 76.
- Harsanyi, John C.: "Cardinal Utility in Welfare Economics and in Theory of Risk-Taking". En *Journal of Political Economy*, 61 (1953).
- Harsanyi, John C.: "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility". En *Journal of Political Economy*, 63 (1955).

²⁸ De maneras diferentes, Warnock, G.: *The Object of Morality*, 1971, y Mackie en *Ethics, Inventing Right and Wrong*, 1977. Véanse también las observaciones de Richard Brandt sobre justificación en el capítulo X de *A Theory of the Good and The Right*, 1979.

- Kant, Immanuel: *The Moral Law*. Traducción al inglés de H. J. Paton. Hutchinson, 1948.
- Mackie, John: *Ethics, Inventing Right and Wrong*. Harmondsworth: Penguin Books, 1977. [Traducción al castellano: *Ética: la Invención de lo Bueno y lo Malo*, Gedisa, 2000.]
- Nagel, Thomas: *Mortal Questions*. Cambridge University Press, 1979.
- Parfit, Derek: "On Doing the Best for Our Children". En M. Bayles (e.d.), *Ethics and Population*. Schenkman Publishing Company Inc, 1976.
- Rawls, J.: *A Theory of Justice*. Harvard University Press, 1971. [Traducción al castellano: *Una Teoría de la Justicia*, FCE, 1995]
- Rawls, J.: "The Independence of Moral Theory". En *Proceedings and Addresses of the American Philosophical Association*, 47, (1974-1975).
- Ross, W. D.: *The Foundations of Ethics*. Oxford University Press, 1939. [Traducción al castellano: *Fundamentos de Ética*, Eudeba, 1972].
- Singer, P.: "Sidgwick and Reflexive Equilibrium". En *The Monist*, 58 (1974).
- Singer, Peter: "Famine, Affluence and Morality". En *Philosophy and Public Affairs*, 1 (1972).
- Warnock, G.: *The Object of Morality*. Methuen, 1971. □