

Rawls y el utilitarismo*

Samuel Scheffler **
(Universidad de California)

En el Prefacio a *Una Teoría de la Justicia*^{1***}, Rawls observa que “durante mucho tiempo la teoría sistemática predominante en la filosofía moral moderna ha sido alguna forma de utilitarismo” (*TJ*, p.vii/xvii rev.). Los críticos del utilitarismo, nos dice, han señalado que muchas de sus implicaciones se mueven de manera contraria a nuestros sentimientos y convicciones morales, pero fracasaron “al tratar de construir una concepción moral practicable y sistemática que se le oponga” (*TJ*, p.viii/xviii rev.). Como resultado de ello, Rawls escribe que “con frecuencia nos vemos obligados a escoger entre el utilitarismo y el intuicionismo”. Al final, especula, nos inclinamos por “adoptar una variante del principio de la utilidad circunscrito y limitado, en ciertas maneras *ad hoc*, por restricciones intuicionistas”. “Tal punto de vista”, agrega, “no es irracional, y no existe garantía de que podamos hacer algo mejor. Pero eso no es razón para no intentarlo” (*TJ*, p.viii/xviii rev.). De manera similar, lo que propone hacer “es generalizar y llevar la teoría tradicional del contrato social representada por Locke, Rousseau y Kant a un nivel más elevado de abstracción”. Rawls cree que de todas las teorías tradicionales de la justicia, la teoría del contrato es “la que mejor se aproxima a nuestros juicios ponderados acerca de la justicia”. Su propósito es desarrollar esta

* Texto publicado en Samuel Freeman (editor). *The Cambridge Companion to Rawls*. Cambridge University Press, 2002, pp.426-459. Se publica con autorización de la editorial y del autor. Traducción de Víctor Alarcón Olgún. (UAM-Iztapalapa / México).

** Profesor de Filosofía y Derecho en el Departamento de Filosofía de la Universidad de California en Berkeley, EUA.

¹ Harvard University Press, 1971; edición revisada, 1999. Citada aquí como *TJ*.

*** *Nota del traductor*. En lo posible, se ha reproducido la traducción de los pasajes de la *Teoría de la Justicia* hecha por María Dolores González en 1979 para el FCE, pero respetando la paginación de la edición inglesa original. Se tomó dicha decisión porque la paginación en castellano varía para reimpresiones posteriores a 1979, debido a que el propio FCE decidió una compactación de la tipografía. Por otra parte, cabe señalar que Rawls realizó una revisión general a su obra en 1999, y dichos cambios están indicados por Scheffler en la versión original de su trabajo. Dado que la edición del FCE no incorpora aún las correcciones de la edición revisada, es por ello que se decidió conservar la paginación inglesa. Las variaciones de traducción que pueda encontrar el lector en los pasajes traducidos son responsabilidad mía. VAO.

teoría de manera tal que “ofrezca una explicación sistemática alternativa que sea superior... al utilitarismo dominante de la tradición” (*TJ*, viii/xviii rev.).

Rawls toca una tecla similar hacia el final del capítulo 1, donde observa que “las diversas variantes de la visión utilitarista han dominado nuestra tradición filosófica por largo tiempo y continúan haciéndolo”, esto a pesar de “los recelos persistentes que tan fácilmente despierta el utilitarismo” (*TJ*, p.52 / 46 rev.). Este es, nos dice “un peculiar estado de cosas”, que debe ser explicado por “el hecho de que no se ha propuesto ninguna teoría constructiva alternativa que tenga virtudes comparables de claridad y sistematización, y que al mismo tiempo disipe tales dudas” (*TJ*, p.52 / 46 rev.). La conjetura de Rawls “es que la doctrina del contrato correctamente elaborada puede llenar dicho vacío” (*TJ*, p.52 / 46 rev.)

La afirmación de Rawls de haber delineado una teoría [“justicia como imparcialidad”] que es superior a la del utilitarismo ha generado extensos debates. A pesar de su oposición al utilitarismo, parece evidente de los pasajes que he citado que él también la considera como poseedora de virtudes teóricas que desea emular. En particular, admira el carácter “sistemático” y “constructivo” del utilitarismo, y piensa que es desafortunado que las perspectivas señaladas por los críticos del utilitarismo no hayan sido comparativamente sistemáticas o constructivas. Este aspecto de la actitud de Rawls hacia el utilitarismo ha recibido menos atención. Sin embargo, ello marca una diferencia importante entre su visión y las de otros prominentes críticos del utilitarismo que han escrito en la misma época, incluso cuando dichos críticos expresan sus objeciones en un lenguaje que es una reminiscencia del suyo. Por ejemplo, donde Rawls dice que “el utilitarismo no considera seriamente la distinción entre personas” (*TJ*, p.27 / 24 rev.), Robert Nozick, citando explícitamente a Rawls, dice que sacrificar a un individuo a cambio del mayor bien social “no respeta en forma suficiente ni toma en consideración el hecho de que es una persona distinta, y de que la suya es la única vida que tiene”². Y Bernard Williams, al desarrollar una crítica diferente, aunque no del todo desligada, argumenta que el utilitarismo vuelve la “integridad personal como un valor más o menos ininteligible”³. Pero ni Nozick ni Williams señalan la importancia de proporcionar una alternativa *sistemática* al utilitarismo.

He mencionado que el aprecio de Rawls por el carácter sistemático y constructivo del utilitarismo ha concitado menores comentarios que los de su afirmación de haber identificado una teoría de la justicia que es preferible al

² *Anarchy, State and Utopia*. (New York, Basic Books, 1974), p. 33 (Existe traducción al español,. México, FCE, 1988).

³ “A Critique of Utilitarianism”, en *Utilitarianism For and Against*, eds. J.J.C Stuart y Bernard Williams (Cambridge, Cambridge University Press, 1973), pp. 77-150. Ver p. 99.

utilitarismo. Sin embargo, una serie de críticos han argumentado que la posición de Rawls tiene puntos en común importantes con el utilitarismo –rasgos en virtud de los cuales su propia visión está abierta a algunas de las mismas objeciones que él lanza contra los utilitaristas–. Acertados o no, tales argumentos pueden considerarse en parte como respuestas al énfasis puesto en el sistema, lo cual es propio tanto de la teoría de Rawls como de la del utilitarismo. En la medida en que ello es así, aquéllos pueden ayudar a explicar la actitud compleja de Rawls hacia el utilitarismo –una actitud marcada por el respeto y por áreas de afinidad, tanto como por profundos desacuerdos–. También pueden ayudarnos a ver que cierta gente pueda sentirse perturbada por los argumentos de Rawls en contra del utilitarismo, no porque simpaticen con aquellos aspectos de la visión que éste critica, sino más bien porque son críticos de aquellos aspectos de la postura con la que él simpatiza.

En este ensayo comenzaré por revisar los argumentos principales de Rawls en contra del utilitarismo. Luego examinaré un argumento de Nozick y otro de Michael Sandel para demostrar que existe cierta tensión entre determinados aspectos de la teoría de Rawls y sus críticas al utilitarismo. Explicaré por qué no tomo este argumento como persuasivo, pero también indicaré cómo esto apunta hacia algunas afinidades más genuinas entre la justicia como imparcialidad y las ideas utilitaristas –afinidades que luego exploraré con mayor profundidad–. Mi expectativa es alcanzar una valoración equilibrada de la actitud de Rawls respecto del utilitarismo. Concluiré discutiendo algunas de las aparentes diferencias entre la posición de Rawls en *Una Teoría de la Justicia* y su perspectiva en *Liberalismo Político*⁴.

Las críticas de Rawls al utilitarismo abarcan una variedad de afirmaciones que dependen del aparato de la posición original en niveles de grado y maneras. Formalmente, su objetivo es mostrar que las partes en la posición original preferirían su propia concepción de justicia –justicia como imparcialidad– a una concepción utilitarista. Por lo tanto, sus argumentos “oficiales” en contra del utilitarismo toman la forma de argumentos orientados a mostrar que esto sería rechazado por las partes. Aunque lo más importante de dichos argumentos también puede ser formulado independientemente de la construcción de la posición original, sin contar con que existen algunos argumentos que en absoluto son ofrecidos desde el punto ventajoso de la posición original.

Rawls ofrece distintos argumentos en contra de dos formas de utilitarismo: la versión clásica y el principio de utilidad promedio. El utilitarismo clási-

⁴ Columbia University Press, 1993 (edición rústica, 1996). [Citada aquí como *LP. Nota del traductor*. Los pasajes aquí traducidos son responsabilidad mía, VAO. Sin embargo, el lector puede tener acceso a las dos traducciones de *LP* hechas por Antoni Doménech para la editorial Crítica-Grijalbo, Barcelona, 1996, y por Sergio René Madero para el FCE, México, 1996].

co, como él lo entiende, sostiene que “la sociedad está correctamente ordenada sólo cuando sus principales instituciones están dispuestas para lograr el mayor balance neto de satisfacción de todos los individuos que pertenecen a la misma” (*TJ*, p.22 / 20 rev.). Es ésa, según Rawls, una teoría teleológica, con lo cual quiere decir que aquélla define lo bueno independientemente de lo correcto, y define lo correcto como maximizador de lo bueno. El principio de utilidad promedio, como su nombre sugiere, “dirige a la sociedad a maximizar no la utilidad total, sino la utilidad media” (*TJ*, p.162/140 rev.). Aunque el utilitarismo clásico y promedio “a menudo puedan tener consecuencias prácticas similares” (*TJ*, p.189 / 165 rev.), y aunque tales consecuencias coincidan completamente en tanto el tamaño de la población permanezca constante, Rawls argumenta que las dos visiones son “concepciones marcadamente distintas”, cuyos “supuestos analíticos subyacentes están muy distantes” (*TJ*, p.161 / 139 rev.).

Rawls observa que la distribución de la satisfacción dentro de la sociedad no tiene un significado intrínseco para el utilitarismo clásico. Si una distribución radicalmente inequitativa –sea en la satisfacción misma o en los medios de satisfacción– resultara en la mayor satisfacción global, la desigualdad de la distribución no es razón para evitarla. Por supuesto, como lo reconoce Rawls, los utilitaristas con frecuencia argumentan que, dados supuestos empíricos plausibles, es improbable que la maximización de la satisfacción se logre en tal modo. Sin embargo, permanece el hecho de que el utilitarismo clásico no atribuye importancia intrínseca a cuestiones de distribución, y no impone límites de principio al terreno donde el razonamiento agregado pueda ser legítimamente empleado en la elaboración de decisiones sociales. Las pérdidas de cierta gente, pueden, en principio, estar siempre sobrevaluadas por las mayores ganancias de otra persona.

Rawls argumenta que este compromiso con la agregación irrestricta puede ser visto como el resultado de extender a “la sociedad como un todo el principio de la elección racional para un hombre” (*TJ*, p.26-7 / 24 rev.). En otras palabras, normalmente pensamos que es razonable para un simple individuo buscar la maximización de la satisfacción en el curso de una vida. Es razonable, por ejemplo, “imponernos a nosotros mismos un sacrificio ahora en aras de una mayor ventaja posterior” (*TJ*, p.23 / 21 rev.). El utilitarista clásico, argumenta Rawls, razona en gran medida de la misma forma acerca de la sociedad como un todo, asumiéndola como legítima para imponer sacrificios a cierta gente en función de obtener mayores ventajas para otros. Esta extensión a la sociedad como un todo del principio de opción para un individuo se facilita, cree Rawls, al tratar la aprobación de un espectador perfectamente solidario, idealmente racional e imparcial, como el modelo de lo que es justo. Dado que el espectador imparcial “se identifica con y experimenta los deseos

de otros como si éstos fueran los suyos”, su función es organizar “los deseos de todas las personas dentro de un sistema coherente de deseos” (*TJ*, p.27 / 24 rev.). De esta manera, “muchas personas son fusionadas en una” (*TJ*, p.27 / 24 rev.). En síntesis, Rawls argumenta que “la visión clásica del utilitarismo de la cooperación social es la consecuencia de extender a la sociedad el principio de la opción para un hombre, y luego, para lograr que dicha extensión trabaje, conjugar a todas las personas en una mediante los actos imaginativos del espectador solidario e imparcial” (*TJ*, p.27 / 24 rev.). Esto es lo que orienta a Rawls hacia la afirmación de que esta forma de utilitarismo “no toma en serio la distinción entre personas”. Esta es una objeción decisiva en virtud “de que asumimos que el principio regulativo correcto para cualquier cosa depende de la naturaleza de dicha cosa, y que la mayoría de las diversas personas con sistemas separados de fines es una característica esencial de las sociedades humanas” (*TJ*. p.29 / 25 rev.).

Vale la pena hacer notar que este argumento en contra del utilitarismo clásico es desarrollado sin referencia al aparato de la posición original y no es dependiente del mismo. Por supuesto, Rawls nos ofrece un argumento adicional en el caso de que las partes en la posición original rechazaran la perspectiva clásica. El argumento es que las partes, conociendo que ellas existen y desean sólo mejorar en sus propios intereses, no desearían maximizar la satisfacción neta agregada, ante todo porque hacer eso podría requerir un aumento en el tamaño de la población incluso a expensas de una reducción significativa en la utilidad promedio por persona. Dicho argumento es directo y parece decisivo. En cualquier caso, ha concitado mucho menos controversia que el reclamo de Rawls de que las partes pudieran rechazar el principio de utilidad promedio. Aunque probablemente es justo decir que ha sido menos influyente, en tanto argumento contra el utilitarismo clásico, que el argumento ofrecido por separado de la construcción de la posición original.

Rawls además sugiere que si los términos de la posición original fueran alterados de manera tal que las partes estuvieran “concebidas como altruistas perfectos, es decir, como personas cuyos deseos se conforman a las aprobaciones” (*TJ*, p.188-9 / 164 rev.) de un espectador imparcial y solidario, entonces el utilitarismo clásico sería adoptado en efecto. Esto le lleva hacia la “conclusión inesperada” de que la visión clásica es la “ética de altruistas perfectos”, en contraste con el principio de utilidad promedio, el cual desde la perspectiva proporcionada por la posición original, emerge como “la ética de un simple individuo racional [sin aversión al riesgo]” (*TJ*, p.189 / 164-5 rev.). Por tanto, al observar las dos versiones del utilitarismo desde el punto de vista de la posición original, se revela un “contraste sorprendente” entre ellas. (*TJ*, p.189 / 165 rev.). Esto le sugiere a Rawls que “incluso si el concepto de la posición original no sirviera para otro propósito, sería un recurso analítico

útil” (*TJ*, p.189 / 165 rev.) que nos permitiría ver “los diferentes conjuntos de ideas” (*TJ*, p.189 / 165 rev.) en que descansan las dos versiones del utilitarismo. Sin embargo, la caracterización del liberalismo clásico como la ética de altruistas perfectos parece intrigante dado que se dice que la visión clásica agrupa a todas las personas en una sola. Parece peculiar suponer que los altruistas perfectos no tomarían en cuenta la distinción entre personas y apoyarían la irrestricta agregación interpersonal que justamente da pauta a dicha omisión. Sin embargo, después de caracterizar al utilitarismo clásico como la ética de los altruistas perfectos, en las páginas siguientes Rawls se pregunta cuál sería la teoría de la justicia preferida por un espectador imparcial que no integrara todos los sistemas de deseos en uno solo. Como respuesta, argumenta que una persona “benevolente” que se ajuste a esta descripción podría preferir en realidad a la justicia como equidad en lugar del utilitarismo clásico. Pero esto vuelve menos claro por qué el utilitarismo clásico debiera estar asociado con el altruismo perfecto. Si dicha asociación no está justificada, entonces el contraste entre las visiones clásica y promedio puede ser menos dramática de lo que sugiere Rawls, y los reclamos de la posición original como esclarecedor recurso analítico podrían reducirse en esa medida.

Por supuesto, esto no refuta que el principio de utilidad promedio tendría mayor atractivo que el utilitarismo clásico para las partes en la posición original. En efecto, mientras que la afirmación de Rawls de que las partes rechazarían al utilitarismo clásico ha concitado escasa oposición, su idea de que su concepción de la justicia sería preferida al principio de utilidad promedio ha sido muy controvertida⁵. Mucha de la controversia se ha centrado en el argumento de Rawls de que sería racional para las partes utilizar la regla maximin de elección bajo incertidumbre cuando decidan seleccionar una concepción de justicia. Si usan esta regla, entonces rechazarán la utilidad promedio a favor de sus dos principios, dado que la regla maximin orienta a los electores

⁵ Ver por ejemplo, Kenneth Arrow, “Some Ordinalist-Utilitarian notes on Rawls’s Theory of Justice”, en *Journal of Philosophy*, 70 (1973): pp. 245-263; Brian Barry, *The Liberal Theory of Justice* (1975) Oxford, Clarendon Press, ver cap. 9 [Nota del traductor. Existe traducción al español, México, FCE, 1993]; Holly Smith Goldman, “Rawls and Utilitarianism”, en *John Rawls’s Theory of Social Justice: An Introduction*. H. Gene Blocker y Elizabeth H. Smith (eds.) (Athens, University of Ohio Press, 1980) pp. 346-94.; R.M. Hare, “Rawls’ Theory of Justice”, en *Reading Rawls*. Norman Daniels (ed.), (Stanford University Press, 1989), pp. 81-107; John Harsanyi, “Can the Maximin Principle Serve as a Basis of Morality? A Critique of Rawls’s Theory”, en *Essays on Ethics, Social Behavior, and Scientific Explanation*, Dordrecht, Holanda, D.Reidel Editor, 1976) pp. 37-63; David Lyons, “Nature and Soundness of the Contract and Coherence Arguments”, en *Reading Rawls*. Op.cit., pp.141-67; Ian Narverson, “Rawls and Utilitarianism”, en *The Limits of Utilitarianism*. Harlan B. Miller y William H. Williams (eds.) (Minneapolis, University of Minnesota Press, 1982) pp. 128-43; Robert Paul Wolff, *Understanding Rawls*. (Princeton University Press, 1977). Ver cap. XV [Nota del traductor. Existe versión al español, México, FCE, 1981].

a escoger la alternativa cuyo peor resultado sea superior al peor resultado de cualquier otra alternativa, y los dos principios sean aquéllos que una persona escogería si supiera que su enemigo fuera a asignarle su lugar en la sociedad. No obstante, como reconoce Rawls, la regla maximin es muy conservadora, y su empleo sería racional sólo bajo ciertas condiciones. A menos que la decisión que enfrentan las partes en la posición original satisfaga tales condiciones, el principio de utilidad promedio puede ser una mejor opción para las partes incluso si ésta es más arriesgada, dado que podría dejar fuera la perspectiva de una ganancia superior (*TJ*, p.165-6 / 143 rev.). Sin embargo, Rawls argumenta que la posición original tiene elementos que vuelven apropiada la confianza en el principio maximin y que las partes rechazarían el principio de utilidad promedio por su excesivo peligro.

En general, el uso del principio maximin se aduce como racional cuando no existe una base confiable para evaluar las posibilidades de resultados diferentes; cuando el elector se preocupa muy poco por las ganancias por encima del mínimo que pudieran estar aseguradas mediante la confianza en el máximo; y cuando las demás opciones tienen posibles consecuencias que el elector hallará intolerables. La estrategia de Rawls es tratar de establecer que la opción entre la utilidad promedio y sus dos principios satisfacen estas condiciones porque [1] las partes no tienen bases para confiar en el tipo de razonamiento probabilístico que soportara una elección del principio de utilidad promedio, [2] sus dos principios asegurarán un mínimo satisfactorio a las partes, y [3] el principio de utilidad promedio podría tener consecuencias que las partes no pudieran aceptar. A pesar de que Rawls primeramente delinea su estrategia en la Sección 26, resulta importante enfatizar que lo que proporciona en dicha sección es solamente un guión de la “estructura cualitativa del argumento que necesita ser elaborado si la posibilidad para ambos principios fuera decisiva” (*TJ*, p.150 / 130 rev.).

Con respecto a la primera condición, Rawls observa en la Sección 28^a que, desde el punto de vista de la posición original, el atractivo *prima facie* de la utilidad promedio depende del supuesto de que uno tiene igual oportunidad de volverse alguien una vez que el velo de la ignorancia es levantado. Sin embargo, Rawls dice que este supuesto “no está fundado sobre factores conocidos de la propia sociedad” (*TJ*, p.168 / 149 rev.). En su lugar, está basado en el principio de la razón insuficiente, el cual en ausencia de algún fundamento específico en la asignación de probabilidades para diferentes resultados, trata todos los resultados posibles como igualmente probables. Rawls dice que, dada la importancia de la elección que enfrentan las partes, sería imprudente para ellas confiar en probabilidades alcanzadas de esa manera. Cabe suponerlo, pues la maximización de la utilidad promedio podría requerir, en sociedades con ciertas características, que los intereses de algunas

personas estén seriamente comprometidos. Por tanto, el riesgo excesivo de apoyarse en el principio de razón insuficiente depende del supuesto acerca de la tercera condición, es decir, de la posibilidad de que el principio de utilidad promedio pueda orientarse hacia resultados intolerables. Esta posibilidad emerge, como sugiere Rawls, porque el utilitarismo se apoya por completo en ciertos “supuestos regulares” (*TJ*, p.159 / 137-8 rev.) para demostrar que sus cálculos normalmente no soportarán restricciones severas en las libertades individuales. Dada la importancia de que las partes se adhieran a las libertades básicas, Rawls mantiene que las mismas “preferirían asegurar de inmediato sus libertades más que hacerlas depender de lo que puedan ser cálculos actuariales inciertos y especulativos” (*TJ*, p.160-1 / 138-9 rev.). Las partes no estarían dispuestas a tomar la opción de que en una sociedad gobernada por principios utilitarios, un cálculo utilitario pueda proporcionar algún día la base para una seria violación de sus libertades, sobre todo desde que tienen a mano la opción más conservadora de los dos principios disponibles para ellos.

Si la conclusión de que las partes asumieran el principio de utilidad promedio como excesivamente peligroso depende del supuesto de que, en ciertas condiciones, se justificaría el sacrificio de las libertades de ciertas personas a fin de mantener el nivel promedio de bienestar dentro de la sociedad tan alto como sea posible, entonces los argumentos de Rawls en contra de la utilidad promedio no son tan diferentes de sus afirmaciones en contra del utilitarismo clásico, como podría sugerir su expresión de un “contraste sorprendente”. En ambos casos se dice de las partes que temen que sus propios intereses se vean sacrificados en nombre de la meta utilitaria más grande. Y en ambos casos, desde la perspectiva de las partes, este argumento corresponde a una crítica independiente del utilitarismo en tanto está excesivamente dispuesto a sacrificar a algunas personas en aras del bienestar de otras.

Defensores del principio de utilidad promedio han desafiado los argumentos de Rawls de diversas maneras. Primero, han afirmado que los “supuestos regulares” son suficientemente robustos, por lo que no sería excesivamente arriesgado para las partes elegir a la utilidad promedio, incluso si ésta significara descansar en el principio de razón insuficiente. Sin embargo, en segundo término, se han preguntado por qué, si Rawls cree que sería sin duda arriesgado para las partes apoyarse en probabilidades que no están sustentadas en información sobre su sociedad, falla en proporcionarles dicha información. Mientras el velo de la ignorancia prevenga a las partes de conocer sus respectivas identidades, proveyéndolas con la información relevante acerca de su sociedad, éstas no necesitan comprometer su imparcialidad. En tercer lugar, han cuestionado que los principios de Rawls en verdad puedan garantizar un mínimo satisfactorio a las partes, y que las partes, que ignoran

sus concepciones del bien, se preocupen apenas por las ganancias por encima de dicho mínimo. En cuarto lugar, han argumentado que los propios principios de Rawls sobre la justicia no están exentos de riesgo, dado que la “concepción general” de la justicia como equidad permitiría la vulneración de libertades básicas en condiciones extraordinarias. Por ende, han mantenido que no existe mayor diferencia entre lo que Rawls señala por utilidad promedio y su propia postura respecto a su peligrosidad. En condiciones normales nadie permitiría restricciones serias a la libertad, pero sí lo harían en situaciones extraordinarias. Finalmente, sus críticos han señalado que existe una opacidad fundamental en la postura de Rawls acerca de la manera en que las partes evalúan el riesgo. Rawls dice que la elección de principios no debería depender de las “actitudes especiales” de las partes respecto al riesgo y que, por tanto, el velo de la ignorancia debería prevenirles de conocer “si ellos tienen una aversión característica a tomar opciones o no” (*TJ*, p.172 / 149 rev., enunciado corregido). En cambio, el propósito es mostrar que “elegir como si uno tuviera una aversión así es racional dados los rasgos únicos de... [la posición original] con independencia de cualquier actitud especial hacia el riesgo” (*TJ*, 172 / 149 rev., enunciado corregido). Sin embargo, los defensores de la utilidad promedio han cuestionado que tenga sentido suponer que exista una actitud hacia el riesgo que sea racional si uno es ignorante de las propias “actitudes especiales” frente al mismo.

Estos temas han sido ampliamente discutidos, por lo que aquí simplemente afirmaré que, a pesar de algunos aspectos desafortunados en la presentación de Rawls, creo que está en lo correcto al sostener que las partes preferirían sus dos principios al principio de utilidad promedio. Al llegar a esta conclusión resulta importante prevenirse contra una interpretación excesivamente estrecha y formalista del “argumento maximin”⁶. Como ya se advirtió, la formulación inicial de Rawls en la Sección 26^a de las razones para apoyarse en la regla maximin es simplemente un esbozo de lo que intentará establecer posteriormente. No debería ser interpretado, como en algunas ocasiones lo ha sido, como la presentación auto-contenida de un argumento formal teórico-decisorio que es independiente, por ejemplo, de las apelaciones en la Sección 29^a a la estabilidad, al respeto por sí mismo, y a los vínculos de compromiso. El efecto inevitable de una interpretación como ésa es hacer que el argumento de Rawls parezca tanto más formal como menos plausible de lo que en realidad es.

De hecho, Rawls establece explícitamente que los argumentos de la Sección 29^a “se ajustan dentro del esquema heurístico sugerido por las razones

⁶ Mi discusión sigue la de Steven Strasnick en su recensión de *Understanding Rawls*, de Robert Paul Wolff, en *Journal of Philosophy*, 76 (1979), pp. 496-510; y la de Joshua Cohen, “Democratic Equality”, en *Ethics*, 99 (1989), pp. 727-51.

para seguir la regla maximin. Esto hace que nos ayuden a mostrar que los dos principios son una concepción mínima adecuada de la justicia en una situación de gran incertidumbre. Cualquier ventaja adicional que pueda ser ganada por el principio de utilidad... es enormemente problemática, mientras que la dificultad si las cosas salen mal es intolerable" (TJ, p.175 / 153 rev., enunciado corregido). En otras palabras, los argumentos de la Sección 29^a están pensados para ayudar a ver que la elección que confronta a las partes tiene rasgos que vuelven racional la confianza en la regla maximin. Estos no son argumentos desvinculados.

En efecto, la cosa va más lejos. Los argumentos enunciados en la Sección 29^a explícitamente invocan consideraciones de psicología moral que no se desarrollarán plenamente hasta la Parte III. En otras palabras, las apelaciones de la Sección 29^a a favor de estabilidad psicológica, del respeto a sí mismo, y de los vínculos de compromiso están todos pensados como contribuciones a la amplia empresa de demostrar que los principios de Rawls proporcionarían un mínimo satisfactorio, mientras que el principio de utilidad promedio podría tener consecuencias con las cuales las partes encontrarían difícil vivir. En formas ligeramente diferentes, sin embargo, todas estas apelaciones están suscritas por el contraste que Rawls desarrolla con amplitud en la Parte III entre las psicologías morales de las dos teorías. Rawls postula allí que debido a que sus principios incorporan una idea de reciprocidad o beneficio mutuo, y dado que la reciprocidad es el mecanismo psicológico fundamental implicado en el desarrollo de la motivación moral, los factores que orientarían a la gente a interiorizar y mantener sus principios son psicológicamente continuos con mecanismos de motivación moral más primitivos en su desarrollo. Esto significa que, en una sociedad cuya estructura básica estaba regulada por los dos principios, la adhesión a dichos principios, bajo condiciones favorables, se desarrollaría naturalmente al margen de factores psicológicos preexistentes. En contraste, el utilitarismo no incorpora una idea de reciprocidad. Si la gente se inclina a estar regularmente motivada a sostener los principios y las instituciones utilitarias, incluso cuando tales principios e instituciones no han laborado en beneficio suyo, la capacidad para una identificación de afinidad tendrá que ser el mecanismo psicológico operativo. Aunque dicha capacidad, en tanto regla, ni es bastante fuerte ni está ciertamente bastante situada dentro del repertorio de la motivación humana para ser una fuente confiable de apoyo a los principios e instituciones utilitarias.

Como he indicado, porciones sustanciales de la Parte III están dedicadas a la elaboración detallada de este contraste junto con sus implicaciones para la estabilidad relativa de las dos concepciones antagónicas de justicia y de su relativo éxito para alentar el auto-respeto de los ciudadanos⁷. Más aún, Rawls

menciona explícitamente que mucho del argumento de la Parte II, que aplica sus principios a las instituciones, está dirigido a establecer que éstos constituyan una concepción plausible de la justicia y proporcionen un mínimo satisfactorio (*TJ*, p.156 / 135 rev.). El resultado es que las razones para apoyarse en la regla maximin, lejos de estar plenamente elaboradas en la Sección 26^a, en realidad son la materia de buena parte del resto del libro^{8,9}. En efecto, el “argumento maximin” funciona como un postulado rector dentro del cual varios de los argumentos más específicos del libro están subsumidos. Bajo esta luz, el significado del argumento como una contribución a la crítica del utilitarismo resulta más fácil de apreciar. Por sí mismo, la afirmación de que

⁷ En *Liberalismo Político*, (pp. xvii-xx y xlvi-xlii), Rawls afirma que el argumento de estabilidad dado en la Parte III de la *Teoría* es defectuoso porque pone a prueba a las concepciones antagónicas de la justicia al preguntarse si la sociedad bien ordenada asociada con cada concepción continuaría generando su propio apoyo en el tiempo, y al hacerlo esta postura asume implícitamente que en una sociedad bien ordenada cualquiera suscribe la concepción sobre la base de una “doctrina moral compartida”. A la vista de una inevitable diversidad de doctrinas generales razonables dentro de una sociedad democrática moderna, Rawls argumenta que esto no es un supuesto realista y por tanto, la prueba de estabilidad es inadecuada. Tal vez sea así, pero Rawls no debería conceder mucho aquí. Puede estar en lo cierto al pensar que necesita mostrar cómo una sociedad regulada por su concepción de justicia puede ser estable a pesar de la preeminencia de diversas teorías generales. Ciertamente, empero, si es verdad que la sociedad bien ordenada utilitaria no continuaría generando su propio apoyo incluso si todo el mundo apoyara inicialmente los principios utilitarios de la justicia sobre la base de un compromiso compartido con el utilitarismo en cuanto doctrina filosófica general, entonces eso permanece como una objeción significativa a la postura utilitaria.

⁸ Por tanto, pienso que es desorientador cuando Rawls dice, al final de su discusión de la estabilidad relativa en la sección 76^a: “Estas consideraciones no están pensadas como razones justificativas de la postura del contrato. Las bases principales para los principios de justicia ya han sido presentadas. En este punto, simplemente se está revisando si la concepción ya adoptada es una posible y no tan inestable que podría hacer mejor a otra opción. Estamos en la segunda parte del argumento en la cual preguntamos si el reconocimiento previamente hecho deba ser reconsiderado” (*TJ*, p. 504 / 441 rev.). Después de todo, en la sección 29^a había dicho [a] que el argumento de la estabilidad es uno de los “argumentos principales para los dos principios” (*TJ* p.175 / 153 rev.); [b] que éste encaja “bajo el esquema heurístico sugerido por las razones de seguir a la regla maximin” (*TJ*, p. 175 / 153 rev.) y [c] que ésta depende de “las leyes de la psicología moral y de la disponibilidad de motivos humanos”, los cuales sólo son discutidos “mas adelante (Secciones 75^a-76^a)” (*TJ*, p.177 / 154 rev.) estos puntos implican que la discusión en la sección 76^a es una parte indispensable de la presentación de “las bases elementales para los principios de la justicia”.

⁹ En “Justice and the Problem of Stability”, (*Philosophy and Public Affairs*, 18, [1989], pp. 3-30) Edward McClenen reconoce que “una lectura cuidadosa de *Una Teoría de la Justicia* hace claro que sería un error tratar al argumento de la elección racional de la construcción analítica (esto es, la posición original) y la teoría psicológica vinculada con el sentido de la justicia como completamente disociadas [p. 6-7]. Sin embargo, el propio McClenen cree que la concepción de justicia de Rawls puede ser defendida mejor con independencia de la construcción de la posición original mediante un argumento de elección racional que realice una apelación más directa a consideraciones de estabilidad psicológica.

incluso la versión común del utilitarismo está completamente deseosa de sacrificar algunas personas por el bien de otras no es novedosa. Sin embargo, al apoyar la resistencia de las partes para aceptar los sacrificios asociados con la utilidad promedio en una psicología moral cuidadosamente elaborada y en un desarrollado argumento acerca de cómo un conjunto eficiente y factible de instituciones sociales podría evitar dichos sacrificios, Rawls fortalece y enriquece en forma considerable dicha crítica conocida. Por lo menos, su argumento reta a los utilitaristas a proporcionar una descripción comparativamente plausible y detallada de las instituciones sociales y económicas utilitarias, y de los procesos mediante los cuales, en una sociedad regulada por los principios utilitarios, se desarrollarían motivos que serían capaces de generar un apoyo creciente continuo para dichas instituciones y principios. La fuerza de este reto, sin embargo, está ampliamente al margen de las propuestas de Rawls sobre la justificación del significado de la construcción de la posición original. Incluso si los utilitaristas rechazaran la posición original como un recurso para elegir entre ambas concepciones rivales de la justicia; en otras palabras, este reto no es de los que ellos puedan fácilmente ignorar.

Mientras que el argumento maximin es presentado como una razón del por qué las partes no escogerían al utilitarismo, Rawls desarrolla otra línea importante de crítica cuya relación ostensible con la construcción de la posición original es menos directa¹⁰. Esta línea de crítica resalta un contraste entre las perspectivas que asumen que sólo hay un simple bien racional para todos los seres humanos y las que conciben los bienes humanos como heterogéneos. Podríamos hablar aquí de un contraste entre las posturas *monista* y *pluralista* del bien. Rawls cree que las teorías teleológicas, las que definen el bien con independencia de lo correcto, y lo correcto como maximizador del bien, tienden también a interpretar al bien en términos monistas. Piensa que esto es verdad para aquellas teorías teleológicas que él describe como perfeccionistas, para ciertas visiones religiosas, y también para el utilitarismo clásico en tanto su descripción del bien sea entendido hedonísticamente. Formas no teleológicas del utilitarismo, tales como el principio de utilidad¹¹, también son monistas si se apoyan en una interpretación hedonista del bien. Sin

¹⁰ Para discusiones útiles de esta línea de crítica, ver Samuel Freeman, "Utilitarianism, Deontology and the Priority of Right", *Philosophy and Public Affairs*, 23 [1994]: pp. 313-49. T.M. Scanlon, "Rawls' Theory of Justice", *University of Pennsylvania Law Review*, 121 [1973]: pp. 1020-69.

¹¹ Ver *TJ*, p. 66 /143 rev., donde Rawls dice que el principio de utilidad promedio "no es una doctrina teleológica, estrictamente hablando, como sí lo es la visión clásica", dado que ésta se encamina hacia maximizar un promedio y no una suma. Sin embargo, advierte que en el índice de entrada existe para el utilitarismo promedio (*TJ*, p. 606 / 538 rev.) un subtítulo que dice: "como teoría teleológica, hedonismo, tendencia al".

duda, las formas hedonistas del utilitarismo reconocen que los individuos obtendrán placer en formas muy distintas, por lo que son superficialmente proclives al pluralismo de una manera que no lo son las otras visiones monistas. Por esta precisa razón, Rawls sugiere que el utilitarismo ofrece “una manera de adaptar la noción del bien racional único a los requerimientos institucionales de un Estado moderno y de una sociedad democrática”¹². Sin embargo, en tanto se identifique al bien con una sensación agradable, la postura permanece monista¹³.

Rawls sugiere que las visiones teleológicas pueden ser transformadas en argumentos monistas a partir de un deseo de evitar la indeterminación en el modo de caracterizar el bien, dado que para las visiones teleológicas “cualquier vaguedad o ambigüedad en la concepción del bien se transfiere a la de lo correcto” (*TJ*, p.559 / 490 rev.). También sugiere que parte del atractivo de los argumentos monistas, y de las teorías teleológicas que los incorporan, puede derivar de la convicción de que nos facultan para resolver un problema fundamental acerca de la naturaleza de la deliberación racional. El problema es explicar cómo pueden realizarse elecciones racionales entre opciones aparentemente heterogéneas. A menos de que exista un fin último al cual se dirija toda acción humana, este problema puede parecer irresoluble. No obstante, si existe algún “fin dominante” al cual todos los demás fines estén subordinados, entonces “una decisión racional es siempre posible en principio, dado que sólo permanecen dificultades en computación y falta de información” (*TJ*, p.552 / 484 rev.). Además, si existe en efecto un fin dominante al cual se dirige toda acción humana, entonces no queda sino un simple paso para construir dicho fin como el “bien único intrínseco” (*TJ*, p.556 / 487 rev.) para los seres humanos. De esta manera, podríamos estar encaminados hacia una descripción monista del bien “mediante un argumento desde las condiciones de la deliberación racional” (*TJ*, p.556 / 487 rev.). Una vez que hemos aceptado una postura monista del bien, una visión teleológica dirigiéndonos a maximizar dicho bien puede parecer plausible. Como dice Rawls: “las visiones teleológicas ejercen una profunda e intuitiva atracción, dado que parecen incorporar la idea de racionalidad. Es natural pensar que la racionalidad implica maximizar algo y que eso en moral debe ser maximizar el bien” (*TJ*, p.24-5 / 22 rev.). Además, el hedonismo es “la tendencia sintomática de las teorías

¹² “Social Unity and Primary Goods”, en *Utilitarianism and Beyond*. Amartya Sen y Bernard Williams (eds.) (Cambridge, Cambridge University Press, 1982) pp. 159-85, en p. 182; también en John Rawls, *Collected Papers*. Samuel Freeman (ed.) (Cambridge, Harvard University Press, 1999), pp. 359-87, en p. 384.

¹³ A la luz de esta evaluación de la concepción utilitarista del bien y de su propia defensa de una concepción pluralista, el comentario de Rawls en la Sección 16^a de que el utilitarismo y su teoría conceden que “el bien es la satisfacción del deseo racional” (*TJ*, pp. 92-3 / 78-80 rev.) parece muy desorientadora.

teleológicas” (*TJ*, p.560 / 490 rev.), tanto porque la sensación agradable puede parecer un “medio de cambio interpersonal” (*TJ*, p.559 / 490 rev.) que vuelve posible la elección social, como porque la superficial hospitalidad del hedonismo hacia formas variadas de vida le permite “evitar la apariencia de fanatismo e inhumanidad” (*TJ*, p. 556 / 487 rev.). Así, Rawls cree que existe una cadena de argumentación que comienza con una preocupación acerca de la posibilidad de la decisión racional y concluye con un respaldo al utilitarismo hedonista.

Contrario a esta línea de pensamiento, Rawls argumenta, primero, que simplemente no existe un fin dominante: ningún objetivo general por el que todos nuestros demás fines sean perseguidos. La sensación placentera o agradable, en particular, no puede pensarse plausiblemente que constituya un tal objetivo. Ni tampoco, afirma, la irreducible diversidad de nuestros fines significa que la elección racional sea imposible. Lo que todo esto significa es que los principios formales juegan un papel limitado para determinar tales elecciones. A menudo la elección racional debe apoyarse, más que en el auto-conocimiento, en un intento cuidadoso por discernir cuál, entre un diverso conjunto de fines, importa más para nosotros. Al llevar a cabo semejantes determinaciones, bien podemos emplear la “racionalidad deliberativa” –reflexionar cuidadosamente, en condiciones favorables, a la luz de todos los hechos relevantes disponibles para nosotros–, pero no existe procedimiento formal alguno que seleccione rutinariamente el curso racional de la acción. Y puesto que no hay fin dominante de toda la acción humana racional, continúa Rawls, es implausible suponer que el bien sea monista. En cambio, menciona que “el bien humano es heterogéneo porque los objetivos del yo son heterogéneos” (*TJ*, p.554 / 486 rev.). Esto depura mucho la motivación a favor de una perspectiva teleológica. Mientras la idea de organizar las instituciones sociales para maximizar el bien parecería atractiva si existiera un bien único al cual se dirija toda acción racional, ésta tiene más sentido, a la luz de la heterogeneidad del bien, para establecer un marco justo de cooperación social dentro del cual los individuos puedan perseguir sus diversos fines y aspiraciones.

El significado de esta crítica está sujeto a dudas de dos clases diferentes. Primero, parecería que la crítica simplemente no se aplica a las versiones contemporáneas del utilitarismo, lo cual no implica en general construir el bien de manera hedonista. Cuando tales opiniones defienden la maximización de la satisfacción total o promedio, su tema es la satisfacción de las preferencias de la gente y no algún estado de conciencia asumido. Sin embargo, en “*La Unidad Social y Bienes Primarios*”, donde construye un argumento inicialmente introducido en los cuatro párrafos finales de la Sección 28^a de la *Teoría de la Justicia*, Rawls sostiene que incluso las versiones contemporáneas del utilitarismo son a menudo encubierta o implícitamente hedonistas.

Lo son porque éstas descansan en una especie de “función compartida del orden de preferencias más alto” en cuanto base para las comparaciones interpersonales de bienestar; y dicha función trata a los ciudadanos como adherentes a un escalafón común de deseabilidad relativa de paquetes diversos de recursos materiales y cualidades personales –incluyendo rasgos de carácter, capacidades y habilidades, vínculos y lealtades, fines y aspiraciones–. Los ciudadanos racionales asumen, pues, desear un paquete general con un rango tan amplio como sea posible. Esta premisa, argumenta Rawls, implica “la disolución de la persona en cuanto que conduce una vida que expresa carácter y devoción a fines específicos”, y esto es “psicológicamente inteligible”¹⁴ sólo si uno piensa el placer como un fin dominante en nombre del cual una persona racional está deseosa de revisar o abandonar cualquiera de sus otros fines o compromisos. Si este análisis es correcto, entonces el argumento de Rawls puede aplicarse a un más amplio radio de teorías utilitarias de lo que era inicialmente manifiesto.

Sin embargo, la relación oblicua del argumento con la construcción de la posición original puede dar paso a dudas de otra especie. En la posición original, el argumento no se presenta a las partes como una razón para rechazar el utilitarismo o las visiones teleológicas en general. Antes bien, parece jugar un papel en motivar el diseño de la propia posición original. Ayuda a explicar por qué se niega a las partes el conocimiento de cualquier concepción específica del bien, y por qué, en cambio, se las orienta a aceptar la teoría débil del bien, con todo lo que ello implica. Dados estos puntos de partida, parece improbable por anticipado que las partes acepten cualquier teoría de la justicia que descansen en una concepción hedonista o en otra monista del bien. Como dice Rawls, “las partes... no saben qué objetivos finales tienen las personas, y todas las concepciones del fin dominante son rechazadas. Por tanto, no se les ocurriría reconocer el principio de utilidad en su forma hedonista. No existe más razón para que las partes estén de acuerdo con este criterio que maximizar cualquier otro objetivo particular. (*TJ*, p. 563 / 493 rev.). Pero esto sugiere que las partes rechazan teorías de la justicia que incorporan visiones monistas del bien debido a que el argumento de Rawls a favor del pluralismo les ha orientado a diseñar la posición original de tal forma que garantice que ellos lo harán. Esto en cambio puede generar dudas en el significado y la justificación de las preferencias de las partes. Y dado que sus opciones representan el núcleo de la posición “oficial” de Rawls en contra del utilitarismo, un efecto de la forma en que despliega el argumento en contra del monismo puede poner en riesgo dicha posición. En sus últimos escritos, Rawls mismo manifiesta recelos respecto del papel jugado en *Una Teoría de*

¹⁴ “Unidad Social y Bienes Primarios”, p. 181.

la Justicia por su defensa de una teoría pluralista del bien. En *Justicia como Equidad: Política, no Metafísica*, describe como una de las “fallas” de la teoría de la justicia el que la versión de la idea del bien desarrollada en la Parte III “a menudo se lee como una explicación del bien perfecto a favor de una concepción moral comprehensiva”¹⁵. Y en *Liberalismo Político* reformula el argumento en contra de las concepciones monistas del bien; el punto ya no es que estén equivocadas, sino que ninguna corriente puede servir como base para una adecuada concepción de la justicia dentro de una sociedad pluralista”¹⁶.

Estas consideraciones implican algunas cuestiones generales significativas –acerca de la función justificativa de la posición original y acerca de los cambios en las visiones de Rawls en el tiempo– que escapan al objeto de este ensayo. Sin embargo, incluso si el papel del argumento en contra del monismo en la *Teoría de la Justicia* suscita preguntas acerca del significado justificativo de la construcción de la posición original, e incluso si el carácter filosófico del argumento está en tensión con el giro “político” tomado en los últimos escritos de Rawls, yo creo que el argumento puede apoyarse en sí mismo como un desafío importante al pensamiento utilitarista. Lo mismo es verdad, como ya he sugerido, respecto de la postura de Rawls de que el utilitarismo tolera acuerdos interpersonales inaceptables. En efecto, considero que ambos argumentos representan sus críticas más importantes y duraderas a la tradición utilitarista.

Pese al vigor de sus argumentos en contra del utilitarismo, empero, algunos de sus críticos han sostenido que la propia teoría de Rawls desarrolla algunos de los mismos elementos que él critica a la posición utilitarista. Por ejemplo, Robert Nozick mantiene que existe cierta tensión entre la afirmación de Rawls de que el principio de diferencia “representa, en efecto, un acuerdo para vigilar la distribución de los talentos naturales como un recurso común y para compartir los beneficios de esta distribución” (*TJ*, p.101 / cfr. *TJ*, p.87 rev., enunciado corregido) y su acusación de que el utilitarismo clásico no asume seriamente la distinción entre personas. Nozick sugiere que Rawls puede evitar esta tensión tan sólo concediendo un peso no plausible a la distinción entre las personas y sus talentos¹⁷. Michael Sandel, apoyando el punto de Nozick, argumenta que Rawls tiene una “teoría de la persona” en virtud de la cual los talentos son sencillamente “atributos enteramente insustanciales y

¹⁵ *Philosophy and Public Affairs*, 14 (1985): pp. 223-51; en particular p. 251 nota; y Rawls, *Collected Papers*, pp. 388-414, en particular, p. 414, nota. Ver también *LP*, pp. 176-7, nota.

¹⁶ Ver por ejemplo, *LP*, pp. 134-5; ver también “Justicia como Equidad: Política no Metafísica”, pp. 248-9.

¹⁷ *Anarchy, State and Utopia*, p. 228

contingentemente dados”, más que “*constituyentes* fundamentales” del yo¹⁸. Por esta razón, argumenta Sandel, Rawls no ve la distinción de *personas* como vulnerada por la idea de tratar la distribución de talentos como un recurso común. Sin embargo, Sandel cree que la teoría subyacente de la persona adolece de “incoherencia”¹⁹, y no puede por tanto proporcionarle a Rawls una respuesta satisfactoria al cargo de que también él es culpable de menospreciar la condición distintiva de las personas.

Sandel sostiene que “la única manera de salir de las dificultades que propone Nozick” sería argumentar que lo que apoya el principio de la diferencia es una concepción ‘intersubjetiva’ de la persona según la cual “la descripción relevante del yo puede abarcar más que a un simple ser humano empíricamente individualizado”²⁰. Esto posibilitaría a Rawls poder afirmar que las demás personas que se benefician de mis talentos naturales no necesitan violar la especificidad de las personas, no porque mis talentos no sean en realidad parte de mí, sino más bien porque dichas personas no pueden, en un sentido relevante, ser distintas de mí. Por supuesto, decir lo anterior sería conceder que Rawls adopta las distinciones convencionales entre “los seres humanos empíricamente individualizados” aún menos en serio de lo que lo hace el utilitarismo. En efecto, se considera una gran moraleja del análisis de Sandel el que el principio de diferencia sea una noción suficientemente comunitarista de la justicia que requiera de una completa concepción comunitarista del yo.

No obstante, considero que el análisis de Sandel magnifica innecesariamente los intereses metafísicos y que la tensión entre los principios de Rawls y su crítica del utilitarismo pueden ser disueltas sin apelar a cualquiera de las dos teorías de la persona que Sandel invoca. El utilitarismo clásico identifica la vida buena para un individuo como una vida de felicidad o de satisfacción. Sin embargo, ello nos encamina a ordenar las instituciones sociales de una forma tal que maximice la satisfacción o el bien agregados, incluso si esto significa que la habilidad de algunos individuos para tener vidas buenas –en términos utilitaristas– estuviera seriamente comprometida, y aun cuando no exista un ser viviente que experimente la satisfacción agregada o cuyo bien se identificara con dicho agregado. En este sentido, el utilitarismo clásico otorga (a lo que considera como) al agregado su buena prioridad sobre (lo que considera como) los bienes de los distintos individuos. Para Rawls, en cambio, la vida buena para un individuo consiste en una realización exitosa de un plan de vida racional, y sus principios nos encaminen a ordenar las instituciones so-

¹⁸ *Liberalism and the Limits of Justice*. (Cambridge, Cambridge University Press, 1982), p. 78. [Nota del traductor. Existe versión al castellano en la editorial Gedisa, Barcelona, 2000].

¹⁹ *Liberalism and the Limits of Justice*, p. 79.

²⁰ *Liberalism and the Limits of Justice*, p. 80.

ciales de manera tal que protejan la capacidad de cada individuo para conducir una tal vida. La idea de que la distribución de los talentos naturales debería ser asumida como un recurso común no es la idea de un bien agregado que tome precedencia sobre los bienes de los seres humanos individuales. En vez de eso, la idea es que un sistema que trate la distribución de talentos como un recurso colectivo en los términos del principio de diferencia es en realidad requerido si cada persona ha de tener una oportunidad para conducir una vida buena. En síntesis, el utilitarismo otorga al bien agregado precedencia sobre los bienes de distintos individuos, mientras que los principios de Rawls no lo hacen. En este sentido, el utilitarismo asume las distinciones entre las personas menos seriamente de lo que lo hacen aquéllos. Dado que, según eso, no existe inconsistencia alguna entre los principios de Rawls y sus críticas del utilitarismo, para él no hay necesidad de tomar medidas metafísicas drásticas para evitarlo²¹.

Sin embargo, existen algunas coincidencias genuinas entre la concepción rawlsiana de la justicia y el utilitarismo, y dichas similitudes pueden ser en parte responsables de la percepción de que existe una tensión entre su apoyo a la primera y su crítica a la segunda. Deseo llamar la atención sobre tres de estas coincidencias. La primera, ya mencionada, es la aspiración de Rawls a producir una teoría que comparta el carácter “sistemático” y “cons-

²¹ Ha existido una extensa discusión y desacuerdo tanto acerca del significado como de los méritos del argumento de Rawls de que el utilitarismo no toma seriamente las distinciones entre las personas. Esto es parcial, debido a que la formulación de Rawls ha parecido a ciertos lectores que abarca dos o más de los siguientes cargos: [1] *un reclamo de error metafísico* al hecho de que el utilitarismo simplemente falla en advertir que las personas son ontológicamente distintas; [2] *un reclamo de error moral* al hecho de que el utilitarismo tolera intercambios interpersonales inaceptables, y en consecuencia falla en ofrecer suficiente significado moral a las distinciones ontológicas entre personas, y [3] *un reclamo explicativo* al hecho de que el utilitarismo falla para dar suficiente significado moral a las distinciones ontológicas entre las personas debido a que éste extiende hacia la sociedad como un todo el principio de elección para una persona. Para una discusión pertinente, véase Derek Parfit, *Reasons and Persons*, (Oxford, Clarendon Press, 1984). Ver cap. 15; James Griffin, *Well-Being*, (Oxford, Clarendon Press, 1986), pp. 167-70; H.L.A. Hart, “Between Utility and Rights”, en *The Idea of Freedom. Essays in Honor of Isaiah Berlin*. Alan Ryan (ed.) (Oxford University Press, 1979), pp. 77-98; Leslie Mulholland, “Rights, Utilitarianism, and the Conflation of Persons”, en *Journal of Philosophy*, 83 (1986): pp. 323-40; Will Kymlicka, “Rawls on Teleology and Deontology”, en *Philosophy and Public Affairs* 17 (1988): pp. 173-90; Samuel Freeman, “Utilitarianism, Deontology and The Priority of Right”; Joseph Raz, “The Morality of Freedom”, (Oxford, Clarendon Press, 1986: pp. 271-3). Considero que Rawls no está haciendo el primero de dichos reclamos, que él está haciendo el segundo y que también está adelantando el tercero, al menos como explicación parcial. La cuestión planteada por Nozick y Sandel, como yo la entiendo, es si los propios principios de Rawls adscriben más peso moral a las distinciones entre personas de lo que lo hace el utilitarismo. Mi argumento en el texto es que en efecto lo hace y que su habilidad para realizarlo de esa manera no está condicionada por la concepción rawlsiana del yo o como “desligada” o como “intersubjetiva”.

tructivo” del utilitarismo. La segunda es su acuerdo con la visión utilitarista de que los preceptos de justicia del sentido común sólo tienen una condición “derivada” (*TJ*, p. 307 / 270 rev.), y que deben ser vistos como “subordinados” (*TJ*, p. 307 / 270 rev.) a un “criterio superior” (*TJ*, p. 305 / 268 rev.). Y la tercera es que tanto las posturas de Rawls como las del utilitarismo sobre la justicia distributiva son, en un sentido que debe ser explicado, holistas en su naturaleza. Estos tres puntos de acuerdo, tomados en conjunto, tienen implicaciones de largo alcance. Ayudan a explicar por qué puede ser tentador pensar que los principios de Rawls muestran las mismas fallas que él critica al utilitarismo. Y aunque, como he argumentado, esta tentación debe ser resistida, aquéllos nos ayudan a ver que Rawls sí comparte con los utilitaristas algunos rasgos que son genuinamente polémicos y que están dirigidos a generar alguna resistencia fuerte hacia ambas posturas.

El deseo de Rawls de proporcionar una concepción “constructiva” de la justicia es parte de su deseo de evitar el intuicionismo. El intuicionismo, como lo entiende Rawls, sostiene que hay una pluralidad de principios primeros de justicia que puede causar conflicto en situaciones particulares. Los intuicionistas no creen que existan reglas de prioridad que puedan facultarnos para resolver tales conflictos; en su lugar, no disponemos de más opción que la de apoyarnos en nuestro juicio intuitivo para alcanzar un equilibrio apropiado en cada caso. En este sentido, los intuicionistas rechazan que sea posible dar una solución general a lo que Rawls llama “el problema de la prioridad”, esto es, el problema de cómo asignar valor a las consideraciones conflictivas de la justicia. Pero “la asignación de valores es una parte esencial y no una menor de una concepción de la justicia”, porque si dos personas difieren acerca del valor a asignar a principios diferentes, “entonces sus concepciones de la justicia son diferentes” (*TJ*, p. 41 / 36-7 rev.). En efecto, una concepción intuitiva de la justicia no es entonces “sino una concepción a medias” (*TJ*, p. 41 / 37 rev.). El objetivo de Rawls, en cambio, es el de reducir nuestra confianza en una intuición sin guía al formular “principios explícitos para el problema de la prioridad” (*TJ*, p. 41 / 37 rev.); es decir, al identificar criterios “constructivos” y “éticamente reconocibles” (*TJ*, p. 39 / 35 rev.) para asignar valores a preceptos de justicia que compiten entre sí. Por supuesto, el utilitarismo alcanza este objetivo al identificar un simple principio como el parámetro definitivo para resolver entre preceptos en conflicto. Rawls espera mostrar que es posible para una teoría ser constructiva sin apoyarse en el principio utilitarista o, de hecho, en cualquier principio único, como el parámetro definitivo. Así, espera producir una solución al problema de la prioridad que ofrezca una alternativa a la solución utilitarista, pero permaneciendo pese a todo como una solución constructiva.

El hecho de que Rawls concuerde con el utilitarismo acerca de lo deseable que resulta identificar una solución clara y constructiva al problema de la prioridad, encamina más o menos directamente al segundo punto de acuerdo. Ambas visiones sostienen que los preceptos de justicia del sentido común deben subordinarse a algún principio o principios “de orden superior”. Pues tales principios entran en conflicto, y a nivel del sentido común ninguna reconciliación es posible, dado que no existe una manera determinada de soportarlos entre sí. Por tanto, si estamos por encontrar una solución constructiva al problema de la prioridad, debemos recurrir a un “principio superior” para resolver dichos conflictos. En teoría, uno o más de los preceptos del sentido común podrían ser por sí mismos “elevados” (*TJ*, p. 305 / 268 rev.) a dicho nivel, pero Rawls no cree que sean candidatos plausibles. “Adoptar uno de ellos como un primer principio es garantía de dirigirnos hacia el menosprecio de otras cosas que deberían ser tomadas en cuenta. Y si todos o muchos de los preceptos son tratados como primeros principios, no existe beneficio con claridad sistemática. Los preceptos de sentido común están en el nivel incorrecto de la generalidad” (*TJ*, p. 308 / 271 rev.). Por supuesto, los utilitaristas creen que el principio de utilidad proporciona el requisito del parámetro superior, mientras que Rawls cree que sus dos principios son el “criterio superior correcto” (*TJ*, p. 305 / 268 rev.). Pero ambos concuerdan en la necesidad de dicho criterio, y en el carácter derivado y subordinado de los preceptos de justicia del sentido común.

Para ilustrar el tercer punto de acuerdo, podemos comenzar con hacer notar que Rawls llama la atención –y tiene una considerable simpatía– sobre el fuerte acento institucional que caracteriza a los grandes escritores de la tradición utilitarista. En su primigenio ensayo “Dos Conceptos de Reglas”, por ejemplo, escribe: “Es importante recordar que aquéllos a quienes he llamado utilitaristas clásicos estaban fuertemente interesados en las instituciones sociales. Figuraban entre los economistas y teóricos políticos líderes de su tiempo, y con frecuencia eran reformadores interesados en asuntos prácticos”²². De manera similar, en el prefacio a *Una Teoría de la Justicia*, deplora nuestra tendencia a “olvidar que los grandes utilitaristas, Hume y Adam Smith, Bentham y Mill, fueron teóricos sociales y economistas de primer orden, y que la doctrina moral que ellos elaboraron fue diseñada para resolver las necesidades de sus más amplios intereses y para encuadrarla dentro de un esquema general” (*TJ*, p. VII /xvii rev.). Dentro de la propia teoría de Rawls, por supuesto, las instituciones constituyen el núcleo central desde el principio, puesto que la estructura básica de la sociedad, que abarca a las principa-

²² *Philosophical Review*, 64 (1955) 3-32, en particular, p. 19, nota; también Rawls, *Collected Papers*, pp. 20-46; en particular, p. 33 nota.

les instituciones, es tratada como primer tema de la justicia²³. En consecuencia, esto lleva a la idea de tratar la cuestión de las cuotas distributivas como un asunto de justicia procedural pura (*TJ*, 84-5 / 75 rev.); dado que la estructura básica es justa, cualquier distribución resultante de bienes también es justa²⁴. Una vez que el problema de la distribución es comprendido de esta manera, los principios de justicia no pueden seguir siendo aplicados a transacciones individuales consideradas por separado (*TJ*, 87-8/76 rev.). Ni tampoco la justicia de una asignación general de bienes puede ser evaluada con independencia de las instituciones que la produjeron. Como Rawls menciona:

Una distribución no puede ser juzgada de manera aislada respecto del sistema del cual es resultante o del que los individuos han creado de buena fe a la luz de expectativas establecidas. Si se cuestiona en abstracto si la distribución de cierto conjunto de cosas a individuos definidos con deseos y preferencias conocidas es mejor que otro, entonces simplemente no existe respuesta a esta cuestión. La concepción de los dos principios no interpreta el problema primario de la justicia distributiva como uno de justicia de asignación (*TJ*, p. 88 / 76 rev.).

Como enfatiza Rawls, el utilitarismo no comparte su idea de que “se requieren primeros principios especiales para la estructura básica” (*LP*, p. 262), a pesar de su amplio énfasis institucional; ni tampoco concede que la cuestión de las cuotas distributivas debiera ser tratada como un asunto de justicia procedural pura (*TJ*, 88-9 / 77 rev.). Son éas diferencias importantes entre las dos teorías. El principio de utilidad, al menos como ha venido siendo interpretado, es un parámetro general que se emplea para evaluar ac-

²³ Rawls da su más amplia defensa de su énfasis sobre la estructura básica en “La Estructura Básica como Sujeto”, la cual está incluida en *LP* como la Lección VII. Para una crítica de dicho énfasis, véase G.A. Cohen, “Where the Action Is: On the Site of Distributive Justice”, en *Philosophy and Public Affairs*, 26 (1997), pp. 3-30; Liam Murphy, “Institutions and the Demands of Justice”, en *Philosophy and Public Affairs*, 27 (1998), pp. 251-91; Nozick, *Anarchy, State and Utopia*, pp. 204-13; para una discusión de la distintividad del interés de Rawls en la estructura básica, ver Hugo Bedau, “Social Justice and Social Institutions”, en *Midwest Studies in Philosophy*, 3 (1978), pp. 159-73.

²⁴ La previsión es esencial. Sólo si la estructura básica es regulada por la concepción substantiva de la justicia de Rawls, la determinación de las cuotas individuales puede ser tratada como un asunto de justicia procedural pura. Así, la confianza de Rawls en la justicia procedural pura no significa que su teoría sea de procedimiento *en lugar de* substantiva. Este es un punto que enfatiza en respuesta a Habermas (*LP*, pp. 421-33) y explica que cuando menciona en el índice a *LP* (p. 455) que “la justicia es siempre substantiva y nunca puramente de procedimiento” –un señalamiento que podría parecer de otra manera inconsistente con el papel que la Teoría asigna a la justicia procedural pura. Para una discusión relevante, ver Joshua Cohen, “Pluralism and Proceduralism”, *Chicago-Kent Law Review* 69 (1994): pp. 589-618, en especial la nota al pie de página nº 44.

ciones, instituciones y la distribución de recursos dentro de una sociedad²⁵. La concentración de Rawls en la estructura básica y su uso de la justicia procedural pura para evaluar distribuciones le otorga a su teoría una mayor cobertura institucional. Sin embargo, estas diferencias, importantes como son, no deberían ser reconocidas para empañar un punto importante de acuerdo, esto es, que ninguna visión pretenda evaluar la justicia o la injusticia de una asignación particular de bienes separada del contexto distributivo más amplio. En otros términos, nadie cree que los principios de justicia puedan ser aplicados apropiadamente “a una simple transacción vista en forma aislada” (*TJ*, p. 87 / 76 rev.). La justicia o la injusticia de asignar un beneficio particular a un individuo dado, dependerá para los utilitaristas de si existe otra forma de asignación que pudiera conducir a una distribución general con una utilidad (total o promedio) más grande. Para Rawls, esto dependerá de si la asignación es parte de una distribución general producida por una estructura básica conforme a sus dos principios. En este sentido, tanto Rawls como los utilitaristas adoptan una visión *holista* de la justicia distributiva; ambos insisten en que la justicia de cualquier asignación de beneficios siempre depende –directa o indirectamente– de la justicia de una distribución más grande de beneficios y limitaciones dentro de la sociedad.

El holismo acerca de la justicia distributiva gana apoyo a partir de dos convicciones. La primera es que las vidas de todas las personas son de igual valor e importancia. La segunda es que las perspectivas de vida de los individuos están tan densa y variadamente interrelacionados, en especial a través de su participación compartida en las instituciones y prácticas sociales, que virtualmente cualquier asignación de recursos a una persona tiene implicaciones moralmente relevantes para otras personas. Los holistas concluyen que es imposible evaluar la justicia de una asignación de beneficios para un simple individuo sin tener en consideración el contexto distributivo más amplio de dicha asignación. En condiciones de escasez moderada, no podemos decir si una persona debe recibir un beneficio dado sin saber cómo dicha asignación se ajustaría en la distribución más amplia de beneficios y restricciones dentro de la sociedad. Tampoco, para quienes hallan convincente el holismo, parece conceptualmente sólido u éticamente apropiado el proyecto de identificar una supuesta línea de base para la distribución natural y pre-social de las ventajas,

²⁵ Parte del argumento de Rawls, cuando llama la atención en “Dos Conceptos de Reglas” sobre el interés de los utilitaristas clásicos en las instituciones sociales, era enfatizar que el significado del utilitarismo, en tanto proveedor de un patrón comprehensivo de valoración, representa un desarrollo relativamente reciente de la perspectiva; una que en dicho ensayo asocia con Moore. En su trabajo posterior, sin embargo, es la versión general del utilitarismo la que él mismo trata como patrón y con la cual contrasta su propio enfoque institucional de la justicia. Ver, por ejemplo, la Sección 2^a de “La Estructura Básica como Sujeto”, donde asocia la interpretación comprehensiva con Sidgwick (*LP*, pp. 260-2).

como tampoco evaluar la justicia de todas las distribuciones subsecuentes solamente por referencia a la legitimidad de cada acción fuera de la línea de base.

Las instituciones sociales estructuran la vida de las personas en aspectos fundamentales desde el nacimiento hasta la muerte, por lo que no existe un momento pre-social en la vida de un individuo. Siendo ése el caso, no está claro qué podría contar razonablemente como la línea de base “natural” o qué credenciales éticas de dicha línea de base podrían ser pensadas como plausibles²⁶. Además, como el tamaño de la población sigue creciendo, como la escala y la complejidad de las instituciones modernas y las economías continúan incrementándose, y como una infraestructura tecnológica y de comunicaciones cada vez más sofisticada continúa expandiendo las posibilidades de la interacción humana, los obstáculos en el camino para una condición satisfactoria de la línea de base se vuelven más elevados, y la presión para adoptar una visión holista de la justicia distributiva crece todavía más²⁷. En sus diferentes formas, las posturas rawlsiana y utilitarista de la justicia son igualmente sensibles a dicha presión²⁸.

Aunque la posición del holismo tiene una fuerza considerable, y muchas de nuestras intuiciones acerca de la justicia distributiva son en efecto holistas, existen otras ideas no holistas acerca de la justicia que también poseen un amplio apoyo intuitivo. A menudo, por ejemplo, parecemos preparados para decir que un individuo merece o tiene un derecho a cierto beneficio, y por tanto, es justo que debiera obtenerlo, sin cuestionar el contexto distributivo más amplio. En efecto, de acuerdo a una visión familiar y tradicional, la justicia consiste, al menos en parte, en dar a la gente lo que se les ha dicho que pueden merecer de manera independiente. En otras palabras, existe un

²⁶ Ver Thomas Nagel, *Equality and Partiality* (Oxford University Press, 1991), pp. 100-2. [Nota del traductor. Existe versión al castellano en la editorial Paidós, Barcelona, 1996].

²⁷ He discutido algunas temáticas relacionadas en “Individual Responsibility in a Global Age”, *Social Philosophy and Policy*, 12 (1995): pp. 219-36.

²⁸ Para una discusión afín, ver Thomas Pogge, “Three Problems with Contractarian-Consequentialist Ways of Assessing Social Institutions”, *Social Philosophy and Policy*, 12 (1995): pp. 241-66. Pogge escribe que en las últimas décadas se ha experimentado un giro desde pensar “interactivamente” sobre la justicia, a hacerlo “institucionalmente”. Con esto, él da a entender que hay un menor énfasis en las reglas y prácticas fundamentales que estructuran a nuestro mundo social y proporcionan el marco para nuestras acciones. Pogge argumenta que el contractualismo (como se exemplifica en la obra de Rawls) y el consecuencialismo son las “dos tradiciones más prominentes del análisis moral institucional”, y que ambas “no son tan distintas como se suele creer”. Arguye en efecto que, cuando se las ve como “guías para la evaluación de las instituciones sociales, el contractualismo y el consecuencialismo son para la mayor parte no antagonistas, sino presentaciones alternativas de una simple idea: ambas tienden a revisar esquemas institucionales alternativos exclusivamente a partir de cómo cada una afectaría a sus participantes humanos individuales” (p. 246).

parámetro previo del mérito, por referencia al cual han de evaluarse la justicia de las acciones individuales y los arreglos institucionales. La base para una demanda meritaria válida, según dicha visión, siempre debe ser alguna característica de, o dato sobre, la persona merecedora. En este sentido, meritorio entendido al modo tradicional es *individualista* más que holista. Ninguna evaluación de la distribución general de los beneficios y las cargas en la sociedad, o de las instituciones que produjeron dicha distribución, se requiere normalmente para decidir si un individuo particular merece un cierto beneficio. En vez de eso, es una restricción en la justicia de las distribuciones y las instituciones que deban otorgar a cada individuo lo que cada uno independientemente merece en virtud de datos relevantes sobre él o ella.

Ni Rawls ni los utilitaristas aceptan dicha opinión²⁹. Como es bien sabido, el utilitarismo no considera el mérito como una noción moral fundamental. Los utilitaristas pueden convenir, por supuesto, que a menudo existen buenas razones para las instituciones sociales para asignar beneficios y cargas de acuerdo a criterios públicamente anunciados de méritos y deméritos. También pueden acordar que, cuando un esquema institucional de esta clase está vigente, de los individuos que satisfacen dichos criterios quepa decir que merecen los beneficios y las cargas en cuestión. Pero decir esto implica tratar el mérito como un artefacto institucional más que como una limitación normativa en el diseño de instituciones justas. Esto no supone decir que las instituciones justas deben ofrecer a la gente lo que por separado merece. Más bien, el punto es que si las instituciones justas han anunciado que asignarán las recompensas en función de ciertos parámetros, cabe entonces decir que los individuos que cumplan con dichos criterios merecen las recompensas anunciadas.

Rawls concuerda con los utilitaristas en que los individuos que han hecho lo que una institución justa ha publicitado recompensar, están avalados para recibir las recompensas en cuestión. Las instituciones justas deben cumplir con las expectativas que han creado, y en la medida en que hablar de méritos es sólo una manera de referirse a tales derechos institucionales, Rawls es feliz acordando que las instituciones justas deben dar a la gente lo que merece. Para el utilitarismo, en cambio, mérito en este sentido depende de una concepción previa de la justicia y de las expectativas realmente establecidas por las instituciones justas. Rawls pone el punto como sigue:

Es necesario ser claros acerca de la noción de mérito. Es perfectamente cierto que dado un sistema justo de cooperación, en tanto esquema de reglas públicas y de las expectativas erigidas por éste, aquéllos que,

²⁹ Aquí retomo mi discusión en “Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Ethics”, *Philosophy and Public Affairs* 21 (1992): pp. 4-22.

con el prospecto de mejorar su condición, han hecho lo que el sistema anuncia que será recompensado, tienen derechos a obtener ventajas... Aunque este sentido del mérito presupone la existencia del esquema cooperativo, éste es irrelevante para la cuestión... [de cómo] en primer lugar debe ser diseñado el esquema... (*TJ*, p. 103 / 88-9 rev., enunciado corregido).

Rawls sugiere en un par de puntos que su visión es consistente con el entendimiento tradicional del mérito. Afirma, por ejemplo, que no existe conflicto alguno entre su visión de la justicia y la visión aristotélica tradicional. Es verdad que “la definición de Aristóteles claramente presupone... una consideración de lo que pertenece propiamente a una persona y de lo que le es debido”, pero, arguye Rawls, “tales derechos... son a menudo derivados de las instituciones sociales y las expectativas legítimas a las cuales dan pauta”, y dado que “no existe razón alguna para pensar que Aristóteles estuviera en desacuerdo con esto”, la consecuencia es que allí “no hay conflicto con la noción tradicional” (*TJ*, 10-11 / 10 rev.). De manera similar, piensa que puede afirmar “que, en la proposición tradicional, un esquema justo confiere a cada persona su parte; esto es, asigna a cada uno lo que tiene derecho, según lo define el propio esquema” (*TJ*, p. 313 / 275-6 rev.).

Tales formulaciones suprimen las profundas diferencias entre la noción tradicional de mérito y el tipo de posición que tanto Rawls y los utilitaristas parecen apoyar. Es cierto que en sus visiones puede caber el reclamo de que un esquema justo concede a cada persona lo que ella o él merecen. Sin embargo, esto se puede hacer únicamente al otorgar a dicho reclamo un significado muy diferente del que tradicionalmente ha tenido. Según tal versión, el reclamo reposa implícitamente en un parámetro independiente e individualista del mérito, y lo que éste afirma es que las instituciones deben ajustarse a dicho parámetro si pretenden ser justas. Para Rawls y los utilitaristas, en cambio, los principios de justicia distributiva, entendidos de manera holista, se configuran sin referencia a ninguna noción previa del mérito, y lo que el argumento simplemente afirma es que las instituciones justas reconocerán las expectativas a las que dan pauta, expectativas que pueden ser expresadas, si uno desea, usando el lenguaje del mérito.

En síntesis, Rawls coincide con el utilitarismo sobre lo deseable de proporcionar un argumento sistemático de la justicia que reduzca el campo para el balance intuicionista y ofrezca una solución clara y constructiva al problema de la prioridad; sobre la necesidad de subordinar los preceptos de justicia del sentido común a un criterio más alto; y sobre el carácter holista de la justicia distributiva. Tomados conjuntamente, estos tres elementos de su visión significan que, como el utilitarista, está preparado para apelar al principio superior, sin recurrir al balance intuicionista, para proporcionar una justi-

ficación sistemática de los acuerdos interpersonales que pudieran violar las máximas de justicia del sentido común. A la luz de este aspecto de la teoría de Rawls, la tentación por aducir que no les otorga mayor peso que el utilitarismo a las distinciones entre personas es comprensible. Aunque he argumentado que esta tentación debe ser resistida, parece justo decir que los enfoques rawlsiano y utilitarista de la justicia tienen algunos elementos importantes en común, y que tales elementos corren en sentido contrario a una tendencia profundamente arraigada en nuestro pensamiento moral.

Dentro de la filosofía política contemporánea, esta tendencia cobra en el trabajo de Nozick la que quizás sea su expresión más poderosa, y es importante hacer notar que una resistencia respecto al holismo distributivo parece ser parte de lo que se sostiene detrás de su objeción a los “principios fin-resultado”³⁰. Se ha dicho que estos principios evalúan la justicia de una distribución dada (o secuencia de distribuciones) únicamente distinguiendo si la matriz distributiva asociada satisface algún criterio estructural, antes que tomando en cuenta información histórica acerca de cómo la distribución vino a suceder. De manera significativa, Nozick clasifica los principios utilitarista y rawlsiano de la justicia como principios fin-resultado. (Señala en efecto que el diseño de la posición original garantiza que sólo se seleccionen principios fin-resultado). Aplicada a Rawls, esta caracterización no parece correcta, dada la prioridad léxica de su primer principio sobre su segundo, así como su tratamiento de la cuestión de las cuotas distributivas como un asunto de justicia procedural pura. Después de todo, es Rawls quien afirma que “una distribución no puede ser juzgada de manera aislada respecto al sistema del cual es resultado, o del que los individuos han creado de buena fe al amparo de expectativas establecidas”; y quien insiste que “simplemente no hay respuesta” a la pregunta abstracta de si una distribución es mejor que otra. Empero, las posturas rawlsiana y utilitarista son en efecto holistas, y esto puede ser parte de lo que Nozick encuentra objetable en ambas. Porque al menos parte de su queja es que las dos exageran el significado del contexto distributivo general y no confieren importancia suficiente a los aspectos concretos de las transacciones particulares.

Tal queja se vincula con una fuente más general de resistencia al holismo, la cual deriva de una convicción cuyo efecto es validar una tendencia desplorable a que las vidas de los individuos modernos estén subsumidas dentro de estructuras burocráticas masivas y a que sus intereses se hallen subordinados a las demandas de agregados sociales más grandes y al poder bruto de fuerzas impersonales que no pueden controlar. Aceptar una posición holista de la justicia, desde esta perspectiva, es reconocer una erosión en la

³⁰ *Anarchy, State and Utopia*, pp. 153-55.

condición del individuo, lo que es uno de los rasgos más estremecedores de la vida moderna. En efecto, para alguna gente, a eso se debe la alta resonancia la queja de Rawls contra el utilitarismo de no tomar en serio la distinción de las personas. Es irónico, por tanto, que el autor de dicha queja no sólo no se oponga al holismo en relación con la justicia distributiva, sino que sea de hecho uno de sus más firmes defensores. No obstante, una vez que reconocemos que, para cierta gente, las palabras con las cuales Rawls articula su crítica pueden servir como un modo de expresar resistencia al holismo, es comprensible por qué algunos de los que se han hecho eco de dichas palabras no hayan seguido a Rawls en su intento de diseñar una alternativa constructiva y sistemática al utilitarismo. Para ellos, la constructividad, la sistematicidad y el holismo pueden ser todos sintomáticos de un fracaso por otorgar importancia moral suficiente a la distinción de las personas. Sin embargo, si esto es así, entonces sus preocupaciones finales no son las mismas que la suya, aun cuando puedan ser expresadas con las mismas palabras.

Eso no significa que sus preocupaciones sean insignificantes. De manera comprensible, algunas personas aborrecen muchas de las tendencias de la vida moderna que presionan para pensar de forma holista acerca de la justicia, y creen que nuestro pensamiento moral, más que buscar acomodar dichas tendencias, deberían servir como fuente de resistencia a las mismas. Sin embargo, cualesquiera que sean los méritos de su visión, Rawls no la comparte. En esta cuestión, los utilitaristas y él están en el mismo lado. La objeción de Rawls al utilitarismo no es a su holismo, sino más bien al criterio particular que éste usa para evaluar la legitimidad de los acuerdos interpersonales. Su propia teoría de la justicia, uno podría decir, no pretende resistir las presiones hacia el holismo, sino más bien tamizarlas o domesticarlas: proporcionar una forma humana y equitativa a favor de una sociedad liberal y democrática que concilie dichas presiones, bien que preservando sus valores básicos, y que mantenga su compromiso por la inviolabilidad del individuo. Esto es algo que, en su opinión, el utilitarismo nunca hará a pesar de las credenciales liberales de sus más poderosos defensores. Algunas personas pueden pensar que el holismo por sí mismo socava los valores liberales, de modo que el propósito de Rawls resulta en principio inalcanzable. Estas personas concluirán inevitablemente que sus críticas al utilitarismo no irán lo suficientemente lejos, y que su propia teoría exhibe algunas de las mismas fallas que ven en la teoría utilitarista. Pero, de nuevo, no son éstas las mismas fallas que él observa en el utilitarismo, se expresen o no con las mismas palabras.

El hecho de que la actitud de Rawls hacia el utilitarismo esté marcada no sólo por desacuerdos profundos, sino también por zonas importantes de afinidad, puede ayudar a explicar algunas cosas, de otro modo desconcertantes, que menciona acerca de esa doctrina en *Liberalismo Político*. Por supuesto,

en dicho libro, los objetivos de Rawls son distintos a los de *Una Teoría de la Justicia*. Su meta primaria ya no es desarrollar sus dos principios como alternativa al utilitarismo, sino más bien explicar cómo una sociedad liberal justa y estable puede ser establecida y sustentada en circunstancias marcadas por desacuerdos razonables acerca de asuntos morales y filosóficos fundamentales. Dado su interés en esta nueva tarea, el utilitarismo es claramente relegado a la periferia. Una de las pocas ocasiones en las que tiene algo substancial que decir acerca del mismo es cuando incluye el utilitarismo clásico –el utilitarismo de Bentham y Sidgwick, “la doctrina clásica estricta” (*LP*, p. 170)– entre las doctrinas que podrían participar en un consenso convergente en una concepción política liberal de la justicia, el “ejemplo modelo” (*LP*, p. 164) de la justicia como equidad. La posibilidad de un tal consenso reside en el corazón de su respuesta a la pregunta de cómo es posible una sociedad liberal justa y estable en condiciones de pluralismo razonable.

A la vista de ello, sin embargo, la sugerencia de que el utilitarismo clásico podría participar en un consenso de ese tipo resulta sorprendente. Rawls parece estar proponiendo que la supuestamente menos plausible de las dos versiones de la misma teoría, a la que en *Una Teoría de la Justicia*, había tratado como objetivo primario de su crítica, y como rival primario de sus propios principios de la justicia, podría en realidad asociarse en un consenso convergente afirmando dichos principios. Resulta muy difícil ver cómo esto podría funcionar³¹. En primer lugar, los participantes en el consenso que él describe son visualizados como convergentes no sólo en los principios de justicia, sino además en ciertas ideas fundamentales que están implícitas dentro de la cultura política pública, y de las cuales dichos principios son asumidos como su consecuencia. La más importante de tales ideas es la idea de sociedad como un sistema equitativo de cooperación. Aunque Rawls había dicho muy explícitamente en *Una Teoría de la Justicia* que el utilitarismo clásico no acepta esa idea (*TJ*, p. 33 / 29-30 rev.). Además, el argumento que va desde las ideas fundamentales hasta la concepción política está concebido en *Liberalismo Político* como procedente desde la posición original, de la que se dice que modela las ideas relevantes (*LP*, Lección I.4). Sin embargo, sabemos que las partes en la posición original rechazan de manera decisiva el utilitarismo clásico. Por lo tanto, es doblemente incierto cómo podría participar el utilitarismo clásico en el consenso convergente que Rawls imagina, pues éste rechaza las ideas fundamentales que forman la base del consenso, y los argumentos que comienzan a partir de aquellas ideas se dicen destinados a volverse en contra suya.

³¹ Aquí retomo mi discusión presentada en “The Appeal of Political Liberalism”, *Ethics*, 105 (1994), pp. 4-22.

Desde luego, Rawls no sostiene que la concepción política sea deducible del utilitarismo clásico, sino únicamente que la visión clásica podría apoyar la concepción política “como una aproximación satisfactoria y quizás la más plausible [a lo que el principio de utilidad requeriría en balance] en condiciones sociales normales” (*LP*, p. 171). Sin embargo, como se advirtió más arriba, Rawls afirma explícitamente que un consenso convergente es suficientemente “profundo” como para incluir tales ideas fundamentales, como la idea de sociedad como sistema equitativo de cooperación (*LP*, pp. 149, 158-60, 164-6), y la sugerencia de que el utilitarismo clásico podría apoyar la concepción política como una “aproximación plausible” no explica cuál es la actitud que el utilitarista deba tener ahora respecto de dicha idea³². Además, dicha sugerencia es ambigua en cuanto, entre dos actitudes considerablemente diferentes, el utilitarista podría decantarse por principios de justicia en apariencia no utilitaristas. La primera significa “que bajo las condiciones de la sociedad civilizada existe gran utilidad social en seguir las para la mayor parte y en permitir infracciones sólo en circunstancias excepcionales” (*TJ*, p. 28 / 25 rev.). Cuando Rawls discutió dicha actitud en *Una Teoría de la Justicia*, argumentaba que ello significaba un compromiso débil e inadmisible para la prioridad de tales principios. La segunda actitud significa que la utilidad se maximizará en realidad si los principios no utilitaristas son “públicamente afirmados y realizados como la base de la estructura social” (*TJ*, p. 181 / 158 rev.). Cuando Rawls discutió dicha actitud en la *Teoría*, argumentó que era sorprendente abandonar por completo el utilitarismo, ya que, dada la condición de publicidad, el utilitarismo debe ser “definido” como “la doctrina de que el principio de utilidad es el principio correcto para la concepción pública de justicia de la sociedad” (*TJ*, p. 182 / 158 rev.). Por tanto, según los propios parámetros de Rawls, ninguna de esas actitudes parecería calificar a una persona *sea* como utilitarista genuino, *sea* como participante comprometido en un consenso convergente dentro de una concepción político-liberal de la justicia.

Si esto es correcto, entonces resulta difícil ver cómo podría ser incluido el utilitarismo clásico en dicho consenso. Aunque la disposición de Rawls es tratarlo como un candidato para la inclusión, lo cual inicialmente es deseable, podría parecer más comprensible si uno conserva en la mente la complejidad

³² Este es el defecto en la respuesta de Brian Barry a mi discusión previa (en “The Appeal of Political Liberalism”) de la participación utilitarista en un consenso convergente. En “John Rawls and the Search of Stability”, *Ethics* 105 (1995): pp. 874-915, Barry argumenta (en particular, pp. 914-5, nota) que, debido a que Rawls no mantiene que la concepción política debe ser deductivamente derivable desde el interior de cada doctrina general representada en un consenso convergente, no debe pensarse que exija que los participantes en tal consenso apoyen algo más que los propios principios de la justicia. Sin embargo, esto sobrepasa las repetidas menciones de Rawls acerca de la profundidad de un consenso convergente.

de su actitud hacia el utilitarismo en la *Teoría*. He argumentado a lo largo de este ensayo que su indudable oposición al utilitarismo, y su determinación a proporcionar una alternativa al mismo, no deberían permitir que se obscurezcan algunos puntos importantes de acuerdo. Quizás uno podría decir que es precisamente porque concuerda en mucho con el utilitarismo, por lo que Rawls está determinado a proporcionar una alternativa que lo supere en los aspectos en que éste es deficiente. Si bien no habría necesidad de proporcionar una teoría mejor si el utilitarismo no tuviera serias fallas, el esfuerzo difícilmente habría valido la pena si éste no tuviera también virtudes importantes. El utilitarismo, en la perspectiva de Rawls, ha sido la teoría moral sistemática dominante dentro de la tradición liberal moderna. Si él mismo no acordara que necesitamos una teoría clara y sistemática para reducir nuestra confianza en una intuición sin guía, y proporcionar una base adecuada para las instituciones liberales y democráticas, tampoco le preocuparía tanto enfatizar las deficiencias del utilitarismo u por producir una teoría que remedie tales deficiencias, aunque preservando las virtudes de la misma.

Hasta cierto punto, por tanto, Rawls y el utilitarista se hallan comprometidos en una empresa común, y es en contra del fondo que tienen en común que Rawls toma al utilitarismo como principal objetivo de su crítica en la *Teoría*. En *Liberalismo Político*, el contexto de la discusión se ha modificado. El objetivo ahora es mostrar cómo las instituciones liberales pueden lograr la estabilidad en condiciones de pluralismo recurriendo a diversas fuentes de apoyo moral. En este contexto, el utilitarismo, con su lugar prominente en las tradiciones del pensamiento liberal y sus diversas afinidades más específicas con la propia visión de Rawls, se presenta a sí mismo como un aliado natural. Por supuesto, queda todavía un problema, dada su insistencia en la *Teoría* de que ni el utilitarismo clásico o el promedio pueden situar los valores liberales fundamentales sobre una línea de base suficientemente segura. Y el problema se torna más agudo, por las razones señaladas más arriba, cuando se concibe el consenso convergente como aquél que no sólo afirma principios liberales en general, sino la teoría de la justicia de Rawls en particular. Sin embargo, el impulso para tratar alguna forma de utilitarismo como un candidato para la inclusión dentro del consenso, cuando se le considera en el contexto de los objetivos de Rawls en *Liberalismo Político* y se tiene en cuenta su simpatía por ciertos aspectos de la doctrina utilitaria, deja de parecer misterioso³³. Sea o no que las tensiones entre dicho impulso y sus obje-

³³ Es importante advertir que, en su ensayo previo “The Idea of an Overlapping Consensus” [Oxford Journal of Legal Studies 7 (1987): pp. 1-26; en particular, p. 12; También en Rawls: *Collected Papers*, pp. 421-48; en particular, pp.433-4], Rawls mismo había rechazado que el utilitarismo de Bentham y Sidgwick pudiera pertenecer a un consenso convergente, pero había sugerido que quizás alguna forma de “utilitarismo indirecto” podría hacerlo. La última sugerencia

ciones obligadas al utilitarismo puedan ser satisfactoriamente resueltas, proporcionan un saludable recordatorio de la complejidad de la actitud de Rawls respecto de la “teoría sistemática predominante” de la filosofía moral moderna. Al mismo tiempo, supone una evidencia del logro de Rawls el que la condición predominante del utilitarismo haya estado abierta a serios cuestionamientos desde que *Una Teoría de la Justicia* enunciase su poderosa visión alternativa³⁴.

cia parece más plausible, aun cuando no sea fácil de reconciliar por entero con los comentarios citados de *Teoría*.

³⁴ Versiones preliminares de este trabajo fueron presentadas en un coloquio del Departamento de Filosofía en la UCLA, en un simposio en Estocolmo patrocinado por la Real Academia de las Ciencias y la Fundación Rolf Schock, y en la Conferencia 2000 de la Sociedad Internacional de Estudios Utilitaristas, llevada a cabo en la Universidad Wake Forest. Agradezco a las respectivas audiencias las provechosas discusiones. Estoy igualmente en deuda con Samuel Freeman por sus comentarios a una versión anterior, y con Nicholas Kolodny por su valioso apoyo en la investigación.