

Paradojas de la filiación kantiana de Rawls ¹

VICENTE SERRANO MARÍN

Universidad Europea de Madrid

RESUMEN. Innegable en algunos aspectos y en la voluntad declarada de Rawls, la cuestión nunca resuelta de la filiación kantiana de su obra no está exenta de conflictos, o mejor de situaciones paradójicas que, a nuestro entender, iluminan el sentido mismo de su obra. La raíz de esas paradojas parece descansar en que la tradición a la que obedece es la anglosajona y la ética implícita de la que parte es la de un utilitarismo que nunca abandona del todo la expresión más externa y aparente de ese juego de paradojas se evidenciaría en la propia evolución de la teoría, desde su inicial afirmación de kantismo moral, hacia una formulación política y no metafísica, siendo, sin embargo, precisamente en ese momento político donde estaría más próxima al kantismo moral. Tal vez con ello Rawls nos está ofreciendo una clave para otro problema nunca resuelto de la modernidad: el de las relaciones entre ética y política.

ABSTRACT. While it is undeniably so in some aspects and in Rawls' stated intention as well, the never quite resolved question of the Kantian leanings of his works is not free of controversy or rather of paradoxical situations which, to our understanding, illustrate the very sense of his work. The root of these paradoxes seems to lie with the fact that while his work follows the Anglo-Saxon tradition, the implicit ethical system which it forms part of is that of a type of utilitarianism which it never gives up completely. The most visible external expression of these paradoxical situations can be found in the very evolution of the theory, moving from its initial affirmation of Kantian morals toward a political, not metaphysical formulation; nevertheless, it is precisely at this political moment when it is closest to Kantian morals. Perhaps with this, Rawls is offering us the key to another unresolved question of our times: that of the relationship between ethics and politics.

Ya no tiene por qué parecer desesperadamente hostil un mundo en el que deban prevalecer inevitablemente la voluntad de dominio y las crueldades opresivas, alimentadas por el prejuicio y la locura. Situados como acaso lo estemos en una sociedad corrupta, puede que no haya consuelo para nuestra pérdida; pero podemos pensar que el mundo no es en sí mismo inhóspito para la justicia política y el bien que ella encierra. Nuestro mundo social podía haber sido diferente y hay esperanzas para los que vivan en otro tiempo y otro lugar.

El sentido general de lo que sigue, y que trato de recoger en el título mismo, procede de mi primera aproximación a la obra de Rawls hace ahora ya más de una década. En aquel primer encuentro con Rawls y su *Teoría de la Justicia* lo que me llamó la atención fueron principalmente dos cosas: el título mismo de su obra y la pretensión de Rawls, que por otra parte parecía generalmente asumida, de considerar esa obra como heredera de una ética kantiana. En cuanto al título me pareció ambicioso en exceso porque, procediendo mi primera formación del mundo del Derecho, el concepto de la justicia, más allá del planteamiento del positivismo jurídico, me había resultado siempre enigmático. En realidad en la Facultad de Derecho descubrí que mi propia tendencia a interesarme por ese problema suponía una dificultad para proseguir mi carrera como profesional en ese ámbito. Mi tendencia en esa dirección se expresaba, por ejemplo, en que ante un examen de Derecho procesal yo dejaba los manuales, los apuntes y códigos para tomar notas de la *KrV*, y justamente por ello acabé por estudiar filosofía, un interés que en gran medida y visto retrospectivamente se vincula a la noción misma de justicia. Lo segundo a lo que me refería, es decir, la pretensión de Rawls de haber construido una teoría kantiana, me llamó profundamente la atención y me resultó cuando menos sorprendente antes incluso de haber profundizado en la obra de Rawls. Y ello por dos razones. La primera porque su obra parecía surgir del contexto de las teorías de la elección

racional, lo que la aproximaba inmediatamente a lo que Kant hubiera llamado acciones técnico-prácticas. La segunda porque en mis estudios de la moral kantiana, prolongados luego en el sesgo ético dado por Fichte al primer sistema idealista, la cuestión de la justicia tal como aparecía diseñada por Rawls no parecía coincidir con una concepción moral en términos de Kant, y sí más bien con un problema de filosofía política, sin duda relacionado con el anterior, pero de índole metodológica y sistemática diferente. El problema, entonces, no era tanto que la filosofía política de Rawls no pudiera ser kantiana en términos de contractualismo, sino que más bien se confundían y mezclaban dos órdenes distintos, sin duda relacionados, pero distintos. De hecho, la distinción kantiana entre lo moral y lo jurídico se estudia en la Facultades de Derecho como una de las más ajustadas y precisas. De la combinación de estas cuestiones resultaba entonces que la *Teoría de la Justicia* parecía ser una fundamentación del orden jurídico-político a partir de bases morales, mas no exactamente una moral; pero además era una fundamentación cuyo momento moral, y no tanto el político, se antojaba bastante alejado de las pretensiones kantianas, en la medida en que caía más bien en lo que la tradición kantiana consideraba *eudaimonismo*.

Éstos eran los presupuestos desde los que abordé un estudio sistemático de la obra de Rawls a finales de los ochenta. Mi análisis se limitó entonces casi exclusivamente a la *Teoría de la Justicia* a partir de

la edición castellana de FCE, y debo decir que mis conclusiones no se apartaron demasiado del punto de partida, es decir, que mis presupuestos iniciales se vieron corroborados en gran medida. Lo interesante a este respecto no son tanto mis tribulaciones con respecto a Rawls sino el hecho de que por esas mismas fechas, y sin que yo mismo lo supiera (por esas fechas yo eran un estudiante de la licenciatura, lo que no justifica mi ignorancia pero sí la explica), Rawls estaba explicando un desarrollo que en parte confirmaba mis hipótesis. En el año 85 había publicado Rawls un artículo titulado «Justicia como equidad. Política, no metafísica»². La primera parte del título coincidía con el artículo original del año 58 que constituía el auténtico germen de *Teoría de la Justicia*³. Pero la segunda parte del título trataba de poner de manifiesto, casi treinta años después, precisamente el hecho de que su teoría pertenecía más bien a la teoría política que a la filosofía moral propiamente dicha. Con ello Rawls se hacía cargo en realidad de algunas críticas que había recibido su obra hasta entonces y trataba a la vez de «algunas de las ambigüedades y oscuridades» de la teoría. En definitiva, al explicitar su pretensión de situar la obra en el ámbito de la filosofía política y así evitar las críticas hechas desde un lectura ética (en aquellos años era un tópico que *TJ* suponía una revolución en la ética anglosajona en la medida en que ofrecía finalmente una ética sustantiva), en primera instancia, y filosófica, en última, de su teoría, lo que hacía era asumir una diferencia fundamental con respecto a la moral kantiana de la que se reclamaba heredera: su enfoque, frente al de Kant, trataba de apartarse de consideraciones metafísicas. De hecho en la segunda nota del trabajo y explicando esto mismo Rawls se remite a las llamadas *Dewey lectures* y en concreto a la 3.^a: *Constructivismo kantiano en teoría moral*, de 1980⁴, para señalar que el término constructivis-

mo trata de señalar analogía pero no identidad con Kant, lo que significa literalmente, según Rawls, que su teoría no obedece «a problemas epistemológicos ni metafísicos». De hecho en su artículo del 75 titulado «Una concepción kantiana de la equidad» ya había estilizado considerablemente su filiación kantiana, en cuanto que Kant habría dado el tono al sentido último de las pretensiones de *TJ*, pero esta última estaba lejos de ser sin más una moral como la kantiana. En efecto, allí se hace cargo de que «el abandono de los dualismos (kantianos) tal como él (Kant) los piensa, significa para muchos el abandono de lo distintivo de su teoría»⁵. En realidad en *TJ* misma ya había formulado que la teoría kantiana se separaba de sus bases metafísicas.

Este breve recorrido que nosotros hemos hecho a la inversa desde el 85 al 71 muestra, por tanto, que, en efecto, hay en Rawls una evolución que estiliza cada vez más el carácter kantiano de su obra, entendido este carácter como la moral kantiana sin más. En lugar de ello Rawls parece reconocer en esa evolución que su filiación kantiana lo es sólo en el sentido de la organización política de la sociedad. Primero abandona los elementos metafísicos, después los epistemológicos, finalmente los éticos, para quedarse sólo con la idea de constructivismo que debería remitir a lo determinante de su kantismo, que es el predominio de lo correcto sobre lo bueno, en lo que residiría el sentido general de la moral kantiana que, aplicado a la concepción política y no metafísica de la sociedad, daría el tono a la propia concepción de Rawls.

Planteadas así las cosas el kantismo de Rawls encontraba una segunda piedra de toque precisamente acudiendo a la segunda de las dificultades que yo había encontrado en su obra, a saber, en el momento moral que lo fundamentara, que en su caso se sustanciaba en la pretensión de ser una alternativa a los utilitarismos, precisamen-

te porque en *TJ* se daría prioridad a lo justo sobre lo correcto, o para simplificar, precisamente porque *TJ* era una teoría deontológica. En efecto, en *TJ* se afirma: «El último contraste que mencionaré ahora es que el utilitarismo es una teoría teleológica, mientras que la justicia como imparcialidad no lo es. Entonces, por definición esta última es una teoría deontológica»⁶.

Esta expresión a partir de la definición de las teorías como teleológicas y deontológicas es una constante que no abandona ya Rawls, ni siquiera el último de *Justicia como equidad. Una reformulación*⁷. En ella se centra, pues, el elemento decisivo a la hora de determinar las relaciones de Rawls y Kant, más allá del constructivismo, sobre el que volveremos. Las alusiones explícitas a este respecto son múltiples e ineludibles. A su vez esta cuestión aparece directamente relacionada con el hecho de que Rawls considere explícitamente su *Teoría de la Justicia* como una respuesta y una alternativa a las teorías utilitaristas. Más allá de la cita mencionada, su interés a este respecto parece no haberle abandonado desde la primera versión de *Justicia como equidad* hasta la última, pocos meses antes de su muerte. En el capítulo 6 del artículo del 58 Rawls consideraba ya necesario aclarar su concepción de la justicia por contraste y comparación con las tesis de Bentham y Sidgwick. Allí nos dice que lo propio de ambos es considerar la justicia desde el punto de vista de la eficiencia, a partir de la función de utilidad, y compara la justicia así entendida con una decisión administrativa de nivel superior que no tiene en cuenta «las relaciones morales en que se encuentran los individuos». Para ello acude al ejemplo de la esclavitud, institución que desde el punto de vista del utilitarismo podría ser considerada injusta mediante la determinación de funciones de utilidad del esclavo y del dueño, mientras que para la justicia como imparcialidad siempre sería

injusta: «El utilitarismo no puede dar cuenta de que la esclavitud es siempre injusta.» Para dar cuenta de esto, añade algunas páginas más abajo, «habría que modificar su inspiración completa como concepción moral»⁸. Por no alargar demasiado el trabajo prescindiré ahora de discusiones análogas en *TJ* y en escritos posteriores, y acudiré al último texto que poseemos del propio Rawls, donde la discusión contra el utilitarismo sigue tan viva como en el año 58. Allí afirma:

El objetivo de la justicia como equidad es encontrar una concepción de la justicia política alternativa a la del utilitarismo[...], y al mismo tiempo encontrar una base moral más apropiada para las instituciones de una sociedad democrática moderna⁹.

Parece, pues, fuera de toda duda que la intención consciente, declarada y más constante de Rawls es precisamente la de elaborar una alternativa a la teoría utilitarista, y que el núcleo de esa alternativa está en la noción de justicia como equidad más allá del beneficio y la eficiencia. ¿Por qué entonces despertó en su día la sospecha de que pudiera ser utilitarista? Y más allá de esto: ¿por qué efectivamente se le ha criticado por ese flanco afirmando el carácter utilitarista de su teoría? Porque, en efecto, ya en los años setenta se alzaron voces a este respecto, críticas que, por lo demás, influyeron decisivamente en su evolución a partir de *TJ*. Más allá de la conclusión a la que podamos llegar y a la que llegaremos en este punto, lo cierto es que las dudas a este respecto, o incluso la sospecha, no son del todo infundadas, y obedecen a aspectos decisivos de su propia obra. Es sabido que el utilitarismo como doctrina ética ha conocido una evolución considerablemente compleja, quizá no comparable a ninguna otra filosofía moral. En el seno de la misma no sólo se ha discutido la determinación misma de lo útil, sino incluso también la articulación de la utilidad con respecto a lo correcto, y

siempre para evitar dificultades análogas a las que plantea Rawls. De hecho su caracterización como doctrina predominantemente teleológica queda matizada en formulaciones como la del utilitarismo de la regla, en la que el momento deontológico parece ocupar un espacio casi análogo al propuesto por Rawls. Lo que con ello queremos expresar es que, al menos *prima facie*, en el seno de una concepción utilitarista podría haber una discusión acerca de la justicia como la que Rawls plantea, sin que esto signifique sin más abandonar el utilitarismo como doctrina moral general. Y en este sentido no sería descabellado considerarle utilitarista, siempre y cuando se aclarara en qué sentido lo es. Pero este argumento que acabamos de desarrollar se limita a levantar una restricción, o si se prefiere señala que la mera contraposición entre teoría deontológica y teleológica parece insuficiente sin más para descartar la posibilidad de un cierto utilitarismo en Rawls en el que se pusiera un especial énfasis en el aspecto deontológico, y siempre desde el punto de vista de la justicia. Por sí solo el argumento en nada avalaría el hecho, o sólo la sospecha de que Rawls lo fuera. Para ello, como es obvio, deben aportarse elementos positivos.

La sola posibilidad de considerar esos argumentos positivos pasa por explicar a su vez por qué el utilitarismo, a diferencia de otras filosofías morales, ha conocido tantas formulaciones y qué es lo que hace que todas ellas puedan considerarse utilitaristas. Ello exige dar una definición mínima de utilitarismo que deban compartir todas las formulaciones, y en la que se pudiera además incluir al proyecto rawlsiano de la teoría de la justicia. El utilitarismo como doctrina moral tiene unos contornos bien reconocibles, así como un origen histórico que ofrece pocas dudas. En su fórmula más concisa obedece al principio de la máxima utilidad para el mayor número posible. Pero esta doctrina como bien puede apreciarse por su formu-

lación misma es una doctrina moral que parece asumir como su problema característico una dimensión social, que expresada en otros términos podríamos resumir como el problema clásico de agregar preferencias individuales. Ahora bien, a su vez el utilitarismo contiene en su seno una concepción moral más general que parece descansar en dos principios fundamentales. El primero sería el de la utilidad como criterio material del bien, con arreglo al cual este último se mide en términos de utilidad para los individuos, y que en último extremo lleva a la maximización de la utilidad como máxima expresión del bien. El segundo principio, en realidad implícito en éste, descansaría en la desaparición de una noción de bien objetiva, a partir de bases metafísicas. Una concepción como ésta en todo caso no sería sino la base de otras doctrinas morales, cuya matriz moderna podríamos encontrar en Hume, que iniciaría así una tradición a la que el utilitarismo se adscribe. Obviamente Kant no pudo polemizar con el utilitarismo, pero es esa matriz común a distintas doctrinas la que tiene presente cuando habla de un tipo de morales heterónomas que se basan en el propio interés o en el amor propio o en la tendencia a la felicidad. Cuando construye conscientemente una metafísica de las costumbres lo hace para encontrar un fundamento distinto a esa doctrina moral general, una de cuyas expresiones será años más tarde el utilitarismo. El *factum* moral de la modernidad es el mismo para Kant y para esa tradición *eudeaimonista*, a saber, la imposibilidad de hallar un concepto de bien compartido. Pero la respuesta kantiana ante ese *factum* es precisamente encontrar una posibilidad de fundamentar una concepción moral de validez universal al margen de la determinación empírica del individuo, al margen de la inclinación (*Neigung*) como elemento propio de la condición sensible. Por el contrario, la matriz de la que hablamos, lo que podríamos considerar también la

matriz del utilitarismo, considera que la ética debe prescindir de cualquier construcción metafísica, aunque implícitamente no deja de asumir al menos una determinada concepción de corte empirista: la que considera al individuo como un sujeto que actúa movido por la utilidad y tiende a maximizarla (o el bien como quiera que se determine, o la felicidad), en definitiva la de una racionalidad instrumental, por hablar en términos de la tradición europea continental, y esa misma es la que está en la base del liberalismo económico y de las teorías de la elección racional. Esa tradición es justamente la tradición de la que surge y en la que se enmarca la concepción del Rawls que inicia su trayectoria hacia la teoría de la justicia. Sería injusto identificarla sin más con el utilitarismo, pero lo cierto es que la distinción fundamental entre teorías deontológicas y teleológicas de la que parte Rawls para establecer su divergencia entre su teoría y la teoría utilitarista nos permite decir que en ese punto fundamental Rawls está más próximo del utilitarismo que de Kant.

En nuestra lectura de Rawls éste compartiría esa marco general desde el que afrontar los problemas éticos, y se apartaría en cambio del utilitarismo en la incapacidad que éste tiene para, desde ese marco, resolver determinadas inconsistencias a la hora de agregar preferencias. O dicho de otra forma, desde el punto de vista moral estricto Rawls podría ser utilitarista en la medida en que pertenece a esa tradición moral que Kant llamaría *eudaimonismo*, pero no en cambio desde el punto de vista político. En realidad ésta es la tradición intelectual a la que Rawls pertenece y lo que hace interesante su obra es su esfuerzo por estilizar esa posición inicial hacia una concepción deontológica. No es en este sentido casual que en el artículo del 58, y después de enunciar por primera vez su sentido de la justicia, pueda considerar a la vez como antecedentes del mismo tanto a Kant como a Mill¹⁰. Por lo demás, en su

discusión con el utilitarismo, al final de ese mismo artículo, Rawls se traiciona al hablar de su principio como una modificación del principio utilitarista: «la modificación general al principio utilitarista[...], aun cuando parezca de poca monta a primera vista, en realidad tiene como respaldo una concepción diferente de la justicia»¹¹. Pero una modificación es algo distinto a una alternativa, que es el término utilizado por Rawls en su última versión.

Así las cosas, y por simplificar lo dicho hasta ahora, cabría establecer dos momentos del utilitarismo y señalar a partir de ahí la relación de Rawls con respecto a cada uno de ellos. El primero sería el momento teleológico y *eudaimonista*, con arreglo al cual la noción básica está en la consideración del principio de eficiencia aplicado a individuos, y que se correspondería con una concepción moral básica, y el segundo sería el del utilitarismo en sentido estricto expresado en el principio de eficiencia aplicado a la sociedad o sociedades por ampliación del primero y que obedece al problema moderno clásico de agregar preferencias individuales, y que se correspondería con la teoría política del utilitarismo. Nuestra hipótesis es que el Rawls del 71 coincidiría con el utilitarismo en el primer sentido, y se alejaría del segundo, es decir, precisamente en el orden político. Sería entonces en este punto donde se produciría a su vez el momento kantiano de Rawls. En lo que sigue vamos a tratar de contrastar nuestra hipótesis con los textos de Rawls, y de ahí obtendremos una primera valoración de las relaciones entre Kant y Rawls¹².

Nuestra estrategia será acudir a las dos nociones básicas en torno a las cuales giran las teorías éticas: a saber, a las nociones de bien y deber, o de lo bueno y lo correcto. Obviamente el rasgo distintivo del libro está en la prioridad dada a lo correcto sobre lo bueno y es eso lo que, como sabemos, permite adscribirle a la

tradición kantiana. En efecto, en términos muy esquemáticos la idea fundamental es que en *TJ* los dos principios de justicia se imponen por encima de cualquier concepto de bien, y en ellos lo correcto es, por tanto, prioritario sobre lo bueno. En el contexto de esa prioridad de lo correcto surge otro de los elementos característicos de *Teoría de la Justicia* que es el de la posición original. Ésta, a su vez, constituye, como se sabe, el momento contractualista de la *Teoría de la Justicia*, lo que determinaría a su vez de nuevo un fuerte contraste con el utilitarismo, que, como es sabido, es contrario a cualquier teoría del contrato desde Hume. Es de la posición original en cuanto hipótesis ideal de donde surgen aquellos dos principios de justicia que tienen prioridad sobre el concepto de bien. Estamos, pues, ante una de las piezas claves de *TJ*, en la medida que en ella parecen contenerse tanto el momento deontológico como el contractualista de Rawls, que son precisamente los dos aspectos que más le alejan del utilitarismo. A su vez si nos adentramos en esa posición original encontramos otras dos piezas decisivas que parecen alejarnos del utilitarismo: la noción de persona moral, que serían precisamente los protagonistas de la posición original, y la del velo de ignorancia. La primera ha sido esgrimida reiteradamente por Rawls como prueba de su kantismo¹³, y no cabe duda de que esa noción tiene una importancia decisiva en la concepción kantiana de la moral. En cuanto al velo de ignorancia también parece confirmar en primera instancia la proximidad a Kant, en la medida en que, como se sabe, el velo de ignorancia es un artificio que impide a las personas morales de la posición original conocer sus circunstancias, lo que, en efecto, se acerca mucho al Kant que exige del agente moral abstraer de sus condiciones empíricas, y decidir u obrar a partir de la razón pura práctica. En una aproximación preliminar, pues, nada que reprochar o que objetar al kantis-

mo de Rawls. Pero las cosas cambian si de nuevo penetramos en cada una de esas nociones y analizamos su respectiva función sistemática en el conjunto de la teoría y su juego con respecto a las demás piezas. Como el mismo Rawls nos dice, su teoría es deontológica por la prioridad dada a lo correcto sobre lo bueno, es decir, por los dos principios surgidos de la posición original. Pero ¿cuál es el concepto de bien que maneja Rawls y en qué sentido se somete al de lo correcto?

En *TJ* Rawls ofreció una doble teoría del bien: lo que llamaba la teoría amplia o gruesa, y la que llamó la teoría fina o en sentido estricto. La distinción no es en absoluto baladí, y antes bien, el hecho de que se vea obligado a establecer dos teorías del bien (o una doble teoría) constituye, como veremos, una de las claves de la relación con el utilitarismo, en la medida en que cada una de ellas responde a su vez a los dos momentos del utilitarismo que hemos establecido más arriba. La llamada teoría amplia del bien hace referencia precisamente a eso bueno que se somete a lo correcto, en definitiva a los fines particulares que cada cual puede perseguir a partir de los dos principios de justicia. A ella está dedicada la parte final de teoría de la justicia. En sí misma es enormemente interesante, y aunque puede ser divergente en muchos sentidos del kantismo (hay en ella un cierto tono aristotélico), lo cierto es que sería compatible con el Kant que asume como presupuesto la condición empírica de los sujetos morales y su búsqueda individual de fines propios siempre y cuando las máximas se generalicen correctamente en términos análogos a como se han generalizado en la posición original de Rawls.

Más delicada en cambio, al menos a nuestros efectos, es la teoría fina o estricta del bien, porque no se somete a los principios de justicia, sino que es más bien la que los hace posibles. Y esto formalmente hablando nos sitúa ya en dificultades para

una doctrina que se pretende kantiana. En efecto, para Kant lo bueno (*Gut*) se obtiene siempre a partir del deber, lo que se expresa en la conocida afirmación de que lo único santo es la voluntad santa. Y la cuestión no es secundaria porque si la moral de Kant es autónoma y formal lo es precisamente porque se prescinde de cualquier determinación empírica, lo que la lleva a prescindir de fundamentarse en el bien para hacerlo desde el deber y sólo desde ahí encontrar el concepto del bien. Estamos en la piedra de toque del kantismo. Pues bien, a este respecto el Rawls de *TJ* no puede ser más explícito: «la teoría de la justicia es una parte, quizá la más significativa de la teoría de la elección racional»¹⁴. Y desde esa caracterización inicial añade: «es conveniente considerar el sentido de bueno como análogo al signo de una función»¹⁵, y en otro lugar: «la definición del bien es puramente formal»¹⁶. Lo que todas estas afirmaciones contienen no es otra cosa que la idea subyacente básica al *eudaimonismo* y apunta al elemento que Rawls comparte con el utilitarismo, el de que las acciones éticas están guiadas por un criterio maximizador. Ciertamente esto no es utilitarismo para Rawls, que inserto en su tradición considera que el utilitarismo se identifica sin más con el utilitarismo en sentido estricto, razón por la cual puede definirlo como aquella doctrina ética en la que «el principio de elección para una asociación de hombres es interpretado como una extensión del principio de elección de un solo hombre»¹⁷. Sin embargo, el propio Rawls parece asumir en otro lugar, en sus *Lecciones sobre la historia de la filosofía moral*, otra concepción del utilitarismo que le acerca a esa tradición que es común a su teoría estricta del bien. En efecto, allí nos dice: «por otro lado el utilitarismo clásico arranca de una concepción del bien, como el placer, o la felicidad o las satisfacción del deseo, las preferencias o los intereses. La cuestión es que una doctrina teleológica se propone una concepción del

bien que es anterior a, e independiente de lo justo o de la ley moral[...] así, por ejemplo, el utilitarismo define lo justo como la maximización del bien»¹⁸. Aquí encontramos las dos determinaciones fundamentales del utilitarismo en dos momentos. La primera se vincularía a la definición del bien, mientras que la segunda se determinaría a partir de la determinación de lo justo y es aquí donde Rawls se aparta del utilitarismo, lo que ya sabíamos. Pero ¿cómo y por qué se aparta, por qué obtiene, como decíamos, un criterio de justicia, su dos principios, que se imponen después a las particulares concepciones del bien propias de la teoría del bien en sentido grueso? El problema está en cómo obtiene Rawls sus principios de justicia, si éstos son en verdad anteriores o posteriores a su teoría del bien fina o estrecha. Y lo cierto es más bien que esos dos principios de justicia son derivados de la misma y no al revés. Es éste un punto central de la interpretación de Rawls y es aquí donde encontramos la primera dificultad, y que a nuestro entender se sustancia en la siguiente afirmación:

Hay que notar que las teorías deontológicas se definen como no teleológicas, no como teorías que caracterizan la corrección de las instituciones independientemente de sus consecuencias. Todas las doctrinas éticas dignas de atención toman en cuenta las consecuencias al juzgar la corrección. Si alguna no lo hace sería irracional¹⁹.

¿Qué concepto de racionalidad maneja aquí Rawls? Obviamente no se refiere a la razón pura práctica kantiana, y sí más bien a la racionalidad técnico-práctica, a lo que luego se dio en llamar racionalidad instrumental, es decir, a la teoría de la elección racional. ¿Debemos considerar que la concepción kantiana es irracional para Rawls? Habría un sentido en el que efectivamente sería irracional, aquel que remite a las bases metafísicas de la teoría kantiana, de la que, como sabemos, pretende Rawls separar los aspectos estricta-

mente morales. Porque en efecto, si separamos de esas bases metafísicas, tal vez lo que nos interesa de Kant es el resultado, a saber, la posibilidad de alcanzar un concepto del deber que se anteponga a las distintas concepciones del bien que obedecen a determinaciones empíricas. Y ciertamente es en este resultado en lo que coinciden ambos. Pero con ello Rawls está haciendo suyos los dos rasgos decisivos del utilitarismo definido en sentido amplio: prescinde de cualquier metafísica, y sobre todo llega a ese concepto de justicia a partir de la noción de bien propia del utilitarismo, es decir, a partir de la racionalidad técnica. Y aquí encontramos ya claramente expresada una primera paradoja que nos viene acompañando de manera latente, la que se resume en el hecho de que partiendo de una concepción fundamentalmente teleológica llega a una concepción deontológica y análoga a la kantiana. Pero para mostrar que esto es así es necesario volver atrás y considerar aquellas piezas ya señaladas que harían de su concepción una concepción deontológica a partir de bases teleológicas: el velo de ignorancia y la posición original.

A este respecto debemos hacer las siguientes observaciones:

1) Estamos ante una situación contractual.

2) Los sujetos de la situación contractual tienen la condición de personas, como nos ha dicho Rawls, pero eso se concreta en una doble dimensión: son racionales en términos consecuencialistas propios de la teoría de la elección racional y se dirigen a maximizar.

3) El objetivo que esas personas persiguen en la situación contractual es ordenar una sociedad en términos de justicia.

4) Como es obvio no hay una noción previa acerca de lo que es una situación de justicia, porque en ese caso sobraría el entramado que estamos considerando.

5) Como corolario de la observación anterior, entonces los principios de justicia se obtienen a partir de la interacción de esos sujetos entendidos como seres racionales definidos en 2, pero a su vez ese carácter racional es la estructura de la que surge la teoría estrecha del bien, que, por tanto, es la que debe dar razón de los principios de justicia.

La noción misma de lo justo habrá de resultar de los elementos que estamos considerando, pero entonces aparece como un problema más propio de la teoría política que de la teoría moral, observación que hicimos al principio, y por lo demás como problema político, en la medida en que se sitúa en el ámbito de lo que se llamó el preferidor racional, es un típico problema de agregación de preferencias que es, a su vez, el rasgo característico del utilitarismo político. Situados en este contexto resulta que la discrepancia fundamental con el utilitarismo no lo es sobre «las bases metafísicas del mismo», sino sobre un problema técnico como es el de agregar preferencias. Rawls es explícito a este respecto: «se sabe que algunos métodos de combinar preferencias individuales en preferencias sociales da lugar a inconsistencias»²⁰. Y en este sentido la distancia que Rawls toma desde el principio de la obra del utilitarismo reside precisamente en su incapacidad para resolver esas inconsistencias, las cuales han alcanzado históricamente su punto crítico en el llamado teorema de la imposibilidad de Arrow, según el cual la ordenación agregada de las preferencias individuales no tiene las mismas condiciones de racionalidad que las preferencias individuales mismas. El propio Rawls, haciéndose cargo de este problema, definía el utilitarismo, como vimos, como aquella doctrina ética en la que «el principio de elección para una asociación de hombres es interpretado como una extensión del principio de elección de un solo hombre»²¹.

Uno de los primeros y más penetrantes críticos de Rawls ya señaló en su momento que el origen mismo de la teoría de la justicia tiene su origen y su clave en esta discrepancia, atreviéndose a afirmar incluso que la intención inicial de Rawls no era otra que la de escribir un libro análogo a *Social choice and individual values*, y en el que fuera posible presentar una solución al problema, lo que estaría ya en el punto de partida de «Justice as Fairness», que como se sabe contiene el primer esbozo de la *Teoría de la Justicia* ²². Y en verdad hay muchos elementos que avalan una interpretación semejante.

Sea correcta o no esa interpretación general de la intención de la obra de Rawls, lo que sí parece fuera de duda, puesto que es el propio Rawls quien lo afirma, es que el edificio de la *Teoría de la Justicia* sí parece descansar en esta discrepancia inicial con el utilitarismo. Y ciertamente a la hora de resolver esta discrepancia Rawls no regresa ni a viejas posiciones metafísicas ni a la vieja teoría del contrato, sino que más bien adopta un constructo haciendo suyos los resultados de una metodología que ofrece un modelo de análisis capaz de resolver el inconveniente inicial del utilitarismo.

Desde esta perspectiva, y más allá de las filiaciones que Rawls pretenda establecer y con las que resulta posible arroparla, la posición original aparece como supuesto típico de los analizados por la teoría de juegos y «tiene que ser interpretada como de manera que en cualquier momento se pueda adoptar su perspectiva» ²³, una perspectiva cuyo objetivo fundamental, en palabras del propio Rawls, reside en que los distintos jugadores «aspiran a una situación lo más elevada posible en términos absolutos. No desean para sus oponentes una puntuación ni elevada ni baja, ni tampoco aspiran a maximizar o minimizar la diferencia entre sus éxitos y los de los demás» ²⁴. Aunque a continuación y en esa misma frase afirma Rawls que la idea de

un juego no se aplica totalmente, ya que las partes no están interesadas en ganar sino en obtener el mayor número de punto de acuerdo con su sistema de fines.

Este matiz que Rawls introduce señala precisamente la tensión esencial en la que se mueve desde la metodología de la teoría de juegos. Porque si la metodología pertenece a ésta y el problema mismo parece situarnos en una situación típica de deliberación colectiva, lo que por lo demás el propio Rawls reconoce de modo explícito, en realidad parece difícil resolver únicamente desde ahí los problemas a los que la teoría se enfrenta. Desde una perspectiva semejante el único modo de interpretar la posición original sería considerarla como la situación típica de un juego de cooperación que funcionaría sólo eficazmente en el caso de que todos los agentes fueran utilitaristas ²⁵.

Ahora bien, esta última hipótesis exigiría que cada uno de los jugadores conocieran sus respectivas diferencias y posiciones, lo que justamente queda excluido mediante la caracterización que se hace la posición original, y en concreto mediante el velo de ignorancia. Este último introduce unas condiciones de riesgo e incertidumbre en las que el modelo de la maximización fracasa. Frente a éste cabe entender en cambio que el velo de ignorancia introduce una situación en la que operaría más bien el axioma de la utilidad esperada o el de la cosa segura, mediante cuya aplicación los intervinientes en la posición original no obtienen una ordenación de estados sociales desde el punto de vista de la posición de cada cual, sino más bien un *ranking* de ordenaciones sociales con permutaciones interpersonales ²⁶. Pero esta interpretación parece desplazarnos del marco de la teoría de la deliberación colectiva en sentido estricto para reconducirnos a una situación de elección individual en condiciones de riesgo e incertidumbre en la que varios agentes llegarían a adoptar unánimemente los principios de justicia.

Pero no nos interesa tanto la determinación precisa de tipo de problema que desde el interior de la teoría de la elección racional entraña el velo de ignorancia y con él la posición original, cuanto el hecho de que ambas dependen en definitiva de la teoría de la racionalidad como modelo teleológico. Con ello quedaría confirmada esa primera paradoja de la filiación kantiana de Rawls, aquella según la cual lo que en Kant es deontológico es el punto de partida, mientras que en Rawls lo es el resultado, y ese resultado además pertenece en Rawls a la teoría política y no a lo moral. En nuestro recorrido hemos hecho algunas referencias a la evolución de Rawls que llega hasta *TJ* y a la evolución posterior a *TJ*, pero tales referencias eran sólo incidentales. En lo que sigue, y que constituye la segunda parte de nuestro trabajo, esa evolución será nuestro punto de mira, porque en el despliegue de la misma está encerrada la segunda paradoja fundamental que queremos traer a la luz.

Nadie duda hoy de que se puede hablar de un primer y un segundo Rawls. El primer Rawls sería el de *TJ* o incluso el de los escritos anteriores que llevan hasta el año 71. El segundo Rawls es el que culmina en *Liberalismo Político* por primera vez, una obra del año 93, pero que resume el resultado de una lenta evolución en la que se va separando de *TJ*. La idea fundamental que preside este segundo Rawls es justamente la pretensión de eliminar de su teoría el de fundamentación en términos morales y más bien el de buscar lo propio de liberalismo político. En la última versión de su obra Rawls señala tres órdenes de alteraciones en su propia evolución, las dos primeras afectan a los principios mismos de justicia, a la argumentación en favor de los mismos en la posición original, y finalmente la última al «modo en que la propia justicia como equidad ha de entenderse, a saber: como una concepción política de la justicia en vez de como parte de una doctrina moral comprensiva»²⁷.

Las dos últimas son, pues, las que afectan directamente a nuestros intereses en este trabajo, porque inciden en las relaciones con Kant, y además en los dos aspectos que articulaban la paradoja en cuestión. En cuanto a la segunda el mismo Rawls parece hacerse cargo de la cuestión relativa a la posición original, cuando afirma: «en contraste con lo que pueda sugerir la posición de *Teoría*, esta división del argumento muestra que las razones que justifican el principio de diferencia no descansan (como nada irrazonablemente han pensado K. J. Arrow y J. C. Harsanyi, entre otros) en una gran aversión a la incertidumbre considerada como una actitud psicológica»²⁸. En cuanto al último punto Rawls señala que a pesar de que en *TJ* nunca se defiende una doctrina moral, la formulación misma podía dar a entender que estábamos ante una doctrina moral comprensiva. Demostrar que esto no es así exige, según nos dice Rawls, toda una nueva familia de ideas.

En realidad y bien mirado las líneas en las que Rawls modifica su teoría recaen directamente en la paradoja que señalamos. No queremos decir con ello que de modo consciente Rawls articule su evolución únicamente a partir de a su relación con el kantismo, sino sólo más bien que esa relación con el kantismo era ya desde el principio el síntoma mismo de su teoría moral subyacente (sobre todo teniendo en cuenta que esa relación determina a la vez como la otra cara de la moneda su relación con el utilitarismo), y sigue siéndolo, como no podía ser de otro modo, en los desarrollos posteriores. Por tanto, parece obvio que esas relaciones habrán de reaparecer al afrontar esa nueva familia de ideas que propone.

La idea básica del nuevo Rawls es la ya conocida de que su teoría no pretende ser una teoría moral comprensiva, y que es más bien un fundamento para el liberalismo político, de forma que en su seno quepa una sociedad justa independiente de

cualquier teoría moral o filosófica o religiosa comprensiva. Con ello a la vez confirma algo que ya hemos señalado, a saber, que su teoría no es kantiana al menos en el sentido moral. La pregunta es si no se ve obligado a esto precisamente en parte por las críticas recibidas, a las que en cierto modo da por buenas, y además desde su posición básica a partir del individualismo metodológico. De ser esto así, entonces la evolución de Rawls se explica precisamente a partir de su paradójico kantismo. Creemos que esto es así. Para ello echa mano de nuevo del kantismo, pero ahora ya no en términos de una teoría moral, que por definición nos ha dicho que queda fuera de su campo, y sí sólo en términos de un procedimiento, lo que él llama el constructivismo kantiano en la teoría moral. Su kantismo se ha adelgazado, pues, considerablemente, y en el fondo la tendencia de Rawls parece ser una respuesta a las objeciones que en su momento se le hicieron. Los nuevos elementos que va a incorporar son la pareja racional-razonable, las nociones de consenso entrecruzado y la idea de equilibrio reflexivo.

Veamos en primer lugar su concepto de constructivismo kantiano, tal como lo formula en el artículo del 80. Allí resume esa concepción en el énfasis en la persona moral inserta en un proceso razonable de construcción. La expresión razonable aparece aquí por contraste a racional, y es una novedad importante, porque parece haber desplazado la vieja teoría del bien en sentido estricto que aparecía en *TJ* para sustituirla por ésta, en la que la persona además de racional es razonable, se reconoce a sí misma como tal y como libre, igual y autónoma, lo que daría lugar al concepto de sociedad bien ordenada. El concepto de lo razonable parece, pues, constituir una premisa nueva que represente un avance con respecto al utilitarismo de *TJ*. Más aún, parece ir directamente contra ese utilitarismo y situarnos en un plano que no depende sólo de la maximización, por lo que

resulta extraordinariamente importante en nuestro trabajo, en cuanto paso en la evolución que buscamos. En otro lugar lo precisa con más detalle al afirmar que lo razonable subordina a lo racional²⁹. Pero ahí también se dice que presupone lo racional, con lo que se viene a confirmar lo que en su momento dijimos. Pero a continuación añade Rawls que justamente el constructivismo kantiano consistiría en la unidad de la razón práctica mediante la subordinación de lo racional a lo razonable³⁰. En la última versión ese concepto aparece muy vinculado, junto con los otros elementos, a la idea de persona que ya aparecía en *TJ*, pero que ahora especifica más y mejor y parece deslizarse a un concepto más asumible por la concepción kantiana. En efecto, nos dice de las personas razonables que «están dispuestas a proponer ciertos principios (que son los que definen los términos equitativos de la cooperación), así como a acatar esos principios a expensas de sus propios intereses»³¹. Ahora sí parece que estamos ya lejos del utilitarismo en cualquier consideración que del mismo se tenga. Ahora sí parece que a la base de la teoría encontramos una concepción moral básica que se aleja de la razón técnico-práctica, puesto que de ellos nos dice Rawls que pueden ser irrazonables aunque a partir de una posición negociadora fuerte y a partir de sus intereses dados se les considera racionales. Esta distinción, que no es original de Rawls, sino que él mismo nos recuerda que procede de W. M. Sibley, evoca, por lo demás, la establecida entre una razón instrumental y una sustantiva, que a su vez es un trasunto de la distinción entre las reglas técnico-prácticas y la razón pura práctica, y debe ser completada con la concepción de personas libres e iguales que desarrolla Rawls en el párrafo 7 de la última versión, y de la que nos dice (y es esto algo que faltaba en *TJ*) que estamos ante una concepción normativa: es decir, construida por el pensamiento y la práctica³², lo que nos acerca de nuevo al

Kant que postula (léase aquí construye) la libertad.

A su vez esa idea de lo razonable se va elaborando al hilo de otra idea que es complementaria de la anterior, la de justificación pública. Esa idea, presente ya en *TJ* (parágrafo 87), es desarrollada a lo largo de los años ochenta y en *Liberalismo Político* se formula como un elemento político en los siguientes términos: «justificar nuestros juicios políticos ante los demás significa convencerles mediante el uso de la razón pública, es decir, mediante formas de razonamiento e inferencia apropiadas a las cuestiones políticas fundamentales y apelando a creencias, razones y valores políticos de los que cabe razonablemente esperar que los otros también reconocerán»³³. Esta definición evoca, sin duda, el concepto kantiano de uso privado de la razón en *Respuesta a la pregunta qué es Ilustración*³⁴, y la centralidad que Rawls la da en su nueva sistemática parece haber desplazado la centralidad dada a la posición original con toda su carga técnica a partir de la teoría de la elección racional, para sustituirla por un concepto de razón en sentido de razonable. A este respecto resulta muy ilustrativa la nota que Rawls introduce al hilo de la comparación entre racional/razonable y categórico/hipotético en su última versión de *Justicia como equidad*. Allí nos dice:

Corrijo aquí una observación de *Theory*, donde se dice que la teoría de la justicia es parte de la teoría de la elección racional. Según lo que acabamos de decir, esto es simplemente un error, e implicaría que la justicia como equidad es en el fondo hobbesiana (tal como a menudo se interpreta a Hobbes) más que kantiana. Lo que debería haberse dicho es que para describir a las partes, y su razonamiento, se usa la teoría de la elección (decisión) racional, pero que esta teoría es en sí misma parte de una concepción política de la justicia...³⁵.

La justificación pública hace referencia, por tanto, a un contexto que va más

allá de la teoría de juegos y de la racionalidad instrumental. En *Liberalismo Político* da una triple caracterización de la misma que parece apuntar hacia un horizonte que va más allá de los fines individuales, o incluso de determinadas razones colectivas. En todo caso estamos lejos de la razón pura práctica kantiana, en la medida en que su lugar sistemático es el de las llamadas por Rawls esencias constitucionales, y, por tanto, no el ámbito moral estricto. Sin embargo, el hecho de que la razón pública sea ajena a las concepciones comprensivas y tienda a desentenderse de los intereses concretos lo sitúa de nuevo en un ámbito análogo al razonamiento kantiano que obliga a generalizar las máximas subjetivas.

El resultado de ese uso público de la razón es lo que Rawls denomina el consenso entrecruzado. La idea del consenso entrecruzado tampoco faltaba en *TJ*, pero es a partir de *Liberalismo Político* cuando se desarrolla y potencia como una pieza fundamental del sistema. Lo que viene a expresar es que en una sociedad dada el uso de la razón pública está por encima de las doctrinas morales, filosóficas o religiosas comprensivas, y el resultado es un acuerdo básico sobre la ordenación social. Ahora bien, Rawls insiste en *Liberalismo Político* en que esa idea de consenso entrecruzado está lejos de ser un mero acuerdo en términos de simple transacción, y en este sentido es algo distinto a un mero *modus vivendi*. A este respecto pone el ejemplo de las comunidades católicas y protestantes del XVII que llegaron a un acuerdo acerca de la tolerancia a fin de sobrevivir, aunque cada una de ellas estaría convencida de su superioridad y si estuviera en posición de dominio eliminaría a la otra. La idea del consenso entrecruzado va mucho más allá de este acuerdo puntual, y aparece vinculada a lo que Rawls llama la estabilidad, es decir, justamente a un trasfondo que sería fruto de la razón pública.

En esta constelación aparece un último elemento de la teoría que es el del equilibrio reflexivo. Como en los casos anteriores la noción misma estaba ya presente en *TJ*, si bien va cobrando importancia a lo largo de los años ochenta para finalmente ocupar un papel sistemático considerable desde *Liberalismo Político* y en la última formulación del año 2002. La mejor definición la encontramos precisamente en esta última obra, en la que de hecho el equilibrio reflexivo aparece configurado en tres momentos diferentes: equilibrio en sentido estricto, en sentido amplio y en sentido pleno. El equilibrio estricto hace referencia a la situación de un individuo que, después de ponderar los distintos juicios morales en conflicto, alcanza una concepción en la que la mayoría de sus juicios morales armonizan con ella en términos de consistencia. El equilibrio en sentido amplio hace referencia a la posición que alcanza ese mismo individuo después de contrastar nuestra concepción con las diversas teorías morales propias de su tradición. Finalmente el equilibrio reflexivo pleno se da cuando cada ciudadano ha alcanzado un equilibrio amplio que como tal se generaliza mediante la razón pública y afecta al conjunto de la sociedad como sociedad bien ordenada³⁶. Como en las nociones de razón pública y consenso entrecruzado, de las que en realidad es inseparable, se pueden detectar aquí elementos que nos aproximan a la razón pura práctica kantiana, o que al menos nos alejan de la mera razón técnico-práctica, si bien una vez más nos movemos en el ámbito de la teoría política, pero de hecho sitúa la teoría más allá del mero contractualismo, lo que la aproxima todavía más al modelo de la razón pura práctica kantiana, puesto que el equilibrio reflexivo es algo más que un mero acuerdo entendido como transacción de intereses.

Aunque este conjunto de piezas teóricas no tienen identidad sin más con la

razón práctica kantiana (como el propio Rawls reconoce), es indudable que configuran un universo más propio de una teoría deontológica, que si bien se aplica desde el ámbito político, parecen ir empujando a Rawls a una concepción moral implícita, que es lo que denomina constructivismo kantiano. La nueva paradoja entonces estaría en el hecho de que a medida que Rawls va tomando distancia de la teoría moral, incluida la kantiana (entendida como una teoría moral comprensiva) para desplazarse al plano político, va acercándose a una teoría moral kantiana, mientras que cuando afirmaba casi implícitamente estar en una teoría moral kantiana parecía más alejado de la misma y más próximo a la utilitarista, de la que sin embargo se alejaba en el plano político. Dicho de otra manera, parece como si Rawls, queriendo huir de la paradoja inicial, desembocara en otra consistente en que, para abandonar explícitamente una teoría moral comprensiva (lo que efectivamente disolvería la primera paradoja), acudiera a un modelo de fundamentación próximo al de una de esas teorías: la kantiana.

Hay una pieza que subyace a todos estos elementos y que es condición de los mismos y en general de la concepción misma de la teoría de la justicia como teoría política y no metafísica, y que tal vez puede hacer iluminar algo esta situación. Me refiero a la noción de pluralismo razonable. Se trata de una pieza decisiva y que además permite anudar en torno a ella los dos momentos de la concepción rawlsiana. Rawls se ha referido a ella en distintos lugares, y en el fondo jugaba ya un papel importante, si bien implícito, en *TJ*. Estamos en realidad ante una premisa de su esfuerzo teórico y que se expresa en un hecho irrefragable: la diversidad de doctrinas morales comprensivas como rasgo permanente de las sociedades contemporáneas. De hecho Rawls la vincula a la idea misma de liberalismo político. En

Liberalismo Político y los distintos trabajos que conducen a esa obra se pueden encontrar constantes alusiones a esta idea como horizonte irrebalsable de la justicia como equidad entendida en términos políticos. Lo que me parece especialmente interesante de esta idea es que en sí misma contiene implícita una teoría de la modernidad, y que se resume en el hecho evidente de que en el mundo moderno el concepto de bien no es ya único, como ocurría en las sociedades premodernas. Este dato, que en una primera aproximación resulta una obviedad, tiene sin embargo una enorme importancia en orden a nuestros intereses aquí, porque es el que nos va a permitir reunificar los dos momentos de la teoría del Rawls e incidir sobre las paradojas que estamos considerando. En realidad ese pluralismo se vincula al hecho de la separación de la ética y la política, que es justamente el lugar desde el que Rawls puede resultar a la vez kantiano y no kantiano y utilitarista y no utilitarista. En efecto, su concepción moral básica coincidiría en parte con el utilitarismo, pero se apartaría de éste en orden al problema de agregación de preferencias, que es el problema político. Por lo demás precisamente en esto último se aproximaría a Kant tal como hemos señalado más arriba, y en cambio se apartaría de las premisas metafísico-morales de Kant, tal como él mismo se expresa en distintos lugares. De manera que la obra de Rawls se enmarca entonces en la encrucijada misma de los problemas morales de la modernidad, y en este sentido las paradojas que hemos señalado apuntan en realidad al núcleo mismo de la moralidad moderna. Dicho de otro modo: sus tensiones y oscilaciones entre utilitarismo y kantismo serían la expresión de las tensiones propias de la moralidad moderna y su obra el esfuerzo de síntesis entre aquellos dos extremos.

Bien mirado, lo que comparten el utilitarismo, el propio Kant, y desde luego

cualquier teoría del contrato, es precisamente su condición moderna. Pero con ello no nos referimos a un periodo cronológico, sino al hecho que las tres teorías obedecen al pluralismo moral propio de sociedades en proceso de secularización, y que intentan aplicar la razón como único instrumento para solucionar los problemas morales. En el fondo la obra de Rawls anuda las tres teorías. Precisamente su insistencia en el carácter no metafísico de la teoría y su desplazamiento al terreno de lo que llama liberalismo político obedece a su pretensión de otorgar un marco amplio que resuelva ese problema a la hora de ordenar la sociedad con criterios éticos, es decir, de justicia. Esa pretensión fundamental estaba ya dada desde el comienzo (recuérdese que se situaba a la vez en la estela de Kant y Mill), si bien su propia tradición se lo impedía, a pesar de sus pretensiones kantianas. Su evolución posterior le llevó a eliminar a la vez los elementos metafísicos del eudemonismo y del kantismo, mediante lo cual parecía resolver la paradoja inicial, para culminar en una concepción moral que se reviste de concepción política y que tiene toda la apariencia de ser kantiana, pero que no lo es en cuanto que no necesita de las oposiciones metafísicas kantianas. La paradoja resultante podría entonces tal vez resolverse considerando que en Rawls la ética y la política vuelven a unirse en una nueva articulación de los conceptos de bien y deber. Su teoría política sería su teoría ética, una ética que efectivamente es formal como la kantiana, pero que sólo juega en el límite: las esencias constitucionales; una ética que efectivamente no es metafísica como quería el utilitarismo y la tradición *eudaimónica*. La frase de Rawls que encabeza el trabajo parece confirmar efectivamente que el liberalismo político de Rawls pretende ser eso, una teoría moral, pero que sólo puede adoptar una forma política.

NOTAS

¹ El presente texto fue defendido, con este mismo título, en el marco de un seminario dedicado al bicentenario de Kant, dirigido por Ángeles Jiménez Perona en la Universidad Complutense de Madrid. Para su publicación he introducido algunas pequeñas modificaciones.

² En *Collected Papers*, 1999, pp. 388-414. Hay versión castellana en Carlos Gómez, *12 Textos fundamentales de la ética del siglo xx*, Madrid, Alianza, 2002.

³ «Justice as Fairness» se publicó originalmente en *The Philosophical Review*, en abril de 1958. *Teoría de la Justicia* aparece en el 71 como un desarrollo de la idea embrionaria ahí contenida.

⁴ *Collected Papers*, pp. 303-358.

⁵ Ídem, p. 264.

⁶ *Teoría de la Justicia*, Madrid-México, FCE, 1979, p. 48. Adviértase que la lógica del argumento descansa sobre el principio del tercer excluido, lo que exige que asumamos que esa clasificación es la única posible y que no exista un *tercer género*.

⁷ *Justicia como equidad. Una reformulación*, Barcelona, Paidós, 2002.

⁸ Edición castellana de la UNAM, México, 1984, p. 40.

⁹ *Justicia como equidad. Una reformulación*, p. 138.

¹⁰ Cfr. *op. cit.*, p. 7, n. 3.

¹¹ Ídem, pp. 42-43.

¹² Pero adviértase (sobre todo para empezar a hacer justicia al título) que aquí encontramos ya una paradoja, a saber, el supuesto kantismo de Rawls se sitúa en el ámbito político y no en el moral, cuando justamente lo que se pretendía es que fuera la base moral la que hubiera tomado de Kant. Y sería esto lo que permitiría llamar deontológica su teoría.

¹³ El concepto de persona en Rawls es ambiguo y oscilante, y va evolucionando, como toda su doctrina, hacia posiciones cada vez más próximas a Kant, sin que llegue nunca a identificarse con la condición *nouménica* dependiente de la metafísica kantiana.

¹⁴ *Teoría de la Justicia*, p. 34.

¹⁵ Ídem, p. 448.

¹⁶ Ídem, p. 469.

¹⁷ Ídem, p. 42.

¹⁸ *Lecciones sobre la historia de la filosofía moral*, Barcelona, Paidós, p. 240.

¹⁹ *Teoría de la Justicia*, p. 48.

²⁰ Cfr. A. Sen, *Elección colectiva y bienestar social*, Madrid, Alianza, 1976, p. 57.

²¹ *Teoría de la Justicia*, p. 42.

²² Cfr. R. P. Wolff, *Understanding Rawls*, Princeton, University Press, 1977, pp. 4 y 16.

²³ *Teoría de la Justicia*, p. 165.

²⁴ Ídem, p. 162.

²⁵ Lo cual no implica necesariamente que el propio Rawls deba considerarse sin como un más utilitarista.

²⁶ Cfr. A. Sen, *op. cit.*, p. 169.

²⁷ *Justicia como equidad. Una reformulación*, p. 18.

²⁸ Ídem, p. 19.

²⁹ Cfr. ídem, pp. 316-317.

³⁰ Ídem p. 319.

³¹ Ídem, p. 254.

³² Ídem, p. 49.

³³ *Liberalismo Político*, p. 53.

³⁴ Proximidad que es reconocida por Rawls. Cfr. *Liberalismo político*, p. 248, n. 2.

³⁵ P. 121.

³⁶ Estas definiciones se encuentran en *Justicia como equidad. Una reformulación*, pp. 55-57.