

# EL CONCEPTO DE UTILIDAD SEGÚN JOHN RAWLS

Felipe Martín Huete\*

## RESUMEN

Rawls tiene como propósito elaborar una teoría que represente una opción frente al utilitarismo clásico. El bienestar de la sociedad se formaría a partir de la satisfacción de los sistemas de deseos de los muchos individuos que la conforman, es decir realizar en la mayor medida el sistema general de deseos al que se llega a partir de los deseos de sus diferentes miembros. Llegamos así al concepto del principio de utilidad en que una sociedad está ordenada cuando sus instituciones maximizan el equilibrio neto de satisfacción. La justicia social es el principio de prudencia racional aplicado a una concepción colectiva del bienestar del grupo. El autor interpreta el principio de utilidad como concebido para la satisfacción del deseo racional. Para el utilitarismo la distribución correcta en cada caso es la que produce la máxima satisfacción.

## PALABRAS CLAVE

Utilitarismo, justicia, bienestar social, identidad, individuo

## SUMARIO

1. Introducción, 2. La crítica de Rawls al utilitarismo como preferencia compartida de orden supremo, 3. El problema del utilitarismo y el intuicionismo, 4. La "teoría de la justicia" frente al utilitarismo, 5. Rawls ante el utilitarismo cardinal de Harsanyi.

**Fecha de recepción: 19-11-2009**

**Fecha de aceptación: 26-01-2010**

## 1. INTRODUCCIÓN

Para Rawls el «concepto de utilidad», según el sentido tradicional, significa "satisfacción de un deseo", y admite comparaciones interpersonales que pueden al menos ser sumadas al margen. Supone también que la utilidad se mide mediante algún procedimiento independiente de las elecciones que implican riesgo, postulando una capacidad para jerarquizar diferencias entre diversos niveles de satisfacción. Este principio clásico de utilidad requiere que las instituciones estén proyectadas para maximizar la suma absoluta de expectativas de las personas realmente representativas. A esta suma se llega valorando cada expectativa por el número de personas en la posición correspondiente, y luego sumando. Así, al duplicarse el número de personas en la sociedad se duplica la utilidad total. (Por supuesto que según la idea utilitarista, las expectativas que se miden son las satisfacciones totales, disfrutadas y previstas; y no, como en la justicia como imparcialidad, únicamente los bienes primarios). En cambio, el principio de *utilidad media* exige que la sociedad maximice no el total sino la utilidad media (per capita). Este parece ser un punto de vista más moderno y fue sostenido por Mill y Wicksell. Para aplicar esta concepción a la estructura básica se requiere que las instituciones estén establecidas de manera que maximicen las expectativas de los representantes individuales, valorado según el porcentaje de la suma total.

---

\* Felipe Martín Huete es licenciado en Filosofía y Ciencias de la Educación y Doctor en Filosofía por la Universidad de Granada (España).

Pero, ¿cuál de estos principios de utilidad sería preferido en la posición original?. Para responder a esta pregunta, debemos advertir que ambas variantes llevan al mismo resultado si el volumen de la población se mantiene constante. Sin embargo, si la población varía habrá una diferencia. El principio clásico requiere que, en tanto las instituciones afecten al tamaño de las familias, la edad para el matrimonio y cosas semejantes, deberán estar conformadas de forma que obtengan el máximo de utilidad total. Esto implica que mientras la utilidad media por persona desciende lo suficientemente despacio cuando el número de individuos crece, se debería alentar un crecimiento indefinido de la población sin importar lo que haya disminuido el porcentaje. Los dos principios serían equivalentes sólo si se supone que el promedio de bienestar desciende siempre lo suficientemente rápido de manera que no exista ningún conflicto serio entre ellos. Por tanto, la utilidad se entiende en el sentido tradicional como la "satisfacción de deseos", lo cual hace posible que las discusiones cardinales entre personas sean posibles. Sin embargo, esta noción de utilidad ha sido abandonada en buena medida por la teoría económica de las últimas décadas; se ha pensado que es demasiado vaga y que no desempeña ningún papel esencial para explicar la conducta económica.

En la actualidad, la utilidad se entiende como una manera de representar las decisiones de los agentes económicos y no como una medida de satisfacción. El tipo de *utilidad cardinal* reconocido actualmente es sobre todo el derivado de la construcción de Newman-Morgenstern que se basa en la toma de decisiones entre diversas posibilidades que impliquen riesgos. A diferencia de la noción tradicional, esta medida toma en cuenta las actitudes frente a la incertidumbre y no intenta proporcionar una base para comparaciones interpersonales. Aún así, es posible formular el principio de utilidad media utilizando este tipo de medida, aunque deben tomarse algunas precauciones, por ejemplo, estas funciones de utilidad no pueden tomar en cuenta todo tipo de puntos de vista sino que deben reflejar las estimaciones que hacen las partes de aquello que promueve su bien. Si estuviesen influidas por otras razones, entonces no tendríamos una teoría teleológica.

Si se observan estas restricciones puede, sin embargo, formularse una teoría de la utilidad media que tome en cuenta el alto grado de aversión al riesgo que parece que tendría cualquier persona normal en la posición original; y conforme mayor fuese esta aversión al riesgo, más se asemejaría esta forma de principio de utilidad al principio de diferencia, al menos a lo relativo a la evaluación de los principios económicos. Rawls considera que estos principios no son iguales puesto que existen diferencias importantes entre ellos. Sin embargo, existe una semejanza: bajo una perspectiva suficientemente general, el riesgo y la incertidumbre conducen, a

ambos puntos de vista, a dar más valor a las ventajas de aquellos cuya situación es menos ventajosa. Por tanto, una aversión razonable al riesgo puede ser tan grande que la ponderación utilitarista puede ser, a efectos prácticos, tan cercana al principio de diferencia como para hacer que la simplicidad de este último sea decisiva en su favor.

## **2. LA CRÍTICA DE RAWLS AL UTILITARISMO COMO PREFERENCIA COMPARTIDA DE ORDEN SUPREMO**

Lo que pretende Rawls es demostrar en base a la superioridad de su concepción de la persona moral y de la sociedad, la superioridad del criterio de los bienes primarios frente al del utilitarismo, puesto que la doctrina de Rawls :

- 1) Admite que los ciudadanos tengan cada uno una idea de lo que es una vida buena diferente, opuesta y hasta inconmensurable.
- 2) Considera que los ciudadanos cooperan juntos en una sociedad porque comparten principios de justicia, no porque comparten concepciones del bien.

Según Rawls, el utilitarismo no hace esto.

Por tanto, la cuestión que tiene que resolver Rawls para defender la utilización del criterio de los bienes primarios en contra de un criterio utilitarista es: ¿establecen los bienes primarios más adecuadamente que la utilidad los modos de reclamar justicia a los ciudadanos, haciéndose cargo del hecho de que los ciudadanos cooperan por adhesión a principios de justicia y no porque compartan concepciones del bien? Su respuesta consiste en decir que todos ellos tienen ciertos intereses en común y los bienes primarios son asimismo de común necesidad para todos.

El utilitarismo sostiene que justa es aquella actuación, política o institución social que maximiza el bienestar agregado a los ciudadanos, sin embargo, para poder maximizar un "agregado" primero hay que conocer el bienestar de cada uno de los individuos, darle un número según una escala de medición y, luego, establecer cómo se agregará. El bienestar de cada individuo lo dan las preferencias que tiene por las políticas –instituciones, acciones, etc-.

Lo que interesa a Rawls señalar es que para hacer comparaciones entre ordenamientos individuales de preferencias se requiere que tengamos un patrón común de valores que nos permita decir con seguridad que "ser el individuo *i* con la política *A* es mejor que ser el individuo *j* con la política *B*". Ese patrón común de valores es lo que Rawls llama "función de preferencia compartida de orden

supremo", que si, por un lado, tiene la ventaja de facilitar el cálculo de qué política sería la que produciría mayor bienestar en la sociedad, por otro significa que hay una unanimidad en considerar que ciertas cosas son buenas y que la justicia consiste en tener más de ellas que menos.

Lo que molesta a Rawls de esta manera de realizar los juicios de justicia es que obliga a todos los individuos a querer lo mismo y esto es contrario a la tesis liberal que él defiende de que todos los individuos tienen una concepción diferente de lo que es deseado.

La noción de preferencia compartida de orden supremo es claramente incompatible con la concepción de la sociedad bien ordenada propia de la concepción de la justicia como equidad. En la sociedad bien ordenada de la concepción de la justicia no existe por tanto una preferencia compartida de orden supremo sobre cuya base pudiera hacerse una evaluación compartida de las situaciones globales de las personas.

Rawls señala que si el utilitarismo puede operar de esa manera imponiendo una concepción de lo valioso es porque no tiene una concepción adecuada de lo que es ser persona. Pues, esta depende de tener una identidad; y tener una identidad es equivalente a tener una concepción del bien propia, puesto que, en primer lugar, la noción de preferencia compartida de orden supremo implica que tales personas no tienen una determinada concepción del bien con la que estén comprometidos, sino que consideran sus distintos deseos y capacidades como rasgos que han de ser ajustado a la búsqueda del lugar más alto posible de la jerarquización pública. Las personas de la concepción utilitarista en realidad son "*personas vacías*", puesto que no tiene convicciones o carácter, algo que es permanentemente valioso, sino que solamente tienen deseos que pretenden satisfacer. Pero lo que más preocupa a Rawls es que cuando las personas son tratadas como si no tuvieran identidad, al mismo tiempo son tratadas como si no tuvieran autonomía o estructura moral. Esto puede llevar a configurar una sociedad democrática y pluralista, pero inaceptable.

Rawls pretende por tanto, señalar la diferencia filosófica de concepción de la persona y de la sociedad que hay tras su propuesta de la justicia como equidad y la del utilitarismo, destacando que su teoría de la justicia como equidad es una teoría liberal, mientras el utilitarismo no lo es.

### **3. EL PROBLEMA DEL UTILITARISMO Y EL INTUICIONISMO**

Según Rawls, las principales virtudes del utilitarismo son:

1) Su terminante afirmación del valor fundamental de la felicidad humana.

2) Su carácter constructivo, es decir, su enunciación de una norma o procedimiento mediante el cual las cuestiones éticas encuentran una respuesta y las disputas éticas una solución.

3) Su aptitud como principio para el arreglo de las cuestiones de política social.

Los dos defectos más evidentes del utilitarismo son:

1) Su imposibilidad de explicar el grado de racionalidad con que han de ser guiados los interesados en elevar al máximo su propio placer, a fin de que sustituyan, como objeto de sus acciones, su felicidad personal por la felicidad general.

2) Las implicaciones manifestantes contra-intuitivas, y, a veces, auténticamente aborrecibles, de su principio fundamental.

En estas disputas, las simpatías y las antipatías de Rawls se hallan divididas por igual entre el intuicionismo y el utilitarismo. Moralmente, él se encuentra más cómodo, sin duda, con los intuicionistas que con los utilitaristas; pero metodológicamente, su corazón esta con los utilitaristas y con los economistas neoclásicos que adoptaron el utilitarismo como fundamento moral de sus construcciones teóricas. El utilitarismo, incluso en sus versiones más elaboradas y complicadas, apoya el sacrificio de ciertas personas a la felicidad de otras<sup>1</sup>. Rawls sostiene que el utilitarismo trata de seres humanos individuales como otras tantas dimensiones a lo largo de las cuales puede distribuirse la felicidad, más que como agentes morales autónomos cada uno de los cuales sigue, independientemente, una vía de acción con razón y dignidad, libremente elegida. Pero el intuicionismo es una simple expresión de convicción, una confesión del fracaso de la razón práctica. Incluso al más honesto y bienintencionado agente moral que se pregunta lo que ha de hacer o que percibe un conflicto entre dos obligaciones reconocidas, el intuicionista sólo puede decirle: "consulta a tus intuiciones". Si mis intuiciones difieren de las del intuicionista, él no puede presentar, literalmente, ninguna argumentación para convencerme, por muy dispuesto que yo esté ha escucharle.

---

<sup>1</sup> Rawls considera que la idea principal del utilitarismo es que cuando las instituciones más importantes de la sociedad están dispuestas de tal modo que obtienen el mayor equilibrio neto de satisfacción distribuido entre todos los individuos pertenecientes a ella, entonces la sociedad está correctamente ordenada y es justa.

Llegado a este punto, Rawls se vuelve a una tercera tradición, no precisamente la filosofía moral o social, sino más bien la teoría política: la tradición del contrato social. Desde el punto de vista de la teoría moral, hay, en realidad, dos tradiciones en torno al contrato:

La primera, representada por *Locke*, supone una teoría moral como una de las premisas del razonamiento relativo al contrato, donde la sustancia y la validez del contrato descansan en la propia teoría.

La segunda tradición relativa al contrato procede de *Rousseau*. El contrato cambia, no sólo la sustancia, sino también la forma de razonamiento moral de los hombres. Pero al establecer el convenio, cada individuo adquiere una voluntad general, o adquiere por primera vez la posibilidad de tener o de ejercer una voluntad general. La transformación es moral y no psicológica.

Rawls propone superar el punto en que las teorías morales del utilitarismo y el intuicionismo se han atascado, invocando una versión de la teoría del contrato social en su forma rusioniana. Al hacerlo así, quiere unir la filosofía moral y la social, de un modo que recuerda más a Platón que a Locke, Bentham o Mill.

#### **4. LA "TEORÍA DE LA JUSTICIA" FRENTE AL UTILITARISMO**

Argumentos a favor del utilitarismo:

Las instituciones básicas de la sociedad no deben distinguirse simplemente por ser ordenadas y eficientes: ellas deben ser, sobre todo justas<sup>2</sup>. Y si no lo son, entonces deben ser reformadas o abolidas. Rawls define como el primer objetivo de sus escritos el de "elaborar una teoría de la justicia que sea una alternativa viable a las doctrinas que han dominado largamente nuestra tradición filosófica". Estas doctrinas rivales a las que se refiere este autor son el intuicionismo y, sobre todo, el utilitarismo<sup>3</sup>. El autor defiende el utilitarismo como aquella doctrina que considera que un acto es correcto cuando maximiza la felicidad general. Ya en esta primera formulación, el utilitarismo aparece como más y menos atractivo que el intuicionismo. Por un lado, el utilitarismo posee un método capaz de ordenar diferentes alternativas, frente a posibles controversias morales (según dijéramos, la mejor opción, conforme al utilitarismo, es la que más contribuye al bienestar general). El intuicionismo, en cambio, carece de dicha capacidad. Sin embargo, Rawls tiende a rechazar el utilitarismo, en su carácter de concepción "teleológica" o

---

<sup>2</sup> Rawls, J., *Teoría de la Justicia*, [1971], Traducción de María Dolores González, Fondo de Cultura Económica, México, 6ª reimpresión, 2006, p. 17.

<sup>3</sup> Rawls se refiere al utilitarismo clásico.

“consecuencialista”. Como muchos otros liberales, defenderá una concepción no-consecuencialista (“deontológica”); es decir, una concepción conforme a la cual la corrección moral de un acto depende de las cualidades intrínsecas de dicha acción y no de sus consecuencias.

Pero, ¿cuáles podrían ser las razones que expliquen el atractivo generado por el utilitarismo?. Ante todo, conviene señalar el hecho de que el utilitarismo nos sugiere que evaluemos las distintas alternativas en juego considerando los intereses de los distintos individuos que se podrían beneficiar o perjudicar a partir de tales opciones. Este solo hecho merece ser destacado considerando que otras concepciones de justicia se desentienden de lo que sus propuestas puedan implicar efectivamente a las personas “reales” sobre las cuales va a recaer la solución que discuten. Así, frente a la propuesta de censurar cierto tipo de comportamientos, el utilitarismo nos incitará a preguntarnos: ¿por qué adoptar tal curso de acción?, ¿qué persona resultará efectivamente afectada o beneficiada por tal decisión?, ¿por qué censurar tales conductas sin no perjudican a nadie?. Esta peculiar mirada, que toma como punto de partida la suerte de los individuos “reales”, de “carne y hueso”, sitúa al utilitarismo como una postura en principio atractiva, al menos, frente a otras alternativas opuestas. Por otra parte, el utilitarismo resulta atractivo porque prejuzga sobre los deseos y preferencias de los distintos individuos cuya suerte se encuentra en juego: a la hora de elaborar sus propuestas, el utilitarismo sugiere tomar en cuenta las preferencias de cada uno de los afectados, con independencia del contenido específico de los particulares reclamos de cada uno de ellos.

Por ejemplo, a la hora de pensar como organizar la economía de una sociedad, una propuesta destinada a orientarla según preceptos socialistas, se situará en pie de igualdad con otra destinada a organizar un sistema de libre mercado. Del mismo modo, una petición de mantener un medio ambiente libre de impurezas contará tanto como la del empresario que proponga privilegiar el crecimiento industrial aun a pesar de los costos ambientales que ello involucre. El utilitarismo no dejará fuera de juego a ninguna de tales peticiones. Nos obligara a preguntarnos, en cambio, ¿cuál es la propuesta que satisface la mayor cantidad de intereses?. A través de este tipo de actitudes, libres de prejuicios, el defensor del utilitarismo va aún más allá del criterio señalado: su propuesta no sólo pretende mostrarse ciega (libre de prejuicios) frente al contenido de las distintas peticiones, sino también respecto de los particulares titulares de tales deseos o intereses.

Otro argumento a favor del utilitarismo es el de su carácter igualitario. Para muchos autores liberales, como R. Dworkin, el

igualitarismo de esta posición representa el dato más interesante de la misma. Este invocado igualitarismo aparece en el hecho de que el utilitarismo considera como iguales las distintas preferencias en juego, frente a un particular conflicto de intereses. Por ejemplo, en una sociedad en donde la mayoría de los habitantes prefiere utilizar los recursos existentes para distribuirlos entre los más pobres, mientras que el grupo restante -más rico- prefiere construir campos de golf, el utilitarismo privilegiará la pretensión de la mayoría.

En este sentido, el utilitarismo muestra su estricto compromiso igualitario: no hay nadie cuyas preferencias cuenten más que las de los demás cuando de lo que se trata es de reconocer cuál es la preferencia que consigue acaparar mayor respaldo social. Finalmente podría agregarse lo siguiente: en su habitual recurrencia a cálculos de costos y beneficios ¿cuántos se benefician de tal medida?, ¿cuántos resultan perjudicados?, el utilitarismo lleva adelante una operación que todos tendemos a realizar en nuestros razonamientos cotidianos: a todos nos parece razonable recurrir a la realización de balances que pueden acabar en la aceptación de ciertos sacrificios presentes en pos de mayores beneficios futuros.

#### Argumentos en contra del utilitarismo:

Hasta aquí tenemos una serie de argumentos que ayudan a ver el utilitarismo como una postura más bien irreprochable. Sin embargo, lo cierto es que cada una de las consideraciones presentadas a favor del utilitarismo parece tener una contracara poco atractiva, hecho éste que terminará mostrándonos el utilitarismo como mucho menos interesante de lo que hasta aquí resultaba. Según estas consideraciones, el utilitarismo simplemente reproduce, en una escala "social", nuestra tendencia a aceptar ciertos sacrificios presentes, con el objeto de obtener mayores beneficios en el futuro.

Una de las críticas más interesantes que Rawls realiza del utilitarismo es que el utilitarismo tiende a ver la sociedad como un cuerpo, en donde resulta posible sacrificar a unas partes en virtud de las restantes. Y dicha operación puede ser tildada como ilegítima porque desconoce lo que Rawls denomina independencia y separabilidad entre las personas: es decir, el hecho de que cada individuo debe ser respetado como un ser autónomo, distinto de, y tan digno como, los demás. Este ejercicio globalizante propio del utilitarismo, nos habla de una especial y muy sólida justificación adicional. Frente al utilitarismo, Rawls objeta también al presupuesto según el cual el bienestar es el aspecto de la condición humana que requiere atención normativa. Y critica esto por dos razones:

1) Por un lado, esta perspectiva implica tomar como relevantes lo que podríamos llamar los "gustos caros" de las personas. Por

ejemplo, una persona que se considera satisfecha con una dieta a base de leche, pan, y garbanzos, frente a otra que reclama platos exóticos y los vinos más caros. Una postura como el utilitarismo del bienestar deberá dotar al último con más recurso que al primero, para evitar que aquél obtenga menor satisfacción final que el que se conforma con la dieta más modesta. Esta es la razón por la que Rawls va a defender una métrica objetiva (los bienes primarios) y no subjetiva, a la hora de determinar cómo distribuir los recursos de la sociedad de un modo justo, igualitario.

2) Por otro lado, critica el utilitarismo por darle cabida a las preferencias o gustos "ofensivos". Con esto quiero decir que, en el "conteo" propuesto por el utilitarismo, pueda resultar computado, por ejemplo, el placer que una persona obtenga de discriminar a otra, o de dejar menos espacio de libertad a otros.

Dworkin se ocupa de mostrar el modo en que el utilitarismo termina frustrando su original promesa igualitaria. El argumento de Dworkin se basa en la idea de las preferencias "externas", esto es, preferencias acerca de la asignación de bienes hacia otras personas. La idea es que el utilitarismo deja de mostrarse como una postura igualitaria cuando permite que ingresen en el "cálculo maximizador" preferencias externas, y no, exclusivamente, preferencias personales, esto es, preferencias relativas a los bienes que reclamo para mí.

De acuerdo con Dworkin, el único modo en que el utilitarismo puede asegurar el mismo respeto a cada individuo es a través de la incorporación de un cuerpo de derechos, capaces de imponerse a reclamos mayoritarios basados en preferencias externas. Los derechos funcionarían como límites destinados a impedir que alguna minoría sufra desventajas en la distribución de bienes y oportunidades, en razón de que una mayoría de individuos piense que aquellos pocos son merecedores de beneficios menores de los que la mayoría recibe.

Estas observaciones nos muestran que el utilitarismo incumple sus promesas originales. En verdad, el utilitarismo no nos presenta una alternativa atractiva cuando nos dice que no prejuzga ni sobre el contenido de determinadas preferencias ni sobre sus particulares titulares.

Autores como Rawls y Dworkin sostienen que el utilitarismo no parece garantizarnos genuinamente aquello que nos prometía, al decirnos que las distintas soluciones en juego habrían de evaluarse a partir del impacto que ellas causen sobre los propios individuos. Lo interesante de dicha propuesta consistía en tomar como punto de partida a los sujetos "reales", en lugar de partir de meras abstracciones que desconsideraban la situación de los mismos. Por

tanto, vemos como el utilitarismo es compatible con la producción de ciertas violaciones de derechos (los derechos de una minoría), en nombre del bienestar general y mayoritario. Este tipo de críticas al utilitarismo puede acompañarse con otras similares, como la siguiente: muchas veces, lo que una persona prefiere puede ser contradictorio con aquello que resultaría más valioso, y ello, no en razón de haberse “adaptado” o “resignado” frente a situaciones injustas sino por causas tales como la ausencia de una información empírica adecuada. Esta observación nos lleva a la brutal recomendación de desconocer las preferencias de los individuos.

Otra nueva objeción al utilitarismo, también sugerida por Rawls, es que una propuesta ofrecida por el utilitarismo no sería capaz de reunir apoyo en una situación contractual hipotética. Esto es, si tuviéramos la oportunidad de discutir acerca de qué teoría de la justicia debería organizar nuestras instituciones, tenderíamos a dejar el utilitarismo de lado, y ello porque el utilitarismo termina mostrándose como una doctrina demasiado exigente. Al adoptarse una concepción como la utilitarista, es posible esperar que surjan situaciones en las cuales los derechos fundamentales de algunos resulten puestos en cuestión en nombre de los intereses de la mayoría. Por eso, el sistema no sería estable a menos las que tienen que sacrificarse se identifique fuertemente con intereses más amplios que los suyos. Lo dicho es lo que lleva a Rawls a decir que dicha doctrina no es capaz de “asegurar las bases de su propia estabilidad”.

## **5. RAWLS ANTE EL UTILITARISMO CARDINAL DE HARSANYI**

La propuesta de Harsanyi viene a ser una elaboración de la teoría del observador imparcial. Alguien que desee hacer un juicio de bienestar tiene que situarse en la actitud de ignorancia con respecto a su posición personal en las distintas situaciones sociales alternativas; para ello tiene que considerarse como si tuviera una igual probabilidad de ocupar todas las posiciones sociales en cada una de las situaciones que se le ofrecen (equiprobabilidad). Pero ésta es ya claramente una elección bajo incertidumbre y viene exigida por la propia naturaleza de los juicios de valor.

El modelo de Harsanyi adopta los elementos que ya eran ineludibles para el problema de la fundamentación. El núcleo de esta diferencia radica en la dicotomía entre lo individual (la utilidad) y lo social (el bienestar) que ha dejado de ser clara o no problemática. Por un lado, el bienestar social ha perdido su status de pura cantidad objetiva; por otro lado, la individual ya no es una dimensión tan subjetiva. Se podría decir que han nacido nuevos conceptos que llevan el mismo nombre que los antiguos. Así, el bienestar social se concibe ahora con Harsanyi como un juicio que cada individuo puede formular y en el que legítimamente puede discrepar con sus

conciudadanos; el bienestar social es una multiplicidad de juicios sobre el bienestar social. Lo mismo ha ocurrido con la utilidad individual: De ser una preferencia que no tenía en cuenta otra cosa que el beneficio personal del que la formulaba se ha convertido en un reflejo de la situación económica de todos los individuos afectados.

La función de utilidad de un individuo se realiza en base de sus propias preferencias personales (que para Harsanyi incluye tanto los gustos como los valores) refleja, pues, preferencias subjetivas que mostrarían “una actitud egoísta en el caso de un altruista”. En cambio, las funciones del bienestar social reflejan lo que prefiere “sólo aquellos posiblemente raros momentos cuando se fuerza a sí mismo a adoptar una actitud especial impersonal”. Estas son sus preferencias éticas y reflejan aquel compromiso de imparcialidad constitutiva de los juicios sociales.

El problema que surge es ¿cómo amalgamar las funciones de utilidad individual en una función de bienestar social? La construcción de una función de bienestar social en el modelo de Harsanyi se realiza bajo los tres siguientes postulados:

1) *Racionalidad individual*. Las preferencias personales de todos los individuos satisfacen los axiomas de racionalidad.

2) *Racionalidad colectiva*. Las preferencias éticas de un individuo  $j$  satisfacen los axiomas de racionalidad.

3) *Individualismo*. Si todos los individuos son personalmente indiferentes entre dos situaciones sociales, entonces el individuo  $j$  será éticamente indiferente entre ellas.

De este modo, se construye el puente entre la subjetividad y la intersubjetividad de los juicios de valor de un individuo; los postulados excluyen que el juicio sea una preferencia subjetiva e indican la forma que ha de tener la función de bienestar social para que refleje la actitud de imparcialidad que la intersubjetividad de las decisiones sociales requiere.

La equiprobabilidad es la piedra angular del modelo de Harsanyi, puesto que es lo que permite dar un modelo común a la construcción de las funciones de utilidad de otros individuos y a la comparación de esas funciones entre sí. Así, pues, el modelo de Harsanyi se define como un modelo utilitarista que utiliza medidas cardinales y comparaciones interpersonales.

El modelo de Harsanyi articula un punto de vista ético apoyado en la teoría del observador imparcial. Pero, asimismo su intención es defender la posibilidad de una función de bienestar social cuyos

argumentos son funciones de utilidad personales construidas comparadamente. Para ello la pieza clave es también la equiprobabilidad<sup>4</sup>, puesto que al incorporar un punto de vista único e imparcial, permite que las funciones de utilidad individuales se construyan con una misma escala y sean comparables para la persona que se sitúa en ese punto de vista. Gracias a esa peculiar actitud “como si” cada individuo puede construir la función de utilidad de una situación social para cada individuo. Esto ha de ser entendido como la exigencia de juzgar el valor de cada alternativa social según las preferencias, los gustos, los valores, los intereses, etc., de cada uno de los afectados; algo como: “Si yo fuera el individuo A en el estado social x, mi función de utilidad para este estado sería...”. Construidas bajo esa actitud, las utilidades de los distintos individuos ya no encierran un problema de interpersonalidad, sino de intrapersonalidad; es decir, de obtener ordenaciones agregadas que no dependan de reglas por las que se han asignado los números a las utilidades de todos los individuos.

Según Harsanyi, las comparaciones de utilidad han de estar basadas en una información empírica lo más fiable posible. Admitir otra cosa sería introducir la arbitrariedad en los juicios sociales. En concreto, sería reducir las preferencias éticas a preferencias subjetivas, rompiendo la tensión de todo el modelo hacía la convergencia intersubjetiva de todas las preferencias éticas individuales.

## CRÍTICA DE RAWLS AL UTILITARISMO DE HARSANYI

Lo que Rawls objeta con respecto al utilitarismo de Harsanyi es que el principio de utilidad media no sería elegido por los participantes en el contrato hipotético, porque en esa situación no existen bases para calcular las probabilidades de los resultados, sino meras posibilidades de que un principio lleve a un tipo de consecuencias o a otras.

Esta argumentación se basa en tres supuestos que expone Rawls:

1) “... la situación tiene que ser tal que resulte imposible el conocimiento de las alternativas posibles o que, al menos, sea altamente inseguro”

---

<sup>4</sup> El supuesto de equiprobabilidad, produciría la misma clase de homogeneidad de las preferencias éticas y la misma unanimidad con respecto a la elección del mejor estado del mundo que consigue Rawls con su *velo de la ignorancia*. Vid. HABERMAS, J., Rawls, J., *Debate sobre el liberalismo político*, Traducción de Gerard Vila Roca, Paidós- I.C.E. de la Universidad de Barcelona, Barcelona, 1998.

2) "... la persona que escoge tiene una concepción del bien tal que le importa muy poco o nada lo que pueda ganar todavía por encima del mínimo que seguramente obtiene al seguir la regla maximin".

3) "... las alternativas rechazadas tienen resultados que difícilmente pueden aceptarse".

En primer lugar, Rawls considera también que una elección influida por las actitudes ante el riesgo de los participantes carecería de cualquier significado moral. Con todo, la situación de la elección inicial se caracteriza por ser una elección incierta y arriesgada. En segundo lugar, señala la no admisión en su modelo de otros juicios de probabilidad que aquellos que tengan una base objetiva, por ejemplo, las probabilidades empíricas. Por último, Rawls se dirige en contra del núcleo mismo del modelo de Harsanyi: al señalar que su función de bienestar social la realiza un solo individuo que observador imparcial; de donde se sigue que "las utilidades representan el valor de las alternativas para este individuo, estimadas en función de su esquema de valores. Rawls cree que esperar de ello un juicio intersubjetivo tiene tanto sentido como esperar que alguien tenga valores, fines, expectativas que no le sean propios.

Pero la crítica de Rawls es matizada. No argumenta que las comparaciones interpersonales tal y como las piensa Harsanyi bajo el mecanismo de "ponerse-en-el-lugar-del-otro" sean un sin sentido. Este proceder es adecuado cuando lo que se trata de comparar y evaluar son casos como la posición social, la riqueza, el bienestar. La posición original de Rawls es precisamente una situación diseñada pensando en que no hay una única concepción del bien, sino que hay muchas y conflictivas. En consecuencia, el acuerdo al que se pretende llegar no consiste en que finalmente haya una expectativa única para todos los miembros de la sociedad, sino unos principios que favorezcan las múltiples expectativas de dichos miembros. Esto es lo que no consiente el procedimiento de Harsanyi y es por esto por lo que se rechaza la posibilidad de las comparaciones en la situación de elección social.

## **RÉPLICA DE HARSANYI A RAWLS**

Harsanyi responde una por una a las objeciones que Rawls presenta a su procedimiento en cada uno de esos tres supuestos<sup>5</sup>:

En primer lugar, Harsanyi señala que, incluso si el decisor se rige por la regla maximin, ello no le evita hacer una elección entre

---

<sup>5</sup> Vid. HARSANYI, J., "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking", *Journal of Political Economy*, 1953.

probabilidades subjetivas, al menos implícitamente. Sobre esta base intenta destruir el argumento rawlsiano de que sólo el principio maximin podría justificar una elección frente a las generaciones futuras, por cuanto no lleva los riesgos que el principio utilitarista inevitablemente supone. Harsanyi supone en contra que si bien la equiprobabilidad puede interpretarse –como hace Rawls- como una variante del principio de razón insuficiente, otra interpretación es posible; puede ser la expresión del principio normativo de dar a los intereses de cada individuo el mismo peso. Si se acepta tal principio – y no aceptarlo es éticamente muy problemático-, asignar la probabilidad de uno o casi uno a la posibilidad de ser el individuo peor situado supone ponderar esa situación de una manera difícilmente justificable. En segundo lugar, Harsanyi explica la introducción de las actitudes ante el riesgo en las elecciones sociales. Si las funciones de utilidad se construyeran en base a las preferencias subjetivas de los individuos, ciertamente contendrían un elemento arbitrario de actitud individual ante el riesgo que vaciaría de valor moral a las decisiones sociales. Finalmente, con respecto a las comparaciones interpersonales, Harsanyi le concede a Rawls que...

“... la idea de evaluar las circunstancias personales de otro individuo en función de su propia función de utilidad –y no en términos de la nuestra- resulta un concepto difícil”.

Mientras que Rawls señala la imposibilidad de reducir los sistemas de valores de las distintas personas a un común denominador, Harsanyi insiste más bien en que las “preferencias y funciones de utilidad de todos los seres humanos están gobernados por las mismas leyes psicológicas básicas”. Esto es lo que posibilita el funcionamiento del mecanismo del “ponerse-en el-lugar-del-otro”, cuyo resultado es que cualquier individuo sea capaz de reconstruir la función de utilidad de cualquier otro individuo. Ello reduciría el problema de la interpersonalidad a un problema de comparaciones de las funciones de una misma persona: la propia y las que se tienen cuando uno se considera en la piel de cada uno de los demás.

La argumentación de Harsanyi en contra de los supuestos de Rawls tiene como objetivo la defensa del principio de maximización de la utilidad esperada en contra del principio maximin como criterio para la elección del principio de justicia social. El diseño de la posición original contractual responde a los mismos objetivos que el de la situación imparcial de equiprobabilidad: la fundamentación de los principios. Harsanyi argumenta en contra de la utilización del principio maximin en la posición original mostrando que el principio de diferencia que el selecciona produce decisiones sociales con la misma característica de irracionalidad que produce la aplicación directa del criterio maximin.

Curiosamente, Rawls había previsto esta objeción y su réplica consiste en señalar la diferencia entre la regla maximin y el principio de diferencia: "... *el problema de la justicia no consiste en distribuir ad libitum entre ciertos individuos diversas cantidades, sea dinero, propiedades o cualquier otra cosa*".

Es decir, el problema de la justicia no es un problema de distribución pura. La justicia social es un conjunto de dimensiones complejas que está presente en cualquier problema; correspondientemente, la organización social a la que dan forma los principios de justicia también en un sistema complejo de situaciones. Por eso, la regla maximin es inaplicable a las situaciones sociales; en realidad, no puede aspirar siquiera a ser un principio de justicia. Por lo contrario el principio de diferencia sí lo es y tratarlo como una simple regla maximin es una reducción errónea.

La discusión de Rawls con Harsanyi viene a mostrar algo sobre los principios que sus procedimientos seleccionan y que no es muy agradable para ellos. Se trata de que, en uno o en otro, alguien siempre parece tener una razón poderosa para no aceptarlos. El individuo A podría no considerar racional elegir el principio de diferencia en la posición inicial por razones similares a las que presenta Harsanyi. A su vez, el individuo B –o cualquiera de los situados en las peores situaciones sociales- tendría argumentos del tipo presentado por Rawls para no aceptar el procedimiento de Harsanyi. Pero hay que recordar que toda la operación contractual sólo tiene por objetivo encontrar las bases para un acuerdo unánime sobre los principios sociales. Pues bien, esto es lo que no se consigue; los principios no son aceptables por todos.

## **BIBLIOGRAFÍA UTILIZADA.**

Dworkin, R., *Taking rights seriously*, Cambridge: Harvard University Press, 1977.

Rawls, J., *Debate sobre el liberalismo político*, Paidós Ibérica, 1998.

Rawls, J., *La justicia como equidad*, Tecnos, 2002.

Rawls, J., *La justicia como equidad: una reformulación*, Paidós Ibérica, 2001.

Rawls, J., *Lecciones sobre la historia de la filosofía moral*, Paidós Ibérica, 2006.

Rawls, J., *El liberalismo político*, Crítica, 2004.

Rawls, J., *Sobre las libertades*, Paidós Ibérica, 1990.

Rawls, J., *Teoría de la justicia*, Fondo de Cultura Económica, 2006.

Harsanyi, J., "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking", *Journal of Political Economy*, 1953.

Harsanyi, J., "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility", *Journal of Political Economy*, 1955.

Harsanyi, J., *Essays on Ethics, Social Behavior, and Scientific Explanation*, Dordrecht, Holland: Reidel Publishing Company, 1976.