

UNIVERSIDAD DE GRANADA
FACULTAD DE FILOSOFIA Y LETRAS
SECCION DE FILOSOFIA

CUESTIONES DE FUNDAMENTACION EN JUSTICIA ECONOMICA

Tesis para la obtención del grado
de Doctor en Filosofía, presentada
por el Licenciado en Filosofía
D. Damian Salcedo Megales, y rea-
lizada bajo la dirección del Ca-
tedrático de Etica Dr. D. Gilberto
Gutierrez Lopez.

Granada, 1985

If we want to do that we must
come out of the cave.

E. M. Forster

CONTENIDOS

	pg.
NOTA PRELIMINAR	4
INTRODUCCION	8
CAPITULO PRIMERO: ECONOMIA DEL BIENESTAR Y ELECCION SOCIAL	27
CAPITULO SEGUNDO: TEORIA DE LA UTILIDAD Y ELECCION SOCIAL	118
CAPITULO TERCERO: RACIONALIDAD COLECTIVA	198
CAPITULO CUARTO: JUSTICIA ECONOMICA	274
CONCLUSIONES	355
BIBLIOGRAFIA	387
INDICE GENERAL	405

NOTA PRELIMINAR

"El individuo ya no se enfrenta a su identidad colectiva a la manera de un contenido tradicional en base al que se pudiera formar la propia identidad, como si de un objetivo prefijado se tratara; lo que sucede es, más bien, que los individuos toman parte en el proceso de formación (y, en un estado ulterior, de formación de la voluntad) de una identidad solo esbozable en común. La racionalidad de los contenidos de identidad se mide entonces sólo en base a la estructura de ese proceso de generación, esto es: a las condiciones formales de realización y comprobación de una identidad flexible en la que todos los miembros de la sociedad puedan reconocerse y respetarse recíprocamente".

J. Habermas, La reconstrucción del materialismo histórico.

El interés de esta investigación es plantear la discusión sobre la justicia económica en el marco analítico de la teoría de la elección social. ¿Por qué consideramos este marco como relevante para el estudio de la problemática de la justicia económica? Hay diversas razones particulares que son objeto de nuestra atención en los capítulos que siguen. Pero hay una razón general que queremos presentar aquí: la crisis de

la convivencia. Probablemente por las causas que explica Habermas en el texto del que hemos extraído la cita hoy hemos llegado a no saber como fundar la convivencia social. Sin embargo, las sociedades modernas se ven obligadas a decidir -bajo el supuesto de que tienen una identidad- qué van a hacer o qué futuro quieren darse; es decir, están obligadas a decidir las formas de la convivencia. Por eso sus decisiones -la toma de decisiones en sí misma- son inevitablemente ético-normativas. Pues bien, lo que la teoría de la elección social refleja es la crisis de valores comunitarios que todos sentimos como una crisis específica de los fundamentos de la toma de decisiones. A través de la teoría de la elección social es posible estudiar fiablemente cómo y por qué hay valores que son inútiles en los procesos de elección social o por qué esos valores son incompatibles. Porque su objetivo primario -en tanto que teoría normativa- es poner en claro la problemática de la definición de una racionalidad moral para los fines colectivos.

El problema tradicional de la distribución económica -recuérdese la polémica de A. Smith contra B. de Mandeville- también es un problema de convivencia. Preguntarse cómo es posible justificar las políticas económicas que producen resultados desigualmente beneficiosos para los grupos sociales que constituyen la sociedad consiste en intentar pensar el problema de la paz

social: ¿cómo hacer convivir a los intereses sociales diversos si estos no están seguros de la justicia del esquema bajo el que cooperan? Fundar la justicia de las decisiones políticas requiere abordar el problema normativo básico de la definición de la racionalidad moral de las elecciones sociales.

Así la teoría de la elección social aparece ante nosotros -ciudadanos de las sociedades post-industriales- como un instrumento muy atractivo para explorar la posibilidad de que exista la justicia en nuestro mundo. Otros enfoques son posibles -notoriamente el de las éticas sistemáticas, como la de Rawls-; pero si hemos preferido éste, ha sido porque nos ofrece un instrumental formal para comprobar la validez, los límites y las consecuencias de prestar nuestro apoyo a cualquier valor social -incluidos lo que proponen las éticas sistemáticas. En definitiva, lo que se obtiene del estudio de la teoría de la elección social es la comprensión de las condiciones a las que ha de atenerse la definición de la racionalidad moral para los fines sociales; y, por tanto, de las condiciones de posibilidad de la justicia social.

Debo al profesor Gilberto Gutierrez Lopez la posibilidad de haber comprendido de esta manera la problemática de la filosofía política y del momento histórico en que vivimos. Le debo a él y la profesor José García Leal el esfuerzo, la paciencia y el afecto que

han puesto en ayudarme para que esta Tesis fuera una realidad. A ellos quiero dedicar el valor que la misma tenga como muestra pública de mi gratitud.

He de manifestar mi agradecimiento a la Fundación "Oriol-Urquijo" que durante tres años ha sustentado e económicamente la investigación aquí realizada. Asimismo, a Antonio Ruiz Molina, profesor de la Facultad de Ciencias Económicas de Málaga, por su ayuda a la hora de conseguir una buena parte de la bibliografía económica utilizada. Por último, a Juan Cañavate Toribio, amigo, que realizó las representaciones gráficas que se incluyen en los capítulos I y IV. A todos ellos mi sincero agradecimiento.

INTRODUCCION

Este trabajo está dedicado al estudio de aquellas decisiones tomadas por la sociedad que nos plantean cuestiones de justicia económica. Pero, el ámbito de la justicia económica no está claro. Frente a los autores que sostienen que el objeto propio de la justicia económica es la distribución de un bien muy especial, a saber, el dinero, nuestra posición es menos exclusiva. Si la teoría de la justicia distributiva es la que se ocupa genéricamente de las particiones de beneficios y cargas entre los miembros de la sociedad -y no sólo de bienes-, entendemos que el ámbito propio de la teoría de la justicia económica es el de la división de los beneficios y cargas que genera específicamente la renta global de una comunidad. No, por tanto, exclusivamente la renta, sino más bien el binomio beneficio-renta. A ese beneficio que produce una renta dada, si satisface una preferencia, lo llamaremos bienestar y será el principal interés de esta investigación. Para nosotros, las cuestiones de la justicia económica son aquellas que tienen que ver con la distribución social del bienestar que produce la renta comunitaria.

Ahora bien, la investigación de la justicia distributiva ha solido hacerse en base a la comparación de los distintos fundamentos que a lo largo de la historia de las ideas políticas se han presentado para recomendar determinados estados sociales. Lo tópico es enfrentar entre sí algunas formulaciones de criterios -el de necesidades, merecimientos, eficiencia, derechos, etc.- con el objetivo de proclamar a uno de ellos como vencedor y resolutivo de todos los problemas que se plantean en este ámbito. Del fracaso de tantos esfuerzos por obtener la varita mágica de la justicia distributiva hemos aprendido que cada uno de esos criterios es el fundamento apropiado para un área social determinada. Así sabemos que el criterio de necesidades es el adecuado para el caso de la asistencia hospitalaria; pero que el de merecimientos lo es para el acceso a los puestos de la administración. Ello nos ha permitido centrarnos en otro tipo de problemas: las cuestiones de fundamentación. Porque si es inevitable que haya muchos fundamentos particulares para la división justa, lo que nos conviene averiguar son las características que hacen apropiado considerar un criterio como un fundamento adecuado de justicia distributiva. En concreto, las cuestiones de fundamentación en justicia económica surgen cuando uno se pregunta cómo se logra una decisión social válida sobre las distribuciones alternativas de bienes-tar-renta.

En el párrafo anterior hemos hecho equivaler en tanto que cuestiones de fundamentación aquellas que plantea la identificación de los fundamentos y las que plantea la validez de una decisión social. Esta equivalencia es esencial para nuestra investigación y quisieramos justificarla. En los casos en que se presenta un problema de distribución -económica o no-, lo que aparece inmediatamente es una profusión de argumentos apoyando diferentes criterios de distribución para esos casos. Con cada uno de tales argumentos se propone la validez de un determinado juicio social para las circunstancias concretas del caso. Al final el problema de encontrar las características apropiadas de un criterio para que sea fundamento de una distribución justa consiste en decidir cuál es es juicio socialmente válido para la situación de distribución planteada. De este modo, las cuestiones de fundamentación en justicia económica son aquellas que nos propone la búsqueda de un fundamento para la decisión socialmente justa.

La única exigencia que se impone a ese fundamento es que las decisiones por él fundadas sean "socialmente válidas". Una exigencia así proviene de la propia naturaleza de la cuestión. Lo que tratamos de alcanzar es una distribución que afecta a algunos o a todos los miembros de la sociedad. Los criterios que se aducen como relevantes fundan juicios sociales; juicios que se imponen a todos los afectados por la distribución.

Inexcusablemente la decisión que aquellos apoyan ha de serlo de la sociedad. Por consiguiente, lo que tiene que garantizar el fundamento que buscamos es la validez para la sociedad de la decisión que recomienda. Pero, ¿qué es "válido socialmente"? La validez o invalidez de las decisiones sociales es concedida por el procedimiento por el que se obtienen. Un procedimiento socialmente válido es el que toma en serio a cada uno de los afectados en el litigio de la distribución. Esto significa que ha de satisfacer la condición formal de la justicia de dar a cada uno la consideración que le corresponde. De donde se sigue que la validez social de una decisión se alcanza cuando el fundamento de la misma es también una razón por la que todos los afectados habrían de aceptarla.

En lo dicho ya se intuye lo que van a ser cuestiones de fundamentación en justicia económica. Decisiones, decisiones socialmente válidas, decisiones justas, razones y procedimientos para la decisión, etc., son índices del tipo de problemas que tendremos que abordar. Sucede, sin embargo, que estos problemas no se dan separadamente de una teoría. No existe el problema general de las decisiones justas. Lo que existe es el problema de las decisiones justas para y dentro de una teoría determinada de la justicia -ya sea social, distributiva o económica. Cada teoría se define menos por sus soluciones que por la forma de preguntarse por las cues-

tiones de su ámbito. En consecuencia, también las cuestiones de fundamentación adquieren en ella una especificidad y un orden distintivo que las diferencia de las otras formas que adquieren en el resto de las teorías. Por esta razón, al querer estudiar las cuestiones de fundamentación en justicia económica, la primera tarea es identificar las teorías que se ocupan y se definen por ellas.

Como en la mayoría de los ámbitos de la ética es fácil esta identificación por el recurso de agrupar teorías bajo dos rótulos: teorías teleológicas y teorías deontológicas. En nuestra área de estudio, las teorías sobresalientes en la actualidad son la teoría de la elección social y la teoría de la justicia de Rawls. La primera pertenece a la tradición utilitarista; pero más al utilitarismo de los economistas que al de los filósofos sociales. A su vez la teoría rawlsiana se declara ligada a la tradición contractualista y, en particular, a Kant. No obstante, en ambos casos, el que se diga que pertenece a una u otra corriente histórica sólo es un indicio orientativo. Ninguna de las dos son una manera nueva de decir lo viejo. Cada una de ellas asume y extiende las tradiciones a las que se las une y trata de aprovechar al límite las críticas y aportaciones de la teoría contraria. En definitiva, son teorías muy sofisticadas en las que las cuestiones de fundamentación han adquirido un grado de precisión y de-

licadeza como no pudieron soñar sus predecesores.

En razón de ello hemos tenido que escoger. O tratábamos las cuestiones de fundamentación en la teoría de la elección social, o lo hacíamos en la teoría rawlsiana; ocuparse de ambas hubiera supuesto rebasar los límites concedidos a un trabajo de este tipo. Pues bien, hemos escogido trabajar en la teoría de la elección social. Las razones que nos han inclinado en esta dirección son de varia índole. En primer lugar, nos pareció más interesante tratar de introducir en el ámbito de los estudios filosóficos una teoría proveniente de un campo que nos es usualmente tan ajeno como lo es el de la economía. En este sentido, la teoría de Rawls es mucho más accesible al filósofo y de hecho ya está admitida ampliamente en nuestras universidades. Por otro lado, lo más complicado para nosotros -filósofos- de la obra rawlsiana es precisamente todo el aparato que provienen de la teoría de la decisión -teoría de juegos y teoría de la elección social, particularmente- y que de ninguna manera es un refinamiento del que se pueda prescindir para su comprensión cabal. De aquí que un estudio previo de la teoría de la elección social sea altamente rentable a la hora de enjuiciar a su contrario polémico.

En segundo lugar, nos movió el interés por conocer la opinión de los economistas sobre un tema de justicia que precisamente se adjetiva de "económica". Por la

misma razón anterior, es decir, por la ausencia en nuestro curriculum de estudios que nos acerquen a la ciencia económica, ignoramos todo, o casi todo, de lo que los economistas -qua profesionales- opinan sobre las cuestiones de la justicia económica. Nos interesa saber cómo se plantean ellos ese problema que hemos enunciado como la búsqueda de un fundamento para la decisión justa entre distribuciones alternativas.

Por último, y no obstante lo dicho, hemos de precisar que si bien nos hemos decantado por la teoría de la elección social, la teoría de Rawls no puede y no está ausente de nuestro estudio. Hemos incluido suficientes referencias a su obra como para perfilar los contrastes entre las dos maneras de plantear las cuestiones. Ello lo hemos llevado a cabo guiados no por un prurito académico, sino que su necesidad se nos ha impuesto desde el propio material de la investigación. La comprensión de la propia especificidad de las cuestiones de fundamentación en la teoría que hemos escogido no es completa si dejamos en la oscuridad aquello con lo que contrasta. En este sentido, hemos hecho un esfuerzo para establecer la relación entre los planteamientos últimos de ambas teorías.

Intentaremos ahora explicar de un modo general qué es la teoría de la elección social con el fin de obtener un mejor entendimiento de la discusión particular

que realizamos en los capítulos que siguen. Lo primero que hay que señalar es que no existe un cuerpo homogéneo que pueda ser identificado como la teoría de la elección social. Al igual que toda ciencia interdisciplinar, dicha teoría es un complejo de modelos y enfoques que se sitúan en un espacio de encuentro, pero también de expansión divergente; en el caso de nuestra teoría ese espacio de encuentro se sitúa entre la ciencia política, la economía política, la hacienda pública y la investigación operativa, y desde ese espacio va expandiéndose por otros terrenos de la economía, la política y el derecho. El carácter peculiar y genérico que une a ese complejo de propuestas y estudios es el modo económico de sus investigaciones. La teoría de la elección social se identifica donde quiera que entre por utilizar un instrumental analítico procedente de la economía para abordar sus temas. Más allá de este empeño por constituir una ciencia económica de la política, realmente sólo se puede hablar de tendencias y modelos particulares.

A nosotras, no obstante, nos interesa distinguir dos enfoques que se separan casi desde su nacimiento en ramas divergentes a la teoría de la elección social. Ambas ramas coinciden en ser estudios de situaciones de elección para cuya resolución no es adecuado el procedimiento de mercado. Pero, la interpretación de esas situaciones es muy diferente. Para la teoría positiva

de la elección social tales situaciones han de enfocarse del mismo modo en que las situaciones de mercado son enfocadas por la teoría neoclásica del equilibrio general. Por ello utiliza los mismos supuestos sobre el sujeto de las elecciones - el hombre económico racional-, los procedimientos de elección social funcionan como el mercado a la hora de revelar las preferencias de los individuos y lo que intenta responder en sus investigaciones son las mismas preguntas que se hace la teoría neoclásica para los mercados: ¿cómo se alcanza el equilibrio? ¿Cómo se mantiene? ¿Cuál es la eficiencia de esa situación? El objetivo fundamental de esta rama de la teoría de la elección social es obtener modelos explicativos y predictivos de la conducta de los individuos y de los resultados de la elección. Por eso su campo específico es el ámbito de las situaciones de votación, el estudio de la conducta de los votantes y la adecuación de las distintas reglas de votación para cada situación específica.

Nosotros no nos vamos a ocupar de esta teoría positiva, sino de la teoría normativa de la elección social. Frente al carácter de investigación empírica de los modelos de la teoría positiva, los de la teoría normativa aparecen como investigaciones sobre situaciones ideales. Estos tratan de dilucidar no qué elección se realizaría dada una situación, una conducta racional y una regla de elección determinada; su interés desde el prin-

cipio se ha dirigido a determinar cómo debería realizarse una elección social, bajo qué condiciones normativas habrían de construirse los procedimientos de las situaciones en las que el mercado no es un procedimiento adecuado. La diferencia entre ambas es radical; pero esa diferencia no ha de pensarse como un capricho de los investigadores sociales. El que la teoría normativa dirija sus esfuerzos hacia el ámbito del "deber ser" proviene de una insuficiencia teórica de las predicciones de la teoría positiva puesta de manifiesto con la expansión del estado de bienestar.

La inadecuación del mercado para determinadas situaciones de elección es un resultado que se puso de manifiesto históricamente con el colapso del modelo neoclásico de economía política y la intervención masiva de los gobiernos en las economías nacionales. Sin embargo, la teoría de la elección social no nació al mismo tiempo y con la misma función que otras disciplinas de la ciencia económica (prototípicamente, la econometría) para servir de instrumento al nuevo orden económico que surgió a partir de la crisis del año 29. Los problemas de la elección social no se plantearon al planificador social hasta que la ordenación estatal de la economía no fue un hecho. Porque, ciertamente, las inversiones del estado para superar la crisis económica de entreguerras no obedecían a una intención de satisfacer una preferencia social. Por el contrario, la necesidad de

definir una preferencia social sólo apareció cuando el estado ya había creado tantos bienes sociales --bienes adquiribles por los individuos y no divisibles-- que se hicieron apremiantes los problemas de su producción y distribución.

Teniendo en cuenta este origen, no es extraño que lo que marque distintivamente a la teoría de la elección social frente a otras ramas de la economía (p. ej., la economía del bienestar) es su total compromiso con los procesos de decisión política. Porque, desde el momento en que la atención sanitaria, la educación, el mantenimiento de las ciudades o la defensa nacional son institucionalizados, el problema de la producción y distribución de esos bienes ya no puede dejarse a la competencia --más o menos intervenida por el gobierno-- entre los demandantes de los mismos. Cuánto y cómo se producirán esos bienes, de qué manera se asignarán, bajo qué criterios se redistribuirán entre los miembros de la sociedad son cuestiones estrictamente de decisión política. Si además esa decisión se concibe dentro de una sociedad democrática que reconoce el derecho de los ciudadanos a controlar su destino, tenemos un problema de decisión política muy concreto: ¿cómo saber lo que la sociedad quiere? Esta es la cuestión con la que nace la teoría de la elección social.

Ahora bien, cuando en los años cuarenta se abordó la cuestión de saber lo que la sociedad quiere, el enfoque positivo planteó el problema de la elección como

un problema de mercado para bienes públicos. Los individuos, racionales qua maximizadores, pujan entre sí por esos bienes a través de sus votos y el resultado tiene que determinar un punto óptimo para todos. El problema es que esa conducta no determina un punto óptimo socialmente, sino una línea de puntos óptimos. En concreto, dado que el único juicio de valor que se introduce para juzgar los resultados es el principio de Pareto -un estado social es óptimo paretianamente si nadie pierde y al menos uno gana-, lo que se obtiene no es un sólo resultado que mejora a alguien sin perjudicar al resto de la sociedad; lo que se obtiene es una curva de puntos óptimos de Pareto, ^{donde} cada uno de los puntos depende de las circunstancias particulares de que se parte. La teoría de la elección social positiva puede hacer recomendaciones de cambios de la producción y asignación de bienes sociales, siempre que esos cambios supongan un movimiento de una situación no paretiana hacia una situación de optimalidad paretiana. Pero no puede decir nada sobre estados sociales que sean todos ellos óptimos paretianos.

La teoría normativa trata de producir esa decisión que no alcanza la teoría positiva. Por eso su pregunta ya no es "¿Qué ha de ser elegido?", sino "¿Cómo ha de elegirse?". Ello implica un cambio total de perspectiva sobre lo que es la elección social. Saber qué quiere la sociedad ya no puede ser sencillamente dejar que cada individuo calcule lo que maximiza su interés; aho-

ra se trata de calcular lo que maximiza el interés de la sociedad en conjunto. Elegir un óptimo social consiste en definir un interés general y un fin social. Su preocupación será definir unos valores que representen a la comunidad de electores y utilizará esos valores para concretar una función social a partir de las preferencias individuales. "Cómo ha de elegirse" es un problema que se resuelve con un conjunto de postulados normativos diferentes para cada modelo, pero que se presentan como valores fundamentales de la sociedad.

Si la teoría normativa de la elección social trata de suplir una deficiencia decisional de la teoría positiva, ello lo hace al precio de hacerse acreedora -desde el punto de vista de ésta última- de la objeción de idealidad o, lo que es lo mismo, de inutilidad. En la vida real, las elecciones sociales sobre los bienes públicos y en general sobre el bienestar social se realizan mediante negociaciones de los distintos grupos de presión que tienen poder en la sociedad. Es decir, las decisiones sociales se realizan bajo criterios políticos. Tales criterios no se emplean en razón de su racionalidad social o de su valor ético, sino que dependen más bien de la capacidad de maniobra -de poder- de esos grupos. A estos, en principio, no les importa -aunque lo aduzcan- los intereses iguales de todos los miembros de la sociedad; en la negociación ellos tratan de sacar el mayor provecho para sus propias posicio-

nes, independientemente de cualquier interés más general. La teoría positiva proyecta sus modelos para señalar la conducta más racional a esos intereses de grupo en las situaciones de elección reales. Su reproche, en consecuencia, a la teoría normativa, es el siguiente: ¿qué interés tiene la cuestión meramente teórica de encontrar un modelo normativo para decisiones políticas que, necesariamente, deja a un lado la manera real en que se producen esas decisiones? ¿Qué sentido tiene, en el fondo, preguntarse por lo que desea la sociedad como un conjunto, si ésta no es tal cosa?

No es difícil ver que en este cuestionamiento del enfoque normativo de la elección social hay algo más en juego que la mera cuestión del realismo de unos modelos de análisis. Lo que aquí se plantea no es una cuestión ni trivial ni "meramente" teórica. Porque lo que aquí se presenta crudamente es el divorcio —o, quizás, el antagonismo— entre la realidad y la racionalidad moral de la decisión política. Si hay elecciones —y las hay crecientemente en nuestro mundo— que por su naturaleza tienen que ser hechas colectivamente (porque de otro modo todos los miembros de la sociedad saldrían perjudicados en el sentido que muestra el dilema del prisionero), entonces se necesitan procedimientos para ordenar las preferencias de los individuos de una manera racional para la sociedad. Lo que la teoría normativa trata de hacer para la sociedad es definir

lo que es obvio para el individuo. Un elector individual puede tener un comportamiento racional si, y solo si, tiene unos fines que perseguir. Del mismo modo, para aquellas situaciones en que la sociedad tiene que elegir como un conjunto, si se quiere hablar de elección social, habrá que definir los fines de la sociedad.

La teoría normativa no concibe que se pueda hablar de elección social, si antes no se ha constituido la racionalidad de lo social. La teoría positiva podrá seguir esgrimiendo en su contra la falta de carácter descriptivo y predictivo, la falta de realismo. Podrá incluso decir que la verdadera esencia de la democracia se preserva mejor por los medios actuales de decisión política. Sin embargo, el atractivo poderoso de la teoría normativa -que esas críticas no socavan- reside en señalar las condiciones que deberían regir los procesos de decisión en nuestras sociedades; las garantías racionales y morales de los procesos por los que se nos dan o se nos quitan nuestros derechos como ciudadanos; y, en definitiva, su capacidad para mostrar en negativo las insuficiencias de los procesos políticos en el mundo de hoy.

Nuestro trabajo se dirige, pues, a ofrecer una imagen lo más clara posible de las cuestiones de fundamentación en justicia económica dentro de la teoría de la elección social. (De ahora en adelante, si no especifi-

camos, cuando nos refiramos a la teoría de la elección social ha de entenderse la teoría normativa). Para ello hemos dividido nuestro trabajo en cuatro capítulos. En el primero de ellos nos hemos preocupado de establecer la forma y dimensión precisa de la cuestión general y básica a la que trata de responder la teoría de la elección social. Para establecer su característica especificidad la hemos hecho aparecer en el contexto histórico de las propuestas de la economía del bienestar. La reacción frente a la insuficiencia de esas propuestas marca su nacimiento; pero lo curioso es que en la misma acta se certifican nacimiento y defunción. Nos referimos a la obra de Arrow y a su resultado de imposibilidad por el que se establece que no hay ninguna regla de elección social válida para las sociedades democráticas. Estudiamos la estructura del modelo de Arrow para establecer las líneas básicas que perfilan los puntos de partida de la teoría: la noción de procedimiento, de preferencia, de racionalidad individual y colectiva. En el proceso de explicitación de las dimensiones ética y política de este modelo aparece la manera peculiar de ser utilitarista de la teoría, aunque también sólo de una manera inicial. Por último, confrontamos el modo de ordenar las cuestiones de fundamentación y su significado distintivo en la teoría de la elección social y en la teoría de la justicia de Rawls. En algunos puntos la coincidencia entre ambas se revelará só-

lo aparente y en otros se verá -quizás con sorpresa- que existe mayor acuerdo del que podría preverse.

En el capítulo segundo comenzamos el estudio de un modelo que trata de escapar al resultado de imposibilidad de Arrow. Para ello acude a la tradición utilitarista más pura en busca de una estructura informativa y una racionalidad más adecuada a su forma de ver las elecciones sociales. Así frente al modelo ordinalista simple de Arrow, aquí aparece un modelo cardinalista mucho más sofisticado: el de Harsanyi. Como al de aquel, lo hemos hecho surgir en su contexto teórico: la teoría de la utilidad elaborada modernamente por la teoría de juegos. Su validez como teoría de la elección social depende de la capacidad que tenga para formular diversamente la cuestión básica de Arrow; su especificidad consistirá en ser el más puro defensor de los postulados filosóficos del utilitarismo que renueva de una manera más operativa. Por último, el utilitarismo de Harsanyi es confrontado con el kantismo de Rawls; la polémica nos hará comprender mejor las características profundas y divergentes de ambos procedimientos de fundamentación.

En el capítulo tercero nos ocupamos de un nuevo modelo con el interés puesto en examinar la problemática específica de la racionalidad colectiva en la teoría de la elección social. Aquí también el resultado de imposibilidad de Arrow es examinado en sus determinantes

y, a través de este examen, van apareciendo la noción de racionalidad y la de sujeto racional que subyacen a los modelos tradicionales de las elecciones económicas y lo que se requiere para superarlas en una dirección éticamente valiosa. El modelo de Sen es una propuesta para dotar a las elecciones sociales de una estructura ética que se revela absolutamente necesaria -aunque tal necesidad no suela ser reconocida bien por insuficiencia estructural de los modelos (Arrow), o bien por intención explícita (Harsanyi). El análisis del modelo de Sen nos permite exponer una crítica genérica del modelo consecuencialista de la racionalidad, prevalente en la teoría de la elección social; y una previsión de un nuevo modelo de racionalidad -de fines- que está más próximo a las propuestas de Rawls y Richards que a las tradicionales de la economía neoclásica.

Finalmente, en el capítulo cuarto, nos ocupamos de justicia económica específicamente. Estudiamos por turno las dos tendencias fundamentales que dividen a la teoría en esta materia: la utilitarista y la igualitarista. El problema básico al que se enfrentan es el de la medición normativa de la renta-bienestar de las distribuciones alternativas. Muchos aspectos técnicos de esta medición diferencian y separan a las dos tendencias mencionadas. Pero, veremos que la mayoría de ellos obedecen a una concepción profunda de lo que es la justicia de las relaciones entre los miembros de la sociedad.

Intentaremos describir la naturaleza de estos fundamentos profundos -el individualista y el igualitarista-, aun cuando ello nos ha obligado a salirnos de los límites de la teoría de la elección para acudir a otros ámbitos de la filosofía política en los que se ha explicitado la problemática, los argumentos y la insuficiencia de las concepciones que fundan. Esperamos que ese trasvase de teoría de un ámbito a otro sea más iluminador que ilegítimo.

CAPITULO PRIMERO

ECONOMIA DEL BIENESTAR Y ELECCION SOCIAL

La teoría de la elección social tuvo su origen en una especialidad de la ciencia económica conocida como "economía del bienestar". Esta disciplina se ocupa de los problemas que plantea la definición de la eficiencia tanto en el campo de la producción como en el de la distribución -pero, en éste, más de la distribución funcional que de la proporcional. Cuenta la economía del bienestar con casi cien años de historia y algunos cambios importantes en sus modos de tratar aquellos problemas. Con uno de estos cambios -a principios de los años cincuenta- se producirá una crisis que dará lugar a la teoría de la elección social.

La obra fundacional de la nueva teoría es "Social Choice and Individual Values", de K. J. Arrow(1). En ella se abordan por primera vez de manera explícita los problemas normativos ante los que se detenía la economía del bienestar. Este es su carácter distintivo y gracias a él se clausura la larga etapa en que la economía pensó que todas sus recomendaciones estaban libres de valores. La teoría de la elección social fue la forma en que la ciencia económica tomó en serio su dimensión política.

En este capítulo examinaremos la obra de Arrow. Nuestro objetivo es presentar el mencionado carácter distintivo

de la teoría de la elección social a través de la especificación de su manera peculiar de formular las cuestiones del fundamento de las recomendaciones políticas. El método del que nos serviremos será la exégesis y comparación del modelo de Arrow con aquellos modelos de la economía del bienestar a partir de los que, polémicamente, se definió como especialidad autónoma.

I.1 La pregunta por el fundamento (I).

Si se sigue la historia del pensamiento económico, la obra de Arrow aparece en él como algo raro o sorprendente; y ello por dos razones: el modo en que plantea las cuestiones de la economía del bienestar y por la respuesta alternativa que desarrolla. Lo sorprendente de su formulación consiste en que hace aparecer por primera vez en toda su acuciante crudeza la pregunta por el bienestar como una pregunta por el fundamento de un juicio de valor: ¿cómo son posibles los juicios de bienestar social? Es cierto que la pregunta ya estaba latente en el ambiente de la disciplina desde que A. Bergson (en 1938) se había propuesto "enunciar de una forma precisa los juicios de valor requeridos para la derivación de las condiciones de bienestar económico máximo" (2). Para obtener tal derivación había propuesto una función de bienestar social que otorgaba un ordenamiento social a los ordenamientos individuales de alternativas. La función de bienestar de Bergson se configuraba bajo aquellos juicios de valor que el artículo de Bergson señalaba; pero,

tanto estos juicios de valor como los valores incorporados en los ordenamientos individuales de preferencias habían de ser decididos antes de poner en marcha la función de bienestar social, la cual no contenía ningún mecanismo de control sobre ellos. Por esta razón, durante mucho tiempo no se supo muy bien cómo era posible realizar de una forma fiable el cálculo de dichas funciones, y cuando los teóricos se referían a ella, lo hacían usualmente como a una elegancia formal que coronaba el entramado entero de la teoría del equilibrio general (3).

Pues bien, es Arrow quien se lanza a la empresa de definir la construcción de funciones de bienestar social para ordenaciones individuales de preferencias. Desde las primeras páginas de su obra ya está planteada la cuestión sin ambages: "(...) nos preguntamos si es formalmente posible construir un procedimiento para pasar de un conjunto de preferencias individuales conocidas a un modelo de formación de decisiones sociales, satisfaciendo el procedimiento en cuestión ciertas condiciones naturales" (4). Se muestra en ello la concepción de la fundamentación propia de la teoría de la elección social; los juicios sociales sobre el bienestar social se fundan en el terreno de la intersubjetividad. Hay que partir de los individuos que componen la comunidad -de sus gustos, demandas, valores- y luego tratar de establecer -a partir de ellos- una preferencia social. Tal modo de fundamentación tiene ya históricamente probado cuál es su obstáculo: ¿cómo partiendo de lo más subjetivo se puede alcanzar lo más intersubjetivo?

La naturaleza específica de este obstáculo en el área

de la economía era lo que se había puesto de manifiesto en el artículo de Bergson. En la medida en que la ordenación social se configura bajo valores subjetivos que no son controlados por la función de bienestar social, si esos valores cambian o son inconsistentes entre sí, determinarían juicios sociales diferentes y no hay manera de establecer la corrección de ninguno de ellos. La función de bienestar social de Bergson está aquejada de un relativismo que la inhabilita desde el punto de vista de la fundamentación.

Arrow comienza señalando este problema: hay tantas funciones de bienestar social como configuraciones de preferencias entre los individuos de la comunidad. Sin embargo, ello no quiere decir que sea un mal procedimiento; sólo significa que es todavía muy tosco. La cuestión a la que se enfrenta Arrow es precisamente la de cómo habrá de ser perfeccionado. Hay que buscar la forma de establecer unas condiciones que -pudiendo ser exigidas a las funciones de bienestar social- formulen los valores consistentes y permanentes de una comunidad.

De este modo, la cuestión del fundamento "¿Cómo son posibles los juicios de bienestar social?" se resuelve en una pregunta por el procedimiento para conformar correctamente -es decir, con validez intersubjetiva- funciones de bienestar social del tipo de Bergson.

I.2 Orígenes de la teoría de la elección social: los problemas de optimización.

El modo arroviano de plantear el tema del fundamento clarificaba -al mismo tiempo que daba un giro- las discusiones sobre el bienestar económico que se habían suscitado a partir de la "revolución marginalista" de los economistas neoclásicos. Básicamente significaba que no se podía abordar la solución de ningún problema de la economía del bienestar si no se especifica consistentemente la forma de la función de bienestar social; es decir, los valores requeridos para la construcción de un juicio de valor social. Para entender la relevancia de tal exigencia fundacional es conveniente que consideremos los orígenes y los términos en que se plantea la cuestión del fundamento en la economía del bienestar.

El tema propio de esta disciplina aparece por primera vez en toda su problematicidad en la historia del pensamiento económico con el intento de justificación -por parte de Walras y Pareto- de la libre competencia sobre la base de que ésta tiene como propiedad inherente la optimalidad, i. e., un máximo de utilidad para las partes competentes.

Su punto de partida es el polémico supuesto -polémico para la tradición utilitarista de la economía neoclásica- de la imposibilidad de hacer comparaciones interpersonales de utilidad en razón de que "las cantidades de satisfacción subjetivas son inconmensurables por naturaleza" (5). Este supuesto obligaba a abordar el tema del bienestar económico

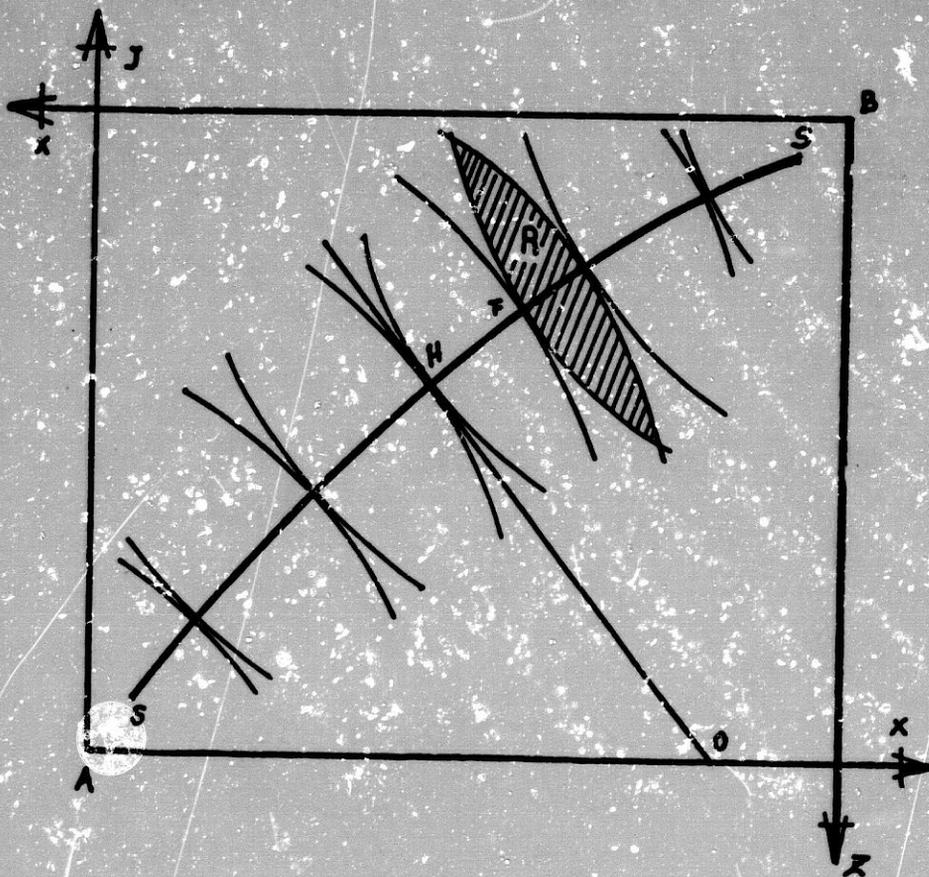
como un problema de asignación de recursos -cualquiera que fuera el objetivo de la misma: la producción óptima, la distribución óptima o, simplemente, la satisfacción óptima. Para Walras la asignación mejor de los recursos se alcanza, en el proceso de libre cambio, en el punto de equilibrio competitivo (6); así, por ejemplo, en el caso de los consumidores, este punto representa la máxima satisfacción, puesto que en él el precio de los productos es igual a su utilidad marginal. No obstante, esto es especialmente válido para los temas de producción; el propio Walras restringió la aplicación del principio de libre competencia -ya que principio de importancia práctica tenía en su sistema- con dos observaciones:

- 1) Sólo se aplica a la producción de objetos de demanda privada; y
- 2) La cuestión de la distribución tiene que decidirse por otros medios (7).

Por su parte Pareto concretó la idea de Walras y legó el concepto más frecuentemente utilizado de óptimo. Dada la imposibilidad de las comparaciones interpersonales de utilidad, el propio concepto de medición de la satisfacción personal por la utilidad queda menoscabado; de ahí que Pareto utilizara las curvas de indiferencia de Edgeworth en donde "una asignación cualquiera se define por los niveles específicos de consumo de cada consumidor y los niveles de inputs de cada producto" (8). A partir de aquí se explicita el concepto de óptimo paretiano: una situación de intercambio es un óptimo paretiano, si no es realizable otra situación de intercambio en que todos los miembros de la sociedad tengan el mismo nivel de utilidad y al menos uno tenga

un nivel mayor, todo ello medido en términos de curvas de indiferencia.

El óptimo de Pareto define un punto de equilibrio, de racionalidad y de eficiencia económicas sobre los que se produjo una vasta literatura en los años treinta (9). Al mismo tiempo, sin embargo, se vinieron repitiendo las objeciones básicas que ya anteriormente formularon Wicksell y Marshall; el óptimo de Pareto es un punto de equilibrio relativo: "(...) esto es, relativo a una gama de posiciones vecinas y a la situación inicial de la que parte el comercio, en especial de la distribución inicial de los artículos entre los individuos" (10). Esto es lo que se muestra utilizando una caja de contratos de Edgeworth como la siguiente (11):



Esta caja representa las curvas de indiferencia de dos individuos, A y B, para dos bienes, x e y , distribuidos entre ellos. Los puntos de tangencia de las curvas de indiferencia de ambos individuos representan puntos en los que las utilidades marginales de x e y son iguales para ambos. La curva SS que une todos esos puntos es la curva de contrato de Edgeworth y refleja puntos de intercambio igualmente beneficiosos para ambos (un movimiento fuera de esa curva significaría que uno pierde para que el otro gane). Pero, cada uno de esos puntos se produce para una distribución dada -así el punto H es el óptimo para una distribución O-; depende, pues, de qué distribución partamos para que obtengamos un óptimo u otro. Obviamente, el criterio de Pareto no nos puede ayudar a elegir entre todos estos óptimos paretianos de la curva de contrato -por ejemplo, entre H o F. Más aún, ni siquiera permite la comparación entre una posición optimal y una sub-optimal (la superficie rayada), pudiéndose dar el caso de que una situación optimal de una mala distribución de renta (por ejemplo, H -optimal) sea peor que una sub-optimal de una buena distribución de renta (por ejemplo, R sub-optimal) y, sin embargo, el criterio de Pareto no tiene los medios necesarios para evaluar el cambio (12).

Con este criterio de evaluación no podemos saber si los resultados, en el orden del bienestar social, son satisfactorios; una sociedad, como vemos, puede ser óptima paretianamente y padecer el mayor desequilibrio entre ricos y pobres. Sólo es capaz de indicar que, para una distribución dada -que el criterio no califica-, ya no se puede hacer nada sin perjudicar a alguien.

Los problemas del óptimo van a ser la cruz de la economía del bienestar en las décadas siguientes. En primer lugar, el modelo supone que la comunidad a la que se aplica es inmutable; no cambia ni en número de individuos, ni estos nacen ni mueren, ni sus gustos cambian, ni están influidos socialmente, etc.. Si, mientras se está decidiendo qué posición es óptima, el individuo que mejora cambia sus preferencias, el resultado paretiano sería incorrecto (12).

En segundo lugar, no parece posible que se puedan separar las cuestiones referentes a la producción y el consumo de las de la distribución. La economía del bienestar se ha empeñado en tomar la distribución de la renta como un dato y en buscar condiciones para posiciones óptimas en la sociedad. Sin embargo, es demasiado obvio que la distribución no se puede reducir a una constante exógena que se introduce en los cálculos; más bien es una magnitud interrelacionada con las otras y que, por consecuencia, produce cambiantes percepciones del bienestar que el modelo tiene que considerar cuidadosamente. Como hemos visto, la inoperatividad del criterio de Pareto proviene en una buena parte de que no es capaz de tener en cuenta que cualquier desplazamiento en la curva de contrato modifica la estructura de la distribución (13). Sin embargo, durante décadas se ha mantenido como un dogma de fe que la cuestión del bienestar social había de plantearse exclusivamente en el terreno de la producción y el comercio; las cuestiones de distribución eran tenidas por cuestiones de ética -por tanto, no científicas- ajenas a la economía.

El planteamiento más acabado de este dualismo producción-distribución, dentro de la economía del bienestar, lo encontramos en Pigou. Su objetivo no fué precisar un óptimo, sino que se interesó en establecer un procedimiento para dilucidar cuándo se puede hablar de un aumento o mejora en el bienestar económico. De modo general puede decirse que, para él, cualquier situación en la que se aumente el ingreso nacional -y donde el ingreso nacional sea una magnitud independiente de su distribución-, aumenta el bienestar económico. Se abría con ello el problema de la medición del producto total y la teoría del número índice que, definida en palabras de Dobb, consiste en saber "cuánto se puede inferir sobre los cambios en la producción total o ingreso nacional -medidos en términos de satisfacción o utilidad del consumidor- a partir de datos de mercado acerca de precios y cantidades" que son recogidos en los números índice (14). Para disminuir las perturbaciones en la medición del ingreso nacional (congénitas a este compuesto heterogéneo) se propone Pigou estudiar la parte del bienestar que puede medirse en dinero de forma que la intensidad del deseo de una persona se expresa en el dinero que está dispuesta a ofrecer por un bien.

En este punto es importante señalar que los trabajos de Pigou tratan de dejar fuera todo tipo de juicios de valor; su intento se dirige solo a las causas y efectos de los cambios en el bienestar que supone definible por los citados indicadores. Cualquier ingerencia en el terreno de los hechos científicamente establecidos de algo -las valoraciones- no controlable por el método es proscrito como una distorsión

que volvería la tarea imposible; los problemas de distribución son tratados separadamente y desde supuestos muy rígidos. En este sentido el método de Pigou para la medición de la renta nacional es especialmente interesante porque presenta explícitamente los presupuestos que han estado implícitos en las investigaciones de la economía neoclásica y que seguiran teniendo vigencia (más o menos matizados) en la economía del bienestar e incluso en algunos teóricos de la elección social. En general se trata el caso de la medición de la renta como si se tratara de un individuo representativo, para lo cual es necesario suponer:

- 1) La distribución del ingreso entre los diversos individuos es constante.
- 2) El sistema básico de necesidades y preferencias está dado.

Como en el caso de Pareto, esto comporta una comunidad y unos individuos inmutables. Esta implicación es la que hace imposible la medición de la renta por números índice. Una producción puede ser considerada mayor o menor según los gustos o la distribución de un periodo; por ello dos periodos no son comparables entre sí y, finalmente, no permite hablar de aumento o disminución de la renta nacional.

Durante dos décadas se estuvo investigando en las direcciones indicadas por Pigou. Por el lado de la producción, se buscó establecer las condiciones del óptimo desglosando los supuestos pigouvianos y se discutió sobre las posibilidades de los números índice como medidas para la renta. Las cuestiones de distribución se desarrollaron sobre la base de la utilidad marginal decreciente del dinero y sobre

la igual capacidad humana para la satisfacción. Se puso de nuevo en vigor alguna forma de comparabilidad interpersonal y de cardinalidad de las satisfacciones personales. Cuando en la década de los treinta el escepticismo cundió y se rechazó la intercomparabilidad, la única propuesta renovadora sobre la cuestión abierta del bienestar la dieron los teóricos de la llamada "New Welfare Economics".

Examinaremos las propuestas de la nueva economía del bienestar en un contexto diferente y más enriquecedor para la comprensión de los planteamientos iniciales de la teoría de la elección social. Ahora conviene prestar atención al artículo auroral de Bergson de 1938. Porque, ciertamente, en la determinación del óptimo el punto de inflexión lo marca Bergson al proponer una función de bienestar social para determinar una especie de curva de indiferencia para la sociedad.

Con la función de bienestar social se trata de establecer las condiciones necesarias para la obtención de un "optimum inter optima" social (15). En concreto, se trata de discriminar entre los diferentes óptimos relativos aquel que socialmente es juzgado mejor. Por primera vez en la literatura económica el bienestar social es un valor que está determinado unívocamente por las unidades de consumo -ya sean individuos o familias. Pero, el interés que para nosotros tiene Bergson reside en que explicita que la construcción de ese juicio de valor que es la función de bienestar social sólo es posible bajo supuestos asimismo valorativos. En primer lugar, si el bienestar social es una función creciente de los bienestares individuales de las unidades de

consumo, hemos de determinar previamente quiénes son las unidades de consumo que va a tener en cuenta la función; ello supone una valoración que separe los individuos que "valen" de los que no "valen". En segundo lugar, para saber qué es "estar mejor" o "estar peor" se requiere asumir un juicio de valor que pondere los niveles de utilidad de las unidades de consumo -para ello Bergson utiliza el criterio paretiano. Por último, es necesario otro supuesto valorativo para establecer la forma de la función de bienestar social -¿utilizaremos la suma de las utilidades, o sus productos, o el producto de sus logaritmos, etc.?-. Una vez adoptados y concretados estos supuestos valorativos, se puede diseñar una función de bienestar social que representa las preferencias sociales en base a los índices de utilidad de las unidades de consumo. La forma genérica de la función es:

$$W = W(U_1, U_2, \dots, U_n)$$

donde las situaciones alternativas se enjuician por los valores que determinan en esa función las utilidades (U) de las unidades de consumo de 1 a n (16).

La cuestión que el planteamiento de Bergson deja abierta (por la que se granjeó durante dos décadas las reticencias de los economistas) y que inaugura el campo propio de la teoría de la elección social~~es~~: ¿cómo se eligen los valores que definen la función de bienestar social? Lo radical del planteamiento de Bergson es que explicita -sin resolver- la cuestión de los valores en la definición del bienestar social. Desde entonces quien tome en serio esta línea de trabajo ya no podrá seguir recurriendo al libre mercado y al equilibrio

competitivo como decisor imparcial. Con la función de bienestar social aparece la figura del individuo que toma decisiones en nombre de la sociedad y que tiene que responsabilizarse de las valoraciones con las que opera ante esa misma sociedad. No obstante, Bergson todavía no es un teórico de la elección social. Su formulación deja la cuestión de los valores fuera del campo de la investigación. Para él, son otros (los políticos, los observadores imparciales, los decisores omniscientes) los que tiene que hacerse cargo por su cuenta y riesgo de decidir esos valores; el status científico de la economía exime al economista de abordar ese tema.

Consecuentemente, el problema de la elección social como problema de una teoría que se desgaja -o es repudiada, según quien cuente la historia- de la economía del bienestar comienza cuando los economistas comprenden que es inexcusable que sus propuestas contengan un procedimiento para controlar los valores que las determinan. Por eso es Arrow (y no Bergson) quien inaugura la nueva teoría que comienza cuando se asume que ya no es posible excluir la ética: "Yo deseo destacar ahora solamente que debemos contemplar el sistema total de valores, incluidos los valores sobre estos valores, al buscar una teoría realmente general del bienestar social"(17). Esta es la tarea que se presenta ahora como la más urgente, puesto que en ella se concentran desde la cuestión de la racionalidad de las propuestas económicas hasta la posibilidad o imposibilidad de una regla -o esquema- de convivencia: la paz social misma.

I.3 La pregunta por el procedimiento de fundamentación.

La pregunta inicial "¿Cómo son posibles los juicios de bienestar social?" se nos ha transformado en una pregunta por el procedimiento de fundamentación. Arrow asume la tarea de investigar las condiciones de formación de las funciones de bienestar social de Bergson. En ello se concentra el trabajo inicial de la teoría de la elección social, puesto que ello supone resolver el obstáculo que todo modelo que parte de los valores individuales enfrenta: alcanzar la intersubjetividad. Lo individual, para esta concepción de la fundamentación, sólo puede ser intersubjetivo -un valor válido socialmente- si dispone de un mecanismo que filtre la "mala" subjetividad y nos entregue lo común y aceptable de esa subjetividad. La busqueda de ese mecanismo es lo fundamental; y en la teoría de la elección social se concreta en la búsqueda de un procedimiento para asignar funciones de bienestar de Bergson a las preferencias individuales sobre alternativas dadas. A tal procedimiento lo llama Arrow en la primera edición de "Social Choice" FUNCION DE BIENESTAR SOCIAL, pero en el "Apéndice" de 1963 y en sus obras posteriores acepta la sugerencia de Kemp y Asimakopulos (18) y lo denomina "constitución" a efectos de evitar confusiones terminológicas. Esta será la denominación que utilizaremos en lo que sigue.

Para abordar esta tarea Arrow reformula el lenguaje de la teoría del consumidor en el que Bergson presentó su función de bienestar social siguiendo la notación de la 16-

gica formal (19); por ello en lugar de hablar de utilidades habla de preferencias y de ordenamientos de preferencias. De este modo, lo que tenemos como argumento directo de la función de bienestar social no son utilidades individuales, sino ordenaciones individuales de preferencia. Por otra parte, señala Arrow, si lo que está en juego es un procedimiento para concretar funciones de bienestar social, lo más conveniente es que éste se defina no para situaciones sociales concretas, sino para todas las situaciones posibles para una sociedad. El conjunto de situaciones alternativas es llamado por Arrow "entorno"; y la constitución es definida como una regla que determina un ordenamiento de situaciones sociales para todo posible entorno según una preferencia social.

La cuestión ahora es averiguar qué tipo de procedimiento ha de ser una constitución, si sobre ella han de descansar (qua fundamento) los juicios de bienestar de la sociedad. Arrow comienza determinando los métodos de elección social (20); según esa clasificación, las decisiones pueden ser tomadas:

- (A) Por individuos o grupos.
- (B) En base a normas consuetudinarias.
- (C) Por mecanismos de voto.
- (D) Por mecanismos de mercado.

Es obvio que donde prevalecen los métodos de decisión (A) y (B) no nos hallamos en un sistema democrático. Sin embargo, esos métodos tienen una ventaja sobre los métodos (C) y (D), propios de las democracias: su precisión. Donde hay una sola voluntad que decide, o una voluntad común,

o una voluntad divina, no se plantea el problema básico que las sociedades democráticas tienen: el conflicto de voluntades. Pues bien, el procedimiento que estamos buscando tiene que ser tal que resuelva ese conflicto, que aúne en una voluntad común -o, al menos, cohesione- las voluntades contrapuestas. De ahí el problema básico que Arrow enuncia en la portada de su obra: ¿puede haber un método para las sociedades democráticas que tenga la misma precisión y consistencia que los métodos de las sociedades dictatoriales o tradicionalistas?

Ahora bien, Arrow es un economista y tratará este problema no sustantivamente, sino formalmente -si bien no es parco en explicitar las implicaciones sustantivas que el tratamiento formal conlleva. De aquí que la formulación de la cuestión del procedimiento de fundamentación se reescriba como la búsqueda y diseño de un procedimiento formal para pasar de las preferencias individuales a las sociales.

I.3.1 Dos procedimientos alternativos: cálculo hedonista y principios de compensación.

En este punto hemos de echar de nuevo la mirada atrás en el tiempo. En el momento en que Arrow escribe el libro del que nos estamos ocupando la economía del bienestar tenía dos modos alternativos de resolver la cuestión del procedimiento. El de más larga tradición era el método de la agregación de las utilidades individuales. La formulación ori-

ginal es de J. Bentham y está diseñado para hacer eficiente la máxima utilitarista. En este procedimiento -conocido por "cálculo hedonista"- la decisión a tomar es aquella que produce la mayor satisfacción al mayor número de personas, lo cual se determina simplemente sumando las satisfacciones de todos los individuos implicados. Ello exige informarse sobre las necesidades e intereses de los individuos para calibrar la gratificación o frustración que un curso de acción les provocaría. Determinar esa gratificación y frustración requiere estipular criterios y este es el sentido de la lista de siete propiedades con las que se cree poder medirlas. El procedimiento es, pues, simple; de acuerdo con esos criterios se pondera el placer y el displacer de un individuo -restando del conjunto de gratificaciones el conjunto de frustraciones-, obteniéndose el balance total individual de satisfacción; sumando a continuación los balances de todos los individuos, obtenemos el balance de gratificación colectivo (21). Por último se comparan los balances de gratificación colectiva de cada una de las situaciones sociales alternativas y se escoge la que mayor balance contenga.

Ante la facilidad con que se presenta el procedimiento, la pregunta que primero se viene a la mente es: "Pero, ¿cómo se hace todo eso?". El procedimiento, ciertamente, exige que demasiadas cosas en la esfera de la subjetividad estén claras, sean además medibles y, finalmente, comparables entre sí y entre los sujetos. La polémica sobre la mensurabilidad de las utilidades individuales, sin embargo, no desanimó a los economistas neoclásicos que tomaron su defensa

a su cargo. Para la teoría del valor-utilidad era esencial refinar la idea de Bentham a fin de hacerla operativa; así debilitaron su exagerada confianza en la introspección, pero no pudieron evitar que la polémica continuara (22). La economía del bienestar, aunque a veces esté apelando a ella de una forma débil, trata de buscar un procedimiento que la haga innecesaria -fue el camino de Pareto. Por eso Arrow la despachará rápidamente diciendo que no hay nada de lo que se trata de explicar con la mensurabilidad de la utilidad que no se explique por curvas de indiferencia y, consecuentemente, "carece de significación alguna relevante para la economía del bienestar" (23). Ya veremos en el capítulo II que este juicio no es muy afortunado.

Alternativamente, la nueva economía del bienestar había elaborado otro procedimiento para fundamentar juicios de bienestar. Se trata de las diversas interpretaciones de los principios de compensación que en los años treinta presentaron Kaldor, Hicks y Scitovsky. Entre estos principios y el cálculo de utilidades media el descrédito levantado por L. Robbins contra las comparaciones interpersonales de utilidad y, en general, contra todo juicio de valor en economía. La nueva economía del bienestar (más que ningún otro modelo) intentará la construcción de un entramado teórico de impecable pulcritud científica; i. e., libre de valores.

Vimos en Pigou que toda posibilidad de comparar utilidades reposa sobre el supuesto de "igual capacidad para la satisfacción" de los seres humanos. En la posición de Pigou -y en muchas otras antes y después de él- sobre la dis-

tribución este supuesto era piedra angular hasta el punto de que sin él no se podía ni empezar a hablar sobre estados de bienestar (24). L. Robbins, rescatando la corriente escéptica de Pareto sobre las utilidades, expuso de forma persuasiva que el postulado de "igual capacidad" era un juicio de valor infundamentado que pertenecía a ese tipo de afirmaciones incapaces de demostración que deben ser excluidas de un "corpus" científico positivo como quiere ser el de la economía (25). El estado en que este ataque a los valores dejaba la capacidad de aconsejar del economista era lamentable. Robbins perseguía este fin intencionalmente. Sobre aquella argumentación defendió que la economía -como cualquier ciencia- sólo ha de informarnos sobre los medios, puesto que no puede pronunciarse sobre qué sea "mejor"; no puede -si es ciencia- decirnos qué situación económica alternativa es preferible. Sólo nos puede dar información acerca de las consecuencias de escogerla: "Frente al problema de la decisión entre esto y aquello, la economía no puede ayudarnos a tomar nuestra última decisión" (26). Tal capacidad judicatoria tendría que estar basada en la posesión de valores que nos permitirían comparar situaciones, preferencias, satisfacciones, etc.; mas, los valores no son ciencia y no tienen lugar alguno en la economía, ciencia "wertfrei" de los medios racionales (27).

N. Kaldor (28) se propuso encontrar un procedimiento que, por ser independiente de los juicios de valor, tuviera una condición científica y, a la vez, sirviera para saber si un cambio económico supone una mejora o no. El principio de compensación hipotética procede de una idea muy simple

y plausible. Una política es recomendable si el cambio que produce permite que los individuos perjudicados por él sean compensados por los que resulten beneficiados. Más precisamente: un estado A es socialmente preferible a un estado B, si aquellos que ganan con el estado A pueden compensar a los que pierden para que acepten el estado A y permanecer en un estado mejor del que disfrutaban en B. Como se ve la rai-gambre paretiana del principio es innegable. Toda política que incremente la productividad -y, por tanto, el ingreso real- es recomendable ya que conlleva "que todo el mundo esté mejor, sin que nadie esté peor" (29). En ello también se observa que el procedimiento de la compensación sólo está dirigido a las cuestiones de eficiencia: una situación es mejor que otra, si hay más ganancias que pérdidas. La cuestión de la distribución de las pérdidas y ganancias es algo que, en cumplimiento estricto del mandato de Robbins, se extirpa de la consideración del científico, arrojándola al infierno de la irracionalidad (30).

A J. R. Hicks (31) le fascinó la idea en la medida en que proporcionaba una base segura para definir la organización óptima del sistema económico. Si se cumplen las condiciones "compensatorias", un sistema está en su punto de eficiencia y el economista puede recomendar movimientos hacia tal punto. No obstante, la formulación suya del principio de compensación es algo diferente de la de Kaldor: un estado A es socialmente preferible a un estado B, si aquellos que perderían con el cambio a A no pueden compensar a los que ganarían con él para que no realicen el cambio. Se supone que si no pueden compensarlos -sobornarlos- es porque

los ganadores ganan más con el cambio de lo que ganarían con la compensación y, por tanto, es un estado que mejora a la sociedad globalmente.

El optimismo que produjeron los artículos de Kaldor y Hicks se vió mitigado cuando T. Scitovsky (32) demostró que el criterio de compensación era ineficiente en la medida en que servía para probar que un estado A es mejor que otro B y, una vez adoptado el estado A, con el mismo criterio podía probarse que B era mejor que A. Esta situación se producía aplicando en el primer caso el criterio según la formulación de Kaldor y en el segundo el criterio según la formulación de Hicks -o viceversa. No obstante, el propio Scitovsky proporcionaba la salida a la contradicción. El estado A es socialmente preferible -en la formulación de este autor-, si los ganadores pueden sobornar a los perdedores para que acepten el cambio, pero los perdedores no pueden sobornar a los ganadores para que no lo hagan. De este modo, si la aplicación del principio señala a un estado social como preferible, puede también asegurarse la eficiencia del mismo (33).

Arrow considera ampliamente los principios de compensación en su obra (34), defendiéndolos de algunas críticas de Baumol (35). No obstante, como a otros muchos autores, a Arrow no le satisfacía ese empecinamiento que tiene la economía del bienestar en separar los aspectos de producción de los de distribución. Cualquier elección, sostiene Arrow, se hace en base a ambos aspectos; y, además, ha de obligatoriamente hacerse así, puesto que el no tener en cuenta la relevancia de la distribución vacía de utilidad a la ordena-

ción de situaciones sociales que resulte; esto es, vacía de operatividad al juicio de bienestar social que el criterio produzca.

Ahora bien, la crítica más radical que Arrow dirige a los principios de compensación es que, considerados como una regla de elección social, no cumplen las condiciones naturales y razonables requeridas para que puedan ofrecer juicios sociales satisfactorios. Este es el núcleo del análisis de Arrow: la investigación de las condiciones de racionalidad de las elecciones sociales. Son ellas las que determinan la forma de la función de bienestar social y de ellas depende, en último término, la posibilidad de los juicios de bienestar social.

1.3.2 La racionalidad de las elecciones.

Tratar de definir los valores que intervienen en la determinación de juicios de valor social conlleva la tarea primaria de determinar en qué consiste la racionalidad de las elecciones colectivas o, más genéricamente, la racionalidad colectiva. La teoría de la demanda ofrece un modelo muy claro de lo que es la racionalidad de las decisiones individuales. El problema que se plantea es averiguar si la elección colectiva no tiene otras condiciones de racionalidad que las conocidas para la elección individual, o bien tiene unas condiciones específicas. Resolver este problema en un sentido u otro marcará la característica distintiva de los diferentes modelos de la teoría de la elección

social (36).

Directamente ligado con la consideración utilitarista de que la norma de conducta racional consiste en "maximizar el interés", el concepto formal de racionalidad que la economía ha elaborado exige que los sujetos de las acciones económicas muestren una conducta "consistente". La conducta de un individuo es consistente -y, por tanto, racional-, si siempre que se le presente un mismo tipo de alternativas ofrece el mismo tipo de elección (37). Tal noción supone que el decisor es capaz de ordenar sus preferencias de acuerdo con ciertas propiedades que aseguren esa consistencia y, notoriamente, según la utilidad que le proporcionan las alternativas. De este modo, en la teoría del consumo aparecen las funciones de utilidad como expresiones de un sistema numérico de preferencia de manera que, dados los lotes de bienes x_1 y x_2 a los que se asigna un número arbitrario de utilidad tal que $U_i(x_1)$ y $U_i(x_2)$ -pero de acuerdo con la curva de indiferencia del individuo i para ese par de lotes- se puede decir

$$U_i(x_1) \geq U_i(x_2) \quad ,$$

esto es, que el individuo i considera al lote x_1 al menos tan bueno como el lote alternativo x_2 . Lo cual, ya en terminología empleada por primera vez por Arrow, se puede expresar en una relación binaria de preferencia

$$x_1 R_i x_2$$

que puede descomponerse en la relación de preferencia estricta, P_i , y de indiferencia, I_i . De esta manera se obtie-

nen los elementos básicos de la estructura preferencial del individuo racional (38).

Con estos elementos se puede definir un ordenamiento de preferencias como una relación preferencial, dentro del conjunto de alternativas presupuestarias (S) de un consumidor, que cumple las propiedades:

- (A) Reflexividad: Para todas las alternativas presentes en un conjunto S se da el caso $x R_i x$; es decir, que x sea considerado al menos tan bueno como el mismo.
- (B) Orden completo: Para todos los pares de alternativas distintas (x_1, x_2) incluidos en S, ya sean $x_1 R_i x_2$ o $x_2 R_i x_1$, o ambas, las relaciones no definidas entre pares pertenecientes a S están excluidas.
- (C) Transitividad: Para todos los casos de comparaciones triples (x_1, x_2, x_3) contenidas en S, si $x_1 R_i x_2$, y $x_2 R_i x_3$, podemos decir también $x_1 R_i x_3$.

Lo que estas propiedades indican son condiciones de consistencia y su función consiste en asegurar que la elección basada en los ordenamientos de preferencias sea racional en el sentido que ellas definen. No obstante, de las tres, sólo la transitividad es una condición estricta de racionalidad, mientras que la reflexividad es meramente una condición de sentido común y la condición de orden completo es un presupuesto necesario a la teoría que asegura que el individuo siempre conoce su decisión cuando se presenta un

grupo de alternativas (39).

El problema, como dijimos, es averiguar si la elección colectiva basada en ordenamientos individuales de preferencias tiene la misma racionalidad que exhiben las elecciones individuales. Si así fuera, el modo natural de comprobarlo sería hacer una trasposición del procedimiento individual al procedimiento social; tratar a la sociedad como si de un solo individuo se formase, según el método de decisión de la regla de la mayoría. Pero, lo que demuestra la paradoja de la votación -con la que Arrow comienza su libro (40)- es que tal cosa no es viable. Tómense tres estados sociales, x , y , z , y tres individuos a , b , c ; supóngase que cada uno de estos individuos tiene ordenadas sus preferencias según se ha definido de tal forma que

$$x P_a y P_a z$$

$$y P_b z P_b x$$

$$z P_c x P_c y$$

Si cada miembro tiene derecho a un solo voto y las decisiones se llevan a cabo sobre cada par de alternativas comparadas, obtendríamos:

$$x P y ,$$

$$y P z ,$$

$$z P x .$$

Consecuentemente, un ciclo de preferencias del tipo

$$x P y P z P x$$

que no tiene la propiedad de racionalidad de la transitividad, i. e., no presenta ninguna alternativa preferida

a cualquier otra.

Lo que este resultado pone de manifiesto es que la elección social requiere algo más que la consistencia para ser racional. O, de otro modo, que la racionalidad colectiva es de una naturaleza diferente de la racionalidad individual. En razón de este resultado, de nuevo nos encontramos ante la pregunta por el procedimiento, pero ahora convertida en una pregunta por las condiciones de racionalidad de las constituciones: "¿Pueden encontrarse otros métodos de agregación de los gustos individuales que impliquen un comportamiento racional de la comunidad y que sean satisfactorios en otros sentidos? " (41).

I. 3.3 Las condiciones de racionalidad de las elecciones sociales.

Arrow define la elección -en la notación ya presentada- en términos de la relación de orden R , para un conjunto de alternativas S (un entorno), como el subconjunto de S , $C(S)$, de todas las alternativas x de S tales que para cada y de S , $x R y$. Definir de este modo la elección en términos de relaciones de orden tiene notables ventajas con respecto a los sistemas utilizados tradicionalmente: los mapas de indiferencia y las funciones de utilidad. Con respecto a los mapas de indiferencia, su ventaja fundamental es que permiten operar mejor con situaciones sociales, dado que admiten fácilmente la multitud de componentes de éstas. En cuanto a las funciones de utilidad, señala Arrow que lo importante de su uso -la representación indirecta

de la ordinalidad de preferencias- las relaciones de orden lo realizan directamente (42).

Además de estas diferencias de método, la teoría de la elección social tiene necesariamente que diferenciarse de la teoría de la decisión individual puesto que sus objetos no son intercambiables en los modelos formales. Ciertamente, las cestas de bienes entre las que elige el consumidor individual no pueden ser tratadas como las situaciones sociales entre las que tiene que elegir la sociedad, dada la variedad de componentes heterogéneos que incluyen estas últimas. La definición precisa que da Arrow de "situación social" es la siguiente:

"Una descripción completa de la cantidad de cada tipo de bienes en manos de cada individuo, la cantidad de trabajo a ser aportado por cada individuo, la cantidad de cada factor de producción invertido en cada tipo de actividad productiva y las cantidades de los diversos tipos de actividades colectivas, tales como servicios municipales, actividades diplomáticas y "su continuación por otros medios" y la erección de estatuas a los hombres famosos" (43).

Lo que esta descripción trata de abarcar es el conjunto de caracteres económico sociales que reúnen dos cualidades específicas:

- (a) Están sujetos a decisión social.
- (b) Algún individuo podría no ser indiferente ante una variación en una de ellos.

Pero, en esta definición hay algo casi más importante implícito. A diferencia de las elecciones individuales, en

las elecciones sociales se puede suponer legítimamente que la decisión del individuo no está guiada por la ganancia personal exclusivamente; hay un lugar para criterios diferentes que son relevantes. En este sentido señala Arrow que si bien estos criterios no son fácilmente distinguibles, sería conveniente separar las ordenaciones de preferencias en dos grandes grupos:

- (i) Las basadas en gustos, i. e., en criterios de consumo directo del individuo.
- (ii) Las basadas en valores, i. e., en criterios generales como el de equidad o el de "emulación pecuniaria" (44).

Arrow urge, a pesar de la falta de nitidez de esta distinción, a decidirse sobre el tema, puesto que sólo las ordenaciones basadas en valores pueden tener en cuenta "todos los deseos de los individuos, incluyendo los altamente importantes deseos socializantes" (45), exigencia inexcusable para la formación de una ordenación social en la que se juega el bienestar de toda una comunidad. Indica Arrow que para separar ambos grupos de ordenaciones el único criterio posible es un juicio de valor y, por consecuencia, que la teoría de la elección social tiene que hacerse cargo no sólo de los valores que rigen las ordenaciones de preferencia, sino también de los valores que deciden qué valores han de tenerse en cuenta en las ordenaciones.

Esta continua aparición de juicios de valor en todos los niveles de la elección social -y el denodado empeño de la teoría por hacerse cargo de ellos- nos pone de nuevo ante el problema de la racionalidad-cientificidad de la teo-

ría. Recordemos que para los autores de la nueva economía del bienestar el puro de hecho de operar en este nivel con valores vaciaría de racionalidad todo el proceso formal. Contrapuestamente, Arrow ve lúcidamente que, si han de derivarse juicios de bienestar social, el procedimiento no debe excluir los juicios de valor -que pueden ser la "buena" subjetividad que nos ayude a construir la intersubjetividad. Por lo demás, la racionalidad del procedimiento ya no se vería empañada por su admisión, puesto que ésta depende de la forma de constituir los propios ordenamientos. Si un ordenamiento -individual o social- de preferencias cumple las condiciones de reflexividad, orden completo y transitividad, puede decirse que es racional, independientemente que la relación de orden se determine por gustos o valores.

Ahora bien, estas son las condiciones para la construcción de ordenamientos de preferencias; la lección que ellas nos dan es que la racionalidad ha de ser formal a fin de que pueda operar con cualesquiera juicios de valor individuales. Pero, la racionalidad colectiva exige más, puesto que lo que está en ella en cuestión es "construir una relación de orden para la sociedad en conjunto que refleje también un proceso racional de elección, de forma que R pueda decirse que satisface los Axiomas I y II" (estos axiomas son equivalentes a las condiciones A-C para ordenamientos antes señaladas) (47). La constitución es una regla por la que producimos un ordenamiento social a partir de los ordenamientos individuales. Las condiciones de racionalidad colectiva lo que hacen es decirnos las características de esa regla. Sólo que estas características

ya están prefijadas de antemano. Por condiciones "naturales" y "razonables" entiende Arrow los valores de la toma de decisiones democráticos. Por consiguiente, las condiciones de racionalidad definen de un modo operativo esos valores y la regla que buscamos ha de cumplir estrictamente las condiciones que los incorporan. Así, a la regla de elección social se le exige no sólo -como a la regla de decisión individual- que produzca ordenamientos consistentes de preferencias, sino también que no se arbitraria para la sociedad, i. e., que responda a la comunidad de valores sociales que están definidos en las condiciones de racionalidad colectiva.

Arrow enumera un grupo de condiciones en las que se encierra un concepto mínimo de elección democrática a fin de que puedan ser unánimemente aceptadas. Para su enunciación nos atenemos a la forma en que aparece en el artículo de 1967, ya que Blau (48) encontró un error en la formulación primera de 1951:

- (A) Condición de dominio no restringido (racionalidad colectiva): La regla de elección social debe poder trabajar con cualquier conjunto lógicamente posible de ordenaciones individuales.
- (B) Principio de Pareto débil (unanimitad): Si una alternativa x es preferida a una alternativa y en todos los ordenamientos individuales, entonces en el ordenamiento social x ha de ser preferida a y . Con esta condición se trata de asegurar la sensibilidad de los ordenamientos sociales a las variaciones en los ordenamientos indi-

viduales, condición que viene impuesta por la propia naturaleza del bienestar social (49).

- (C) Independencia de alternativas irrelevantes: La elección social entre un conjunto de alternativas tiene que depender de las ordenaciones individuales sobre esas alternativas; todo otro ordenamiento ha de ser considerado como irrelevante para la elección social.
- (D) Condición de no dictadura: La constitución no ha de ser tal que si un individuo prefiere x a y , la sociedad prefiera x a y , con independencia de la preferencia de todos los demás miembros de la comunidad.

Bajo las condiciones expuestas el proceso de elección queda determinado como una votación de la siguiente forma. Por la condición (C) se asegura que el ordenamiento social venga determinado solamente por las preferencias de los miembros de la comunidad con respecto a las alternativas presentadas, ello origina un ordenamiento social completo y, por tanto, una elección social para todos los entornos posibles. La condición (B) asegura que el voto para una determinada alternativa favorezca la adopción de esa alternativa; pero, la condición (D) garantiza que una persona no sea decisiva en la votación. Por último, la condición (A) dota a las elecciones de la consistencia necesaria para asegurar su racionalidad formal.

Por lo demás, repetimos, la característica común de estas cuatro condiciones de las constituciones es la de ser

juicios de valor y reflejar un juicio de valor que se considera aceptable por la generalidad de los miembros de la sociedad. Expresan, dice Arrow, "en forma muy general la doctrina de la soberanía y racionalidad de los ciudadanos, permitiéndoles tener un amplio campo de valores"(50). Ciertamente, "de forma muy general" y débil expresan condiciones normativas que al menos en nuestras sociedades democráticas occidentales movilizan un apoyo intuitivo. Sin embargo, lo que ha continuación va a encargarse de demostrar Arrow es que no hay ninguna regla de elección social que satisfaga estas condiciones y que, por tanto, sea una constitución.

Consideremos brevemente la demostración del Teorema de Posibilidad de las Funciones de Bienestar social en el que se enuncia que ninguna constitución puede satisfacer todas las condiciones de racionalidad colectiva (51). Para exponer la lógica de la prueba seguiremos la formulación de W. Vickrey (52) que es más intuitivamente apreciable que la del propio Arrow.

En primer lugar es necesario definir un conjunto decisivo D : un conjunto de individuos D es decisivo para las alternativas x e y en una función de bienestar social dada, si la función prefiere x a y siempre que todos los individuos de D prefieren x a y , y todos los demás y a x .

A continuación utilizaremos una prueba de "cajas chinas" para exponer la demostración:

1. Sea D un conjunto de individuos decisivo para x e y .
2. Supongamos que para todos los miembros de D $xPyPz$, y para el resto C $yPzPx$.
3. Para la sociedad xPy .
4. Para la sociedad yPz .
5. Para la sociedad xPz .
6. Sólo para los miembros de D xPz .
7. La sociedad debe preferir x a z con independencia de otras cosas.
8. D es decisivo para x y z .
9. D es decisivo para todos los pares de alternativas.
10. D debe contener dos o más personas.
11. Dividimos D en dos subconjuntos no vacíos A y B .
12. Supongamos que
para A : $xPyPz$
para B : $yPzPx$
para C : $zPxPy$
13. Puesto que para todos los miembros de A y B , yPz , para la sociedad yPz .

Supuesto.

Condición de dominio no restringido.

Definición de D .

Condición de Pareto.

Transitividad.

Supuesto.

Condición de independencia de las alternativas irrelevantes.

Definición.

Repetición de los pasos 2-8.

Condición de no dictadura.

Supuesto.

Condición de dominio no restringido.

Definición de D .

- | | |
|---|------------------|
| 14. Si para la sociedad xPy , B es decisivo para y y x . | Definición de D. |
| 15. Si para la sociedad xPy , entonces para la sociedad xPz . | Transitividad. |
| 16. Pero entonces A es decisivo para x y z . | Definición de D. |

Los pasos 10-16 pueden repetirse hasta que un subconjunto de D tenga un solo miembro que viola la condición de no dictadura al ser decisivo para todas las alternativas abiertas a la sociedad. ¿Por qué sucede esto? Es el resultado de la combinación de todas esas condiciones -que separadamente son inocuas- unidas a la condición de racionalidad formal de transitividad. La condición de dominio no restringido hace posible cualquier ordenación de preferencias. Entre ellas hay que encontrar una alternativa paretianamente preferida. La condición de independencia de alternativas irrelevantes en un momento dado (en el paso 7) obliga a concentrarse sobre dos alternativas. Entonces se reproduce el problema de la paradoja de la votación cuando se realizan tres votaciones entre pares. Se produce, consecuentemente, un ciclo que la transitividad nos obliga a resolver y la única salida es que esa elección sea dictatorial: "Con la información disponible, las ordenaciones ordinales individuales de pares de alternativas, no existe método que no sea impuesto o dictatorial" (53).

Arrow hace extensivo este resultado a los métodos de agregación de utilidades, con o sin comparación inter-

personal (54), y a los principios de compensación (55), concluyendo con la siguiente interpretación de su Teorema:

"Si excluimos la posibilidad de hacer comparaciones interpersonales de utilidad, entonces los únicos métodos para pasar de los gustos individuales a preferencias sociales, que sean satisfactorios y que estén definidos para un campo amplio de conjuntos de ordenaciones, serán impuestos o dictatoriales" (56).

La búsqueda de un procedimiento que permitiera el paso de la subjetividad a la intersubjetividad se salda con este resultado aparatoso. Arrow piensa la intersubjetividad -qua fundamento de los juicios de bienestar social- como unanimidad. Las condiciones normativas de racionalidad colectiva son naturales y razonables porque las considera como aquellas condiciones que serían elegidas por una sociedad unánimemente. Pero, lo que nos encontramos es que la intersubjetividad fundamentadora a la que tratan de dar cuerpo aquellas condiciones no se alcanza porque entre todas ellas se configura un mecanismo que se autodestruye. La coincidencia de las cuatro condiciones normativas -más la transitividad- excluye aquello que la unanimidad trata de proteger -la soberanía de los ciudadanos- y que está postulado como una condición entre ellas. Si no hay otra forma de concebir la racionalidad colectiva, se puede concluir que es la propia racionalidad de las elecciones sociales quien nos priva de la re-

gla por la cual habríamos de producir juicios de bienestar social, al no ser capaz de realizar lo que le es esencial: la unanimidad. De este modo, la intersubjetividad, pensada desde los valores democráticos de las sociedades occidentales, aparece como un pensamiento imposible.

I.4 Universalidad del resultado de imposibilidad y unanimidad.

El resto del libro de Arrow se consagra a estudiar las posibilidades que se tendrían de obtener una constitución al relajar alguna de las cuatro condiciones. Arrow piensa que la única condición que podría modificarse -si hay que seguir manteniendo que las comparaciones interpersonales no son significativas y que la ordenación social ha de ser sensible a las ordenaciones individuales- es la condición (A) de dominio no restringido. Se trata de hacerla más restrictiva, de hacer que las ordenaciones individuales con las que trabaja la regla de elección social no sean cualesquiera, sino sólo aquellas que se ajustan a unos supuestos.

Estos supuestos son los juicios de valor individualistas. Aunque la razón que da Arrow en favor de estos supuestos es la popularidad que tienen entre los teóricos de la economía del bienestar (57), es obvio que lo que los hace recomendables como solución a la imposibilidad que resulta del Teorema es la esperanza que ofrecen de ser "puente" entre la racionalidad colectiva y la sobera-

nia de los individuos. Estos supuestos se enuncian del siguiente modo:

- (a) Relación de indiferencia: La comparación de dos situaciones sociales alternativas por parte del individuo depende únicamente de los bienes que recibe (y del trabajo que aporta) en cada situación, i. e., el individuo sería indiferente entre dos situaciones sociales cualesquiera en las que su situación de consumo-ocio-ahorro fueran las mismas o, al menos, le fuesen a él indiferentes.
- (b) Relación de preferencia: Al comparar dos situaciones personales en una de las cuales el individuo recibe al menos tanto de cada bien como en la otra (incluyendo como bienes el ocio y el ahorro) y más, al menos, de uno de los bienes, el individuo preferiría la primera situación (58).

Ahora bien, bajo esta restricción Arrow sigue probando que se viola la condición (D) de no dictadura, por la misma intervención de la transitividad que obliga a resolver los ciclos de preferencias(59), y hace extensible este resultado a cualesquiera otras restricciones que establezcan cuasi-ordenamientos. El resultado de imposibilidad para las constituciones tiene, de esta manera, una validez universal.

El capítulo VII de "Social Choice" es una mirada hacia la utopía social. Si toda constitución implica dic-

tadura, pensemos en una sociedad en que la dictadura no supusiera dejar de lado las preferencias de sus miembros. Una sociedad así sería aquella en que hubiera unanimidad completa; i. e., una sociedad en la que los valores de cada individuo fueran los de todos los demás y viceversa. Una sociedad, por tanto, carente de conflictos, homogénea y muy socializada (60). En este caso se disuelve el problema de la constitución; tómesese a un individuo cualquiera y el ordenamiento de sus preferencias es, desde el principio, un ordenamiento de alternativas válido socialmente.

Este caso, según Arrow, puede generalizarse, si suponemos una mayoría de la población con unanimidad de valores (61). Tal caso fue estudiado formalmente por D. Black con la intención de sacar a la regla de la mayoría de la paradoja de la votación (62); si hay un supuesto bajo el cual la regla de la mayoría no produzca ciclos de preferencia, podría considerarsela una regla de elección social satisfactoria, i. e., una constitución. Arrow trae, por ello, a colación el estudio de Black mostrando que bajo el supuesto de la unanimidad de valores es posible construir una constitución: "Matemáticamente, al menos, es posible construir funciones apropiadas de bienestar social (constituciones), si nos sentimos autorizados a decir por anticipado que los gustos de los individuos caen dentro de cierto ámbito de similitud preestablecido" (63). Si se supone unanimidad de valores, ("matemáticamente, al menos") es posible definir una regla de elección social no dictatorial. Pero, entonces, la cuestión es: ¿qué sig-

nificado real para una sociedad con un inmenso y diversificado conflicto no sólo de gustos, sino también de valores -y ello tanto entre los individuos como entre los diversos grupos sociales, culturales, étnicos, religiosos, etc.-, puede tener una regla de elección social formada bajo un supuesto de unanimidad social? Arrow, naturalmente, expresa sus dudas. Sin embargo, apela a una importante tradición de filosofía social que ha respondido afirmativamente a esa cuestión: la tradición contractual.

Es posible, ciertamente, interpretar lo que aquí estamos llamando unanimidad como "voluntad general". Bajo esa interpretación, el supuesto de unanimidad social pierde su carácter abstracto para convertirse en una hipótesis de trabajo para la comprensión de las elecciones sociales. A cambio adquiere todas las objeciones que las teorías contractuales tienen que enfrentar para defender el concepto de voluntad general (64).

Arrow no es el primero que en la historia reciente del pensamiento económico ha expresado y argumentado la necesidad de un supuesto de unanimidad social como fuente última de la posibilidad de las recomendaciones económicas (65). En concreto, Arrow tiene a mano la polémica que a mediados de los años treinta tuvieron Dobb y Lerner sobre lo problemático de este supuesto. Arrow anota las consecuencias de ese debate. En primer lugar, existe el rechazo que una persona educada en los principios liberales siente ante cualquier idea que conlleve un riesgo de "absolutismo" ético -y la voluntad general es de ese ti-

po de ideas (66). En segundo lugar, si se supone la existencia de dicha voluntad, permanece el problema de encontrar el método para hacer coincidir tal voluntad general con las voluntades particulares (67). En tercer lugar, se plantea la cuestión misma de la democracia. Esta exige que una decisión mayoritaria no sea la sacralización de la subordinación de unos intereses a otros, sino más bien el método de discusión entre esos intereses. De otro modo, la voluntad general no ha de homogeneizar manu militari las voluntades particulares, sino que éstas han de poder constituir dialécticamente una voluntad general (68). Por último, se plantea la cuestión de que cualquier idea de unanimidad niega el carácter de "buen" método de decisión al mercado, puesto que este no es sensible a los imperativos categóricos de la voluntad general (69).

La obra de Arrow no se define sobre la aceptación o el rechazo del supuesto de la unanimidad social; tampoco ofrece conclusiones previsoras de cual sería el camino correcto a seguir (70). La última sección del libro deja flotando en el aire la idea de que, si bien es la unanimidad la condición de posibilidad de los juicios de bienestar, el procedimiento para llegar a ella tendría que evitar todo riesgo de "absolutismo" ético. Para ello indica como adecuados los "caminos" que amalgaman las opiniones divergentes que conducen a un consenso general, dejando a un lado los procedimientos que borran el diálogo. La última sección se titula significativamente "El proceso de decisión como un valor en sí mismo", y en ella

nos dice Arrow que siendo menor el valor que ha de darse al proceso de decisión en relación con el objeto mismo de la decisión, no obstante habría de incluirse en el modelo ese valor, de forma que los juicios de bienestar social sean el resultado final de "un acuerdo amplio sobre la desabilidad de las decisiones cotidianas" (71). La teoría de la elección social tenía en ello marcado su programa. Los modelos que luego analizaremos realizarán la tarea de dar forma a ese supuesto de unanimidad social imprescindible para la fundamentación de juicios de bienestar social.

I.5 La dimensión ética de la elección social.

Para hacernos cargo plenamente de algunos de los problemas filosóficos de la elección social tal y como la plantea el modelo de Arrow, así como del alcance de las conclusiones de su Teorema, consideraremos algunas críticas que siguieron a la publicación de su libro (72). Nos centraremos en el trabajo de I. M. D. Little (1952) por ser representativo del tipo de objeciones que se plantearon, pero sobre todo porque permite evidenciar las dimensiones ética y política del planteamiento arrowiano.

Little comienza analizando la relación del Teorema de Arrow con el trabajo de Bergson y, extensivamente, con la economía del bienestar. Trata de probar que el Teorema de posibilidad no se aplica a las funciones de

bienestar de Bergson y, consecuentemente, no dice nada sobre su posibilidad o imposibilidad (73).

Tal y como lo interpreta Little, la función de bienestar social de Bergson es una regla que produce ordenamientos, bajo la condición de que los gustos esten dados. Arrow, por su parte, exige que los ordenamientos y los gustos de los que derivan varien según una correspondencia definida y consistente; Bergson, para las variaciones, sólo exige que luego se puedan volver a ordenar. Esto es, mientras Arrow exige que haya una consistencia a lo largo de las variaciones, Little y Bergson solo piden que haya ordenaciones en los distintos cambios, aunque no exista una relación especial entre ellos. Esta distinción es importante porque constituye la razón por la que, por una parte, Arrow prefiere como materia prima de los juicios de bienestar las ordenaciones individuales hechas sobre valores -a las hechas sobre gustos-; y, por otra parte, es la razón por la que Little considera que el planteamiento de Arrow no tiene relevancia para los problemas de la economía del bienestar. Lo distintivo de las funciones de bienestar social de Bergson es que indican un estado óptimo en la frontera de Pareto, estando las preferencias dadas. Ello significa que ni Bergson ni la economía del bienestar tienen el interés que Arrow manifiesta por los ordenamientos de la sociedad -i.e., por la elección social misma (74).

Arrow responde a esta crítica señalando la necesidad de llevar al terreno de los ordenamientos sociales

los problemas de la economía del bienestar (75). Señala que tal exigencia venía impuesta desde que en la teoría de la demanda se puso a prueba la validez del principio de los indiscernibles de Leibniz: "Solamente las diferencias observables pueden usarse como base de la explicación" (76). Lo que Bergson hizo, en opinión de Arrow, es aplicar al campo del bienestar social este mismo principio, haciendo depender los juicios de bienestar social de la conducta observable. Cuando Arrow exige que haya una consistencia en los cambios de ordenaciones, lo que está exigiendo no es tanto la coherencia de conducta de los individuos como la constancia de los elementos relevantes para las ordenaciones; y ello, naturalmente, requiere seguir todo el proceso de cambio. Lo que Arrow sugiere -y Guibault e Inada se encargaron de probarlo (77)- es la impracticabilidad de un modelo que base sus análisis en ordenamientos sobre puntos sin conexión entre sí. El reto que Arrow ha lanzado a la economía del bienestar -y que, todo hay que decirlo, aquella no ha recogido- es la investigación y formalización en modelos de la génesis de las preferencias sociales.

No obstante la crítica más encarnizada de Little se dirige al propio proceso de formación de los juicios de bienestar social en Arrow. Su desacuerdo es radical: la pretensión arrowiana de que la constitución -definida por las condiciones (A) a (D)- es un procedimiento adecuado para determinar juicios de bienestar social es una pretensión infundada (78).

La constitución -interpreta Little- es una especie de máquina que da una respuesta del tipo "x es mejor que y", cuando todos los individuos han introducido sus respuestas a una cuestión del tipo "¿Es x mejor que y?". La respuesta de la máquina no puede ser un juicio de valor porque para ser tal tendría que ser emitido por una persona: "Todo juicio de valor ha de ser el juicio de valor de alguien" (79). La respuesta de la máquina no es, en opinión de Little, un juicio de valor, sino un "fallo" (en el sentido de una norma) entre los juicios de valor introducidos en ella. Dice solo que la votación, según esa norma, ha sido favorable a tales o cuales juicios de valor presentados.

En esta interpretación de la constitución, continúa Little, la máquina es llamada correctamente "procedimiento de decisión". Pero, lo que no es la máquina y lo que no puede llamarsele es "función de bienestar social" -en el sentido de Bergson-. Porque la función de bienestar de Bergson es un juicio de valor y, por tanto, el juicio de valor de algún individuo. Si se llama "función de bienestar social" a la máquina, se está diciendo que la máquina es un juicio de valor y, por tanto, el juicio de valor de alguna persona. Esto nos aboca a una contradicción. puesto que es condición arroviana de la máquina el que ninguna persona imponga su ordenación (condición (D) de no dictadura). De donde argumenta Little que "es inconsistente llamar a la máquina función de bienestar social y a la vez aceptar la condición de no dictadura"(79).

Este sutil argumento trae de nuevo a la palestra la cuestión que pende siempre sobre los valores unánimes en las teorías del contrato: las condiciones de identificación de la voluntad general. Se podría decir: "De acuerdo; guiémosnos por esa parte más noble que cada individuo lleva dentro de sí mismo. Tomemos a la voluntad general como guía segura de lo que es mejor para la sociedad y aceptemoslo porque es la voluntad de la sociedad". Pero, al menos malicioso se le vendría en seguida a la mente la pregunta que pone en dificultades toda la operación: "¿Cómo sabremos que tal decisión es de la voluntad general y no de la voluntad particular de alguien? ¿Qué identifica a la voluntad general? ¿Quién decide que tal cosa es voluntad general?". Las condiciones de Arrow parecen asegurar que la decisión que se alcance expresa una voluntad general; mas, en la medida en que esas condiciones definen un sistema ético, dependería de que concordarse con los sistemas éticos de los individuos el que fueran reconocidas como las características propias de la sociedad que las constituye. A continuación irían saliendo al escenario todo el conjunto de dificultades que plantea a la filosofía política el tema de la legitimación de la autoridad del estado y su pareja el tema de la fundamentación moral de la obligación política.

En su respuesta Arrow señala la conveniencia de la distinción de Little entre juicios de valor y reglas de decisión (80). Pero, él tiene implícita en su obra una teoría de los valores más sólida de lo que piensa Little.

Arrow se adhiere -y hace constar que implícitamente Bergson también lo hace- a la tradición del utilitarismo. Pensar en valores objetivos e independientes de los valores de los individuos es platonismo acientífico, nos dice; el bien social es el bien de los individuos, el bien que los individuos reconocen como suyo. Por eso, el problema de la elección social es un problema de obtener un ordenamiento social a partir de los ordenamientos individuales, no un problema de ascenso místico hacia el bien. Hasta aquí Little, en opinión de Arrow, lleva razón al señalar que los juicios de valor son juicios de personas y no juicios de instancias supraindividuales.

El criterio del utilitarismo clásico es, no obstante, insuficiente para resolver ese problema. Decir que las alternativas han de ser juzgadas por sus consecuencias para la sociedad no sólo deja abierta la cuestión del método, sino principalmente la problemática de la multitud de juicios que las consecuencias de un curso de acción pueden merecer a los diversos individuos de una sociedad. Bergson señaló el camino de salida: entre los diversos puntos óptimos de la frontera de bienestar de Pareto, se considera "optimum inter óptima" aquel que sea señalado por la función de bienestar social para un estado de preferencias dado de una sociedad. Sólo que aquella función de bienestar social -un verdadero juicio social- se construye bajo determinados supuestos de valor que son decididos enteramente por la persona que emite la función de bienestar social. Por consecuencia, de nuevo se reprodu-

ce la multiplicidad de juicios sociales en la sociedad y no poseemos un criterio para discriminar entre ellos. La constitución es ese criterio; pero, también es un juicio de valor. La discrepancia con Little comienza aquí: las funciones de bienestar social de Bergson y Arrow (en éste, constituciones) no son la misma clase de funciones.

La selección de las funciones de bienestar social de Bergson requiere un criterio o, como lo llama Arrow, "una evaluación de segundo orden". La constitución de Arrow es una regla de elección social que determina un ordenamiento de preferencias para todo posible entorno de alternativas que se presente a la sociedad en base a los propios ordenamientos de preferencias individuales. Con esa ordenación se elige, para cada caso concreto, un ordenamiento social (una función de bienestar social de Bergson) que es, ciertamente, el juicio social de bienestar para esas alternativas. Tenemos, así, dos juicios sociales de bienestar. Uno -el de Bergson- que es concreto; y otro genérico que es la regla por la que se determina el primero -el de Arrow. El juicio social de Arrow se construye (no como el de Bergson, arbitrariamente) bajo unas condiciones normativas de racionalidad que definen el sistema mínimo de valores de la sociedad. Aquí es donde se incardina la crítica de Little: la función de bienestar social de Arrow es un procedimiento de decisión, pero no un juicio de bienestar social porque es impersonal. Arrow, sin embargo, tiene que decir algo a esto: "(...) de la misma manera que para el individuo se identifican valores y elecciones, yo considero los valores sociales significando

nada más que elecciones sociales" (81). ¿En qué basa Arrow esta identificación?

En el artículo de 1967, Arrow sostiene que, para el tema de la decisión, un sistema de valores sería "la regla que un individuo usa para elegir cuál de las acciones mutuamente excluyentes emprendería" (82). Ahora bien, lo interesante para la teoría de la elección social no es elegir entre cursos de acción alternativos, sino entre entornos alternativos. En consecuencia, la elección se hace sin saber qué grupo será el disponible. Si el individuo se hace la pregunta "¿Qué acción elegiría si se diese el entorno X?" para todos los entornos, obtendríamos "una descripción de su sistema de valores en el sentido de una regla que proporciona su elección hipotética para muchos o todos los posibles entornos" (83). Pues bien, no otra cosa es lo que hace la constitución. Definida sobre un sistema ético mínimo racional determina y amplía un sistema de valores sociales de bienestar que daran lugar a los juicios de bienestar social. En razón de lo cual, comenta Arrow, "mientras Bergson intenta localizar los valores sociales en los juicios individuales de bienestar, yo prefiero situarlos en las acciones emprendidas por la sociedad a través de sus reglas de formación de las decisiones sociales" (84).

Desde aquí se puede defender contra Little la no necesidad de que el juicio de bienestar social sea de alguna persona en particular: "Una regla para alcanzar decisiones sociales puede ser acordada por razones de conveniencia y necesidad, sin que haya que tratarlas como

evaluaciones de nadie en particular" (85). Y, a continuación, señala que precisamente lo característico de los juicios de bienestar social es que no son de nadie en particular. Aún más -y aquí se comprueba el abismo que ya separa las concepciones de la teoría de la elección social y de la economía del bienestar- es necesario que no sean de nadie en particular. Si los valores sobre el bienestar fueran individuales, las decisiones sociales no tendrían ninguna efectividad; es porque los valores son sociales por lo que las elecciones son válidas para la sociedad (86). Sobre esta argumentación funda Arrow la capacidad operativa de su modelo. Mientras que la economía del bienestar trata de formular recomendaciones para los productores y los consumidores individuales, la teoría de la elección social trata de formular recomendaciones para los funcionarios del "Estado del Bienestar". Es aquí donde desaparece la contradicción que observaba Little entre aceptar la condición de no dictadura y llamar a la constitución "juicio de valor". El funcionario en el momento de poner en marcha la máquina no es un dictador, puesto que no actúa como un individuo particular sino como "nadie en particular". En este sentido, aunque se aceptase la muy objetable idea de que "la lógica de los juicios de valor es tal que no se puede asumir consistentemente un ordenamiento de valor cualquiera que difiera del propio" (87), sostenida por Little, se puede decir que al menos hay un tipo de personas que en razón de su cargo tienen que asumirlo: los funcionarios del estado.

No obstante, Little lleva más allá sus objeciones: ¿por qué un individuo cualquiera aceptaría las condiciones de Arrow -en concreto la condición de no dictadura-, cuando el sistema de valores público no coincida con el suyo? Little señala que la explicación obvia -la que da Arrow- es que el individuo valorará el procedimiento por sí mismo; considerará un valor en sí mismo un procedimiento que no haga a nadie dictador, sumado a que también valorará el que exista un procedimiento de elección social a que no exista ninguno (i. e., la anarquía)(88). Esta mezcla de razones morales y prudenciales, sin embargo, -insite Little- "no es una razón suficiente para aceptar la condición de Arrow de no dictadura, la cual excepto por el nombre, no contiene nada acerca del poder o la fuerza" (89). Estas razones pueden explicar la conveniencia para las sociedades democráticas de aceptar esa condición; pero, no justifican por qué alguien tiene que aceptar unas condiciones que dan lugar a situaciones que le perjudican y que, por lo demás, no dicen nada de las dictaduras de más de una persona.

El propio Little indica que esta objeción no es un fallo técnico del modelo de Arrow; técnicamente no hay grietas. Es precisamente una objeción que señala la limitación del formalismo para hacerse cargo del problema global de la fundamentación de los juicios de bienestar. No es suficiente decir que los juicios sociales son posibles de ésta o aquella manera; hay que decir también por qué han de ser aceptados por los individuos. Este tema -el de la obligación política- trae de la mano el de la

justicia. Los juicios sociales no sólo tienen que ser racionales, además tienen que ser justos; esto es, cualquiera que sea el contenido que demos a este concepto, los juicios de justicia está obligados a contener razones para convencer a aquel a quién las decisiones sociales no agradan de que no está siendo arbitrariamente maltratado y que, por consiguiente, contienen una obligación moral de aceptación.

La respuesta de Arrow a esta objeción es, obviamente, argumentar para los procedimientos de la elección social lo que se argumenta para los procedimientos democráticos en general: los individuos acatan la decisión no porque sea la mejor desde el punto de vista de sus preferencias particulares, sino porque es fruto de un compromiso. En una situación en que la persecución de esas preferencias individuales llevase a que todos estuvieran peor, la única alternativa es llegar a un acuerdo de aceptación del punto de vista social. El compromiso que cada uno realiza de aceptar el resultado de la decisión social -aunque sea tácito- es parte fundamental de cualquier método de elección social. De este modo, Arrow concluye que "las obligaciones morales son parte de la decisión social a conseguir por la maquinaria de la toma de decisiones colectiva" (90), puesto que éstas no se pueden concebir sin ese acuerdo previo que las conforma como procedimientos de elección social.

Ahora bien, ¿qué razones tienen los individuos para aceptar ese compromiso? Lo que realmente daría fuerza de razón para la aceptación de las decisiones sociales

-para la obligación moral- al compromiso es que este fuera justo. Lo que el resultado de imposibilidad de Arrow significa es que no existe la posibilidad de un compromiso justo. Porque justo sería aquel acuerdo que garantizase que ningún punto de vista individual dominase sobre el resto; es decir, que se adecuara a la condición formal de la justicia de dar igual peso a todos los participantes (91). La imposibilidad de que haya procedimientos de decisión social democráticos aparece aquí como la imposibilidad de que haya procedimientos justos. ¿Es este un resultado universal o es más bien un resultado producido por esa limitación de los modelos formales para hacerse cargo de los valores sociales?

Las investigaciones que siguieron a la obra de Arrow muestran que los teóricos de la elección social creyeron más bien que la imposibilidad de racionalidad, democracia y justicia es un resultado peculiar al modelo de Arrow. En concreto, que la estrechez del marco ordinalista, la exclusión de las comparaciones interpersonales, lo rígido de su noción de racionalidad abocan a un resultado de ese tipo inevitablemente. Pero, lo que no es inevitable es aceptar la estructura del modelo arrowiano. En los capítulos siguientes veremos modelos diferentes al de Arrow y comprobaremos su capacidad para hacerse cargo de las exigencias de la noción de justicia.

I.6 La pregunta por el fundamento (II).

Vamos a reconsiderar en esta última sección los primeros pasos de la fundamentación en la teoría de la elección social, pero ahora de manera menos expositiva y más reflexiva. Para ello nos valdremos de la comparación de este modelo de fundamentación con el que aparece como el modelo alternativo más próximo en el tiempo: la teoría de la justicia de J. Rawls (92).

En primer lugar destacaremos la forma específica de la pregunta por el fundamento en ambos modelos. Como se vió en las primeras secciones de este capítulo, la teoría de la elección social se pregunta por una clase genérica de los juicios de valor. Su pregunta es: ¿cómo son posibles los juicios de bienestar social? El bienestar social es un objetivo de la sociedad y, como tal, un valor social. De aquí que su pregunta en realidad sea una pregunta por los fines de la sociedad: ¿cómo es posible que la sociedad se dé fines a sí misma? Contrastadamente, Rawls se pregunta por un tipo de juicios de valor muy específicos. La pregunta por el fundamento en el modelo rawlsiano es: ¿cómo son posibles los juicios sobre la justicia de la estructura básica de la sociedad? Es intuitivamente apreciable que ambas preguntas no son equivalentes, si bien no sólo por la diferencia de generalidad de los juicios por los que se cuestionan. En último término, la pregunta por los juicios de bienestar social se puede especificar en una pregunta por los juicios de justicia social, y ello

no significaría que en ambos modelos se estuviera preguntando lo mismo.

La diferencia entre ambas preguntas reside en la diferencia de contenido del concepto de juicio de justicia que aparece en ellas. No es un azar el que la teoría de la elección social se pregunte primariamente por el bienestar social (aunque se admita como una especificación posible de su pregunta general la cuestión ulterior sobre la posibilidad de los juicios de bienestar social justos). Ello indica que el bienestar es una cuestión prioritaria y determinante de cualquier otra cuestión, por ejemplo de la cuestión de la justicia. A pesar de lo cual, la teoría contiene una amplia banda de modelos específicos: desde la reducción de toda cuestión de justicia a una cuestión de bienestar social (Harsanyi) hasta el reconocimiento de una paridad en la determinación de los juicios sociales del bienestar y de la justicia (Sen). Con todo, lo característico de la teoría de la elección social es que su objetivo básico ^{sea} la resolución de problemas de bienestar social y que, en sus planteamientos, el concepto de bienestar tiene, cuando menos, un papel privilegiado entre todos los conceptos éticos relevantes.

De modo equivalente, no es un azar que Rawls se pregunte directamente por los juicios de valor justos. Ello indica que, independientemente del contenido que se asigne a la noción de justicia, esta tiene prioridad sobre cualquier otro criterio. Si la pregunta por la posibilidad de los juicios de bienestar social significa en último

término una pregunta por la posibilidad de que la sociedad se dé a sí misma fines, la pregunta rawlsiana se resuelve en una pregunta más precisa: ¿cómo es posible determinar la justicia de los fines alternativos de la sociedad? Esta pregunta que, a primera vista, puede aparecer como complementaria de la primera, es, sin embargo, su antagonista; si se las pone juntas, la pregunta por la justicia termina corrompiendo todas las posibilidades de la pregunta por el bienestar. Porque lo que se esconde en la pregunta rawlsiana es la exigencia de que la sociedad sólo pueda elegir sus fines según principios de justicia. Con ello, la pregunta por la posibilidad de los fines sociales es simplemente disuelta. Puesto que lo que supone la pregunta de Rawls es que los fines son posibles desde el primer momento en que los principios de justicia son aceptados -rigen por la sociedad. Si tal sociedad justa existe, existe como una comunidad con fines no conflictivos. Por tanto, no hay cuestión sobre la posibilidad de los fines; esta cuestión, en todo caso, se reabsorbe en otra de una dimensión diferente: ¿cómo es posible que una sociedad se rija por unos principios de justicia válidos para todos sus miembros?

Ahora bien, del diferente significado de sus preguntas iniciales se sigue la diferente problemática que ambos modelos tienen que enfrentar para obtener una respuesta a aquellas. Por más diversos que sean esos caminos, la contrastación de los caracteres que marcan en las teorías no es inútil. Permite la comprensión cabal de las

dificultades, soluciones e impotencias con las que nos encontramos en sus desarrollos.

Comenzaremos por la teoría de J. Rawls. Como hemos dicho, la diferencia sustancial de su manera de preguntarse por el fundamento implicaba ya una primacía de la justicia sobre cualquier otro tipo de consideraciones. Esto es lo que propiamente encontramos en las primeras páginas de su libro: "(...) no importa cuán eficientes y bien ordenadas estén las leyes e instituciones, si son injustas han de ser reformadas o abolidas" (93). Esta primacía de la justicia es particularmente puesta de relieve con respecto al bienestar, determinando un concepto de justicia que traza venementemente una línea divisoria entre consideraciones de bienestar y consideraciones morales: "Cada persona posee una inviolabilidad fundada en la justicia que incluso el bienestar de la sociedad como un todo no puede atropellar" (93). Esta primacía de la justicia -y esa dicotomía que conlleva entre bienestar y moralidad- es la idea nuclear que la teoría de la justicia rawlsiana tendrá que explicar y, sobre todo, intentar que no se convierta en una fuente de problemas irresolubles.

Una primera medida para que tal dicotomía no se convierta en problemática consiste en restringir el área de aplicación de la teoría de la justicia. Esto se lleva a cabo mediante una peculiar especificación de su objeto: "Para nosotros, el objeto primario de la justicia es la estructura básica de la sociedad o, más exactamente, el modo en que las instituciones sociales más importantes

distribuyen los derechos y los deberes fundamentales, y determinan la división de las ventajas que proceden de la cooperación social" (94). La dicotomía justicia-bienestar no implica que la consideración del bienestar sea arrojada a las tinieblas exteriores. La justicia proyecta su esfera de juicio sobre los derechos y deberes generales; pero, también, sobre los derechos y deberes particulares que rigen el bienestar social (95) con especial atención a la "división de las ventajas que proceden de la cooperación social". La justicia, pues, entiende en cuestiones de bienestar principalmente en la medida en que en ellas se plantean problemas de distribución de los beneficios y las cargas del sistema social. Sin embargo, hay que resaltar que tal preocupación no abarca todas las situaciones de bienestar. Quedan fuera lo que el propio Rawls considera "problemas sociales fundamentales" como son los de coordinación, eficiencia y estabilidad y a los que atribuye el papel teórico de decidir en favor de la concepción de la justicia que mejor los resuelva, cuando haya dos concepciones igualmente preferibles en los demás aspectos (96). Nosotros expresamos nuestras reservas. Ya se verá si esa es una consideración suficiente de la problemática del bienestar para una teoría que quiere determinar el fundamento de los juicios sociales.

Una vez planteadas así las cosas, dijimos, la pregunta por la posibilidad de los juicios de justicia social se hacía equivalente a la pregunta por la posibilidad de que una sociedad se rija por principios de justicia inter-

subjetivamente válidos. Aparece aquí, como ya lo hiciera en el planteamiento de la teoría de la elección social, la problemática de la intersubjetividad. Nos interesa indagar si esa problemática es similar en los dos modelos.

La cuestión de la intersubjetividad se plantea desde el primer momento a la teoría de la elección social como su problema prioritario. El significado propio de la pregunta por la posibilidad de los juicios de bienestar social -o de la pregunta por la posibilidad de que la sociedad se dé a sí misma sus fines- es equivalente a una pregunta por el modo de constitución de esa intersubjetividad. Pero, como la intersubjetividad es concebida por esta teoría como la de un grupo de individuos que -por algún procedimiento- tienen que concordar en los fines que van a perseguir, tales fines sólo pueden pensarse como un valor social que manifiesta las preferencias individuales de los componentes del grupo. Así, la cuestión de la intersubjetividad, por efecto de tomar como única determinación legítima de los juicios de valor social la que llevan a cabo las preferencias individuales, se convierte en la siguiente cuestión: ¿cómo es posible la preferencia social? Esta es ya la pregunta por el procedimiento de las elecciones sociales -que en Arrow, vimos, se resuelve por una votación, pero que en otros autores como Harsanyi se presenta como la suma de las utilidades individuales-; y esta es la cuestión pugnaz con la que se enfrenta la teoría, estando siempre detrás de todos los desarrollos principales o subsidiarios de la misma. Por que, la verdad, es que en la realidad los grupos humanos presentan múltiples pre-

ferencias individuales sobre los fines sociales alternativos; pero, desligadas entre sí y demasiadas veces irreconciliables. ¿Cómo ligar lo desligado? Ese es el reto que lanza el criterio de la preferencia a la inventiva de la teoría de la elección social, marcando estrechamente las posibilidades de alcanzar la intersubjetividad que defina y justifique la persecución de las metas sociales desde y para la subjetividad varia y encontrada.

No es éste el problema que plantea la intersubjetividad a la teoría de Rawls. Para decirlo inexactamente, el problema en ella es por el contrario el paso de la intersubjetividad a la subjetividad. Esto es, los principios de la justicia que son por naturaleza intersubjetivos han de ser también los principios por los que cada cual enjuicie su situación particular en las circunstancias particulares de su sociedad. Para Rawls, el concepto de la justicia es tal que implica que los principios de la justicia o son intersubjetivos desde el primer momento o no son nada. No se puede concebir, en su opinión, que los principios sean meramente prudenciales o que, como a las reglas de la elección social, se llegue por un proceso no justificado como es el de las condiciones normativas para las constituciones de Arrow.

Los principios de la justicia serán intersubjetivos en la teoría rawlsiana porque estarán de acuerdo, explicarán y extenderán nuestros juicios morales ponderados en equilibrio reflexivo sobre casos particulares; y, además, porque se derivarán de una situación de elección inicial cuyas condiciones son proporcionadas asimismo por nuestros

juicios básicos de moralidad del mismo modo ponderados en una situación de equilibrio reflexivo. Si todo esto funciona adecuadamente, la intersubjetividad es la característica esencial de los principios de justicia. Si ese procedimiento de equilibrio reflexivo no es capaz de justificarlos, no son intersubjetivos, pero ni siquiera son principios de justicia. ¿En qué consiste ese "equilibrio reflexivo" que parece ser la clave de bóveda de la concepción rawlsiana del fundamento de los juicios justos?

El método del "equilibrio reflexivo" aparece en la obra de Rawls como un desarrollo evolutivo interno a las propias transformaciones que éste va sufriendo desde su primera publicación en 1951. Nosotros lo expondremos según aparece en "A Theory of Justice" librándolo de algunos inconsistencias y falsas analogías que el propio Rawls ha corregido en sus escritos posteriores (97).

Los conceptos básicos de este método pueden ligarse del siguiente modo. Cualquier persona normal y madura posee un sentido de la justicia en situaciones concretas, el intento de reducir a un sistema coherente la infinita variedad de estos juicios es la labor que ha de realizar cualquier persona que posea tal sentido de la justicia, y su resultado son juicios morales ponderados. El equilibrio reflexivo es, finalmente, el proceso por el que se establece un ajuste entre esos juicios ponderados y algún grupo de principios de justicia.

La concepción rawlsiana del sentido de la justicia implica una teoría bastante compleja del desarrollo moral de la persona de la que no vamos a ocuparnos aquí (98).

Para nuestro propósito nos basta saber lo que el propio Rawls dice sumariamente en el primer capítulo de su obra sobre él:

"Supongamos que cada persona, después de una cierta edad y en posesión de una capacidad intelectual indispensable, desarrolla, bajo circunstancias sociales normales, un sentido de la justicia. Adquirimos una habilidad para juzgar las cosas como justas e injustas y para apoyar esos juicios en razones. Más aún, comúnmente tenemos deseo de actuar conforme a estos pronunciamientos y esperamos un deseo semejante por parte de los demás. Es claro que esta capacidad moral es extraordinariamente compleja. Para verlo es suficiente advertir el número y la variedad potencialmente infinita de juicios que estamos preparados para formar. El hecho de que a menudo no sepamos qué decir y que a veces nos encontremos mentalmente indecisos, no disminuye la complejidad de la capacidad que poseemos" (99).

Tener un sentido de justicia es el primer paso para concebir la posibilidad de juicios morales intersubjetivos. Sitúa algo en la naturaleza del hombre que es probablemente común a todos los miembros de la especie. Pero, como vemos, es una condición de posibilidad; nada indica en él que los juicios que produzca sean comunes a la humanidad. Hay que poner orden en los juicios que proporciona tal sentido, si hemos de concebir lo intersubjetivo. Rawls añade, en esta línea, que la comprensión real de lo que signifi-

ca tener este sentido de justicia no se alcanza hasta que no sabemos que los juicios de justicia concretos son casos particulares que definen el amplio campo de aplicación de unos principios de justicia (100).

Los juicios ponderados representan el resultado de una consideración de esos juicios diversos y espontáneos, Se concibe a la persona como sintiéndose responsable de alcanzar un sistema coherente de juicios morales "en los que sea más probable que nuestras facultades morales se desplieguen sin distorsión" (101). Ello conlleva la exigencia de que cada persona posea la capacidad de eliminar aquellos juicios que hayan sido dictados por circunstancias poco favorables a ese ejercicio, aquellos sobre los que no tengamos seguridad, aquellos otros con los que únicamente pretendamos nuestro beneficio, etc.. Los juicios que permanecen después de esta labor son propiamente juicios ponderados de justicia y expresan, por así decirlo, el sentido común sobre lo que es justo o injusto. Pero, demasiadas veces el sentido común no es tan común a las personas que lo ejercen.

El equilibrio reflexivo, finalmente, es el procedimiento por el que se intenta alcanzar esa intersubjetividad que se niega a nuestras expresiones ponderadas. Hay que ser cuidadoso en la comprensión de este procedimiento, porque Rawls no siempre es claro. El equilibrio reflexivo constituye un argumento de coherencia; es decir, está dirigido a establecer la plausibilidad de una concepción de la justicia, comprobando si es coherente o no con nuestros juicios ordinarios. Ahora bien, se puede hablar de dos ti-

pos de coherencia ampliamente divergentes y no reconciliables (102).

El primer modelo de argumento de coherencia -el naturalista- concibe la comprobación de esta coherencia de un modo análogo a la que se efectúa en las ciencias naturales. En este modelo las intuiciones morales concretas son tomadas como datos observacionales a partir de las que se trata de descubrir los principios generales que las explicarían. Pero, este no es el modelo de Rawls (103). En el segundo modelo, los juicios ponderados no son datos fijos, sino más bien pautas condicionales de una concepción general que no ha de ser descubierta, sino construida. El equilibrio reflexivo como método de esta construcción no concibe puntos fijos objetivos -no lo son los juicios ponderados, pero tampoco los principios de justicia-, sino que va de un extremo a otro repetidas veces en un intento de encontrar una adecuación entre juicios y principios que permita "un programa coherente de acción" (Dworkin). Con tal proceder se alcanza una posición de reposo -de equilibrio- que, por más estable que sea, siempre tendrá un status de provisionalidad. Es, simplemente, una situación válida por el momento:

"Es un equilibrio porque finalmente nuestros principios y juicios coinciden; y es reflexivo puesto que conocemos a qué principios se ajustan nuestros juicios y las premisas de su derivación. Por el momento todo está en orden. Sin embargo, este equilibrio no es necesariamente estable. Está sujeto a ser transformado en un ulterior examen (...)" (104).

La pregunta que este reconocimiento de provisionalidad deja en el aire es si ello no menoscaba la pretendida intersubjetividad de los principios hallados en el equilibrio reflexivo.

En la sección final de "A Theory of Justice", Rawls reconoce que el equilibrio reflexivo solo puede producir principios relativos. En primer lugar, porque la lista de principios que han de ser comprobados tiene que ser necesariamente finita; y finita en el sentido de la disponible a la sociedad que elige. En segundo lugar, los principios que esa sociedad elige son válidos para ella, pero se admite que otra sociedad de características humanas diferentes podría no aceptarlos y elegir otros. ¿Es, entonces, el equilibrio reflexivo un argumento válido para los principios de justicia?

D. Lyons critica la validez del argumento de coherencia en base a su manifiesta circularidad: los principios explican los juicios, los juicios confirman los principios. Ciertamente, en ese ajuste continuo de principios y juicios, cualquier parada es o convencionalismo o intuicionismo (105). Pero, además, ¿cómo saber dónde detenerse en las reformas de unos u otros? ¿Se puede esperar que haya, en realidad, un ajuste entre ambas partes? ¿No estaremos más bien condenados a ir de un extremo a otro sin fin porque siempre y necesariamente exista un desajuste renovado?

R. Dworkin sugiere (106) -y Rawls parece aceptarlo en sus últimos escritos (107)- que tal circularidad sólo se produciría si el equilibrio reflexivo actuara según el

modelo naturalista del argumento de coherencia; esto es, si nuestra meta fuera la verdad objetiva y eterna de los principios alcanzados. Pero, el equilibrio reflexivo de J. Rawls es un argumento constructivo: basta con que se alcance el mejor programa coherente entre principios y juicios disponibles para el grupo humano que reflexiona y elige, aunque ulteriormente pueda ser revisado. Entendido de este modo el equilibrio reflexivo, los principios pueden ser relativos e intersubjetivos al mismo tiempo; dicha intersubjetividad consiste, entonces, no en una resolución tomada de una vez y para siempre, sino en una dinámica disposición de las personas que constituyen una comunidad para estar permanentemente repitiendo la situación de reflexionar sobre los principios de la justicia y sus propios juicios ponderados. De esta forma, los principios de la justicia son los mejores que se pueden alcanzar, pero no para siempre. Con ello la objeción basada en la provisionalidad de dichos principios se convierte en una nota a favor de los mismos, al permitir el progreso y el perfeccionamiento de las comunidades (108).

Si aceptamos el método del equilibrio reflexivo como garantía suficiente de que los principios de la justicia que por su medio se seleccionan son intersubjetivos desde el primer momento, es ahora cuando tenemos que responder a la cuestión que la intersubjetividad propone a la teoría rawlsiana. Ya señalamos que, a diferencia de lo que sucede en la teoría de la elección social, el problema que aquí se presenta es el paso de la intersubjetividad a la subjetividad. Por ello, la cuestión a resolver no es

cómo construir la intersubjetividad -el constructivismo kantiano que propone Rawls es la expresión de que ese no es el problema-, sino cómo concebir que tal intersubjetividad se asiente en el pecho de cada cual. Porque, ciertamente, el método del equilibrio reflexivo no garantiza por sí mismo que, llevado a cabo, tenga el mismo resultado para todas las personas. Para dar una respuesta a esta problemática Rawls tiene que describir las líneas maestras por las que el proceso reflexivo ha de conducirse (109).

Curiosamente en este punto nos encontramos con el problema de la unanimidad que se planteaba en Arrow y, asimismo, encontramos que, como él, Rawls apela a la teoría del contrato para su resolución. En algún sentido se puede decir que Rawls lleva a cabo lo que solo fue enunciado por Arrow. Sobre todo cuando nos percatamos que la situación contractual es descrita por Rawls como una situación de elección social de unos principios generales de justicia -una constitución arrowiana en su forma- que han de regular las instituciones básicas y, por tanto, las ulteriores decisiones sociales. Esta coincidencia, si se piensa bien, entre los dos modelos no debe causarnos sorpresa, teniendo en cuenta todo lo que la obra de Rawls debe a los autores de la teoría de la decisión -teoría de juegos y teoría de la elección social-, deuda por lo demás reconocida en muchas partes por el autor de "A Theory of Justice".

Pues bien, la manera de abordar el problema de la intersubjetividad por la teoría rawlsiana es construir una situación de elección hipotética según las características

normativas que selecciona el método del equilibrio reflexivo. Con ello se espera que cualquier individuo pueda recomponer y hacer suyo el proceso por el cual se eligen los principios de la justicia de una manera igual para todos. Por consecuencia, lo fundamental es que esas características o condiciones que se imponen a la situación de elección sean aceptables por toda persona que tenga el más mínimo sentido de la justicia:

"Para justificar una descripción particular de la situación inicial hay que demostrar que incorpora estas suposiciones comúnmente compartidas. Se argumentará partiendo de premisas débiles, aunque ampliamente aceptadas, para llegar a conclusiones más específicas. Cada una de las suposiciones deberá ser por sí misma natural y plausible; algunas de ellas pueden parecer inócuas e incluso triviales. El objetivo de la propuesta contractual es el establecer que, al considerarlas conjuntamente, imponen límites significativos a los principios aceptables de la justicia. El resultado ideal sería que estas condiciones determinaran un conjunto único de principios" (110).

El resultado ideal sería que esas características diseñaran un camino ineludible para que toda persona que estuviera dispuesta a pensar sobre el tema encontrara que sólo hay una manera admitida por el equilibrio reflexivo para considerar los principios alternativos y -"consecuentemente" quizás- una única elección entre ellos (111). Pero, Rawls es muy cauto, afirmando tan sólo que la virtud del

método queda probada si se obtienen resultados para las circunstancias concretas en las que se plantea la elección hipotética, añadiendo que su interpretación de esas circunstancias no es la única posible y, correlativamente, los principios elegidos pueden variar (112).

Ahora bien, ¿es razonable la confianza que Rawls pone en la aceptabilidad de las condiciones que diseñan la situación de elección como medio suficiente para que toda persona sienta que los principios de justicia en ella elegidos son también sus principios personales? Tal suficiencia está argumentada en base a que producirían una cooperación social ventajosa para todos. Hay, según Rawls, una continuidad entre las condiciones para pensar la justicia y las condiciones de un acuerdo beneficioso, tal y como se expresa en lo siguiente: "Estas restricciones expresan aquello que estamos preparados para considerar como los límites de una cooperación social en términos imparciales" (113). De este modo, todos tenemos buenas razones para pensarnos como partícipes en la situación inicial de elección. En primer lugar, porque expresa las condiciones naturales de la justicia, resumidas en la condición general de la imparcialidad (fairness) (114). Pero, en segundo lugar, porque sólo un ordenamiento de la sociedad elegido en una situación imparcial asegura que las distribuciones de cargas y beneficios que en ella se den serán imparcialmente beneficiosas para todos sus miembros.

Es nuestro deber, sin embargo, preguntarnos si realmente estos dos ordenes de razones son "buenas" razones; es decir, si son suficientes para garantizar la misión

que se les encomienda: conseguir la aceptación personal de los principios interpersonales.

D. Lyons (115) ha mostrado que en la obra de Rawls la imparcialidad (fairness) tiene un poder superior al de la mera delineación de las características de la posición original. Rawls hace descansar la fuerza justificatoria del argumento del contrato en su condición de procedimiento imparcial que incorpora la idea de justicia procesal pura (116). Pero, lo que en ello parece mostrarse es que a la imparcialidad se le asigna un valor independiente -superior o diferente, según se mire- al que se le da cuando se incorpora en la construcción del procedimiento de la posición original qua interpretación "más favorable" de la situación de elección inicial. ¿Qué interés se tiene en subrayar esta independencia? El interés de mostrar un presupuesto evaluativo que es capaz de comprometernos con ciertos principios de justicia independientemente de las consideraciones sobre los hechos científicos básicos y las reglas generales de la condición humana a los que los liga el argumento contractual:

"El argumento contractual, en alguna medida, se apoya sobre un compromiso no argumentado con la "Fairness" y la imparcialidad. Por tanto, el argumento contractual sólo puede retener alguna fuerza bajo el supuesto de que estos son los más "fijos" o los menos "provisionales" de los puntos "fijos" en nuestro sistema de valores compartido -quizás fundamentales a nuestra (o a una) posición moral" (117).

Pero, estemos atentos. Si esto sucede, nos aparece junto al argumento contractual un criterio independiente para elegir principios de justicia, y todo el esfuerzo rawlsiano para basar la fuerza moral del argumento contractual -y de su solución, los principios de justicia- en el concepto de justicia procesal pura, se torna inútil (118). Desde entonces, o los principios de justicia no son principios morales; o sí lo son, hay que explicar y justificar el supuesto normativo de la imparcialidad del que derivan su fuerza moral.

Este entredicho que cae sobre la imparcialidad nos permite preguntarnos por qué habría de aceptarse individualmente un compromiso fundamental con ese valor y, por tanto, por qué habrían de aceptarse los principios de justicia que ahora ya ni siquiera parecen tan intersubjetivos. Es cierto que la obra de Rawls es un continuo alegato en favor del carácter esencial de la "fairness" con respecto a la noción de justicia. No obstante, y a pesar de la compleja argumentación rawlsiana -y, sobre todo, a pesar de la doctrina kantiana que a su través se reinterpreta-, la imparcialidad se nos aparece como una condición normativa de racionalidad colectiva con el mismo status de supuesto "natural y razonable" que las condiciones de Arrow; pero, al mismo tiempo, tan fuerte e infundado como ellas.

Examinemos ahora el segundo orden de razones por las que se nos trata de convencer para que participemos en la situación inicial de elección: las que apelan a nuestros propios intereses. Se plantea aquí el mismo problema

que la crítica de Little presentaba a Arrow: ¿por qué los individuos han de aceptar y ejecutar las decisiones colectivas? Recordemos que la respuesta de Arrow era que el lugar de la obligación moral en la elección social se sitúa en la propia maquinaria de la decisión. La toma de decisiones sociales, por su propia naturaleza, es una especie de contrato -implica un compromiso previo- en virtud del cual se le puede exigir a los participantes que cumplan sus términos aunque no favorezcan los intereses personales de algunos de ellos. Observemos que esta argumentación no es -ni puede ser- la de Rawls. Su contrato es hipotético. No se dirige a personas que realmente han participado en una decisión social; sino que se dirige a personas con el propósito de convencerlas de que sus intereses se beneficiarían si se considerasen parte de ese contrato hipotético. Lo que vamos a ver a continuación es que la separación entre los intereses en la situación hipotética y los intereses reales de una persona destruyen la capacidad del razonamiento para convencernos.

Es, pues, necesario entender bien el argumento de los intereses. Aunque Rawls pueda mantener que una sociedad gobernada por sus principios sea más beneficiosa para todos que cualquier otra, este argumento lo sería en favor de su sociedad, no en favor de que nos incluyamos en la posición original. Puesto que los intereses bajo los que se eligen los principios no son los intereses de las personas reales. En lo que nos beneficia la elección de principios en la posición original es en que establece una cooperación social para conseguir los objetivos básicos -un

cierto nivel de bienes primarios que posibilita la realización de los planes de vida- de cualquier persona (119). Es decir, lo que se favorece son intereses básicos (no específicos) y propios del hombre en general que es el actor de la situación hipotética de elección inicial.

Este argumento sería bueno si, estando ya en la situación hipotética, se tratara de convencernos de que cumpliríamos nuestra obligación con los principios aceptados al pasar a la situación real. Pero, el caso es justo el inverso. Lo decisivo es que las circunstancias de la sociedad real y las de la posición original son muy diversas. Elegir principios de justicia en la posición original implica una apuesta, detrás del velo de ignorancia y con incertidumbre sobre la sociedad y la posición personal en ella, por tales principios. En esa situación mi interés se configura de una manera muy diferente de como se configura en mi realidad actual. Sabiendo esto y conociendo bien nuestros intereses reales, sabemos que unos y otros no coincidirían (120). En tal caso, ¿por qué habríamos de querer prescindir de nuestros intereses verdaderos, simular que no los tenemos, y elegir principios que, aunque estén de acuerdo con un interés genérico, probablemente perjudicarían nuestros intereses reales?

Esta es una pregunta que pone de manifiesto el dualismo presente en Rawls y similar al kantiano entre el yo sensible y el yo noumenal. Si fuésemos puros sujetos noumenales, los intereses de la razón guiarían nuestra conducta y toda ella sería plenamente moral. La posición origi-

nal, ciertamente, expresa el punto de vista del sujeto noumenal (121). Pero, puesto que también somos seres guiados por los intereses heterónomos de las afecciones, ¿por qué habríamos de someter estos a la primacía de aquellos? La respuesta obvia sería algo así como que el objeto a construir -la sociedad- exige que el esquema de construcción sea antes que nada moral. El aspecto polémico de una afirmación de ese tipo nos salva de cualquier otro comentario.

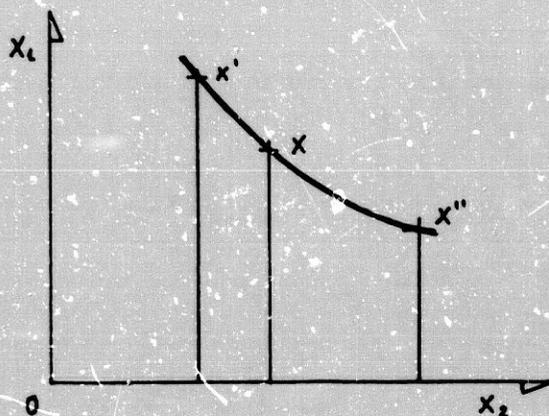
Sin embargo, queremos hacer notar que estamos aquí en presencia de las consecuencias de la concepción genérica de la primacía de la justicia sobre el bienestar. Anunciamos más arriba que probablemente el tema de los intereses -socios indesligables del bienestar- plantearía graves problemas a la teoría rawlsiana de la fundamentación de los juicios sociales. En este sentido, nos gustaría resaltar que el resumen de toda esta problemática podría bien ser la cuestión de la separación de intereses y moralidad llevada a cabo por las teorías deontológicas: ¿por qué los intereses -los intereses heterónomos- no pueden ser morales? Aquí, ciertamente, está el nudo de la polémica entre los que defienden los imperativos y los defensores de los imperativos hipotéticos como base suficiente para una conducta moral y racional. Mas, sobre todo, aquí hay una cuestión de capacidad de la teoría para pensar a los actores reales de los procesos sociales y quizás, propiamente, tales procesos. Desde este punto de vista, el carácter más empírico de la teoría de la elección social parece dejarla mejor situada frente a esa problemática de

lo que queda la teoría de la justicia de Rawls con su apelación a una manera de ser sujetos que no parece ser muy sensible a las conductas oportunas en una sociedad normal.

NOTAS

- (1) K. J. Arrow, (1951-1963²), Elección social y valores individuales. Algunos de los aspectos centrales de esta obra habían aparecido ya, junto a una pequeña autobiografía intelectual, en K. J. Arrow, (1950), "A Difficulty in the Concept of Social Welfare".
- (2) A. Bergson, (1938), "A Reformulation of Certain Aspects of Welfare Economics", en A. Bergson, (1966), pg. 3.
- (3) Un ejemplo de esta "falta" de entendimiento del artículo de Bergson puede verse en J. A. Schumpeter, (1954), Historia del análisis económico, P. IV, c. 7, pp. 1144-65. El clima empezó a cambiar cuando apareció la obra de P. A. Samuelson, (1947), Foundations of Economics Analysis, c. VIII.
- (4) K. J. Arrow, (1951-1963²), p. 53.
- (5) J. A. García Lopez, (1973), "Síntesis y evolución de la economía del bienestar", p. 109.
- (6) Sobre el análisis del equilibrio general, J. A. Schumpeter, (1954), P. IV, c. 7; M. A. Martínez-Echeverría, (1983), Evolución del pensamiento económico, c. XIII; sobre Walras y Pareto son interesantes los artículos de J. A. Schumpeter, (1951), Diez grandes economistas: de Marx a Keynes.
- (7) M. H. Dobb, (1969), Economía del bienestar y economía del socialismo, pg. 14.
- (8) J. A. García Lopez, (1973), p. 109. Como se hablará mucho en este capítulo de curvas de indiferencia de

los consumidores diremos algunas palabras sobre ellas. Una curva de indiferencia es un conjunto de puntos indiferentes entre sí dadas las preferencias de un consumidor. Su forma típica es:



Indica que la relación entre los lotes de bienes (compuestos por combinaciones diferentes de unidades de los bienes x_1 y x_2) x' , x y x'' en la preferencia del individuo i es la de indiferencia. Es decir, que si se ofreciera a un individuo

x' : 4 tomates y 2 patatas

x : 3 tomates y 3 patatas

x'' : 2 tomates y 5 patatas

el individuo i consideraría las tres alternativas como igualmente buenas. Véase, H. A. J. Green, (1971), La teoría del consumo, I.3.

- (9) Los hitos de este desarrollo teórico son los artículos iniciales de Lerner, Kahn y Hotelling de esta década; puede verse un resumen de las ideas allí expuestas en J. A. García-Lopez, (1973), pp. 110-111.

- (10) M. H. Dobb, (1969), p. 17.
- (11) El ejemplo está tomado de J. A. García Lopez, (1978), p. 112.
- (12) I. M. D. Little, (1957²), A Critique of Welfare Economics, pp 84 y ss.
- (13) M. H. Dobb, (1969), p. 32.
- (14) M. H. Dobb, (1969), p. 39.
- (15) Subrayamos "necesarias" porque en ello se diferencia netamente de las tentativas de la nueva economía del bienestar que sólo pretende encontrar condiciones suficientes para el óptimo. Vease, I. M. D. Little, (1957²), pp. 116 y ss.
- (16) El tema de la determinación del óptimo de la producción y el intercambio se desarrolla a partir de estos supuestos. Nosotros no entraremos en ello, pero puede verse su desarrollo en P. A Samuelson, (1947), c. VIII y desde un punto de vista crítico en I. M. D. Little, (1957²), cc. VIII-IX.
- (17) K. J. Arrow, (1951-1963²), p. 79.
- (18) M. C. Kemp- A. Asimakopulos, (1952), "A Note on "Social Welfare Functions" and Cardinal Utility"; otros autores prefieren nombrar con minúscula la función de Bergson y con mayúscula la de Arrow. Debemos notar que la relación entre el concepto de constitución arrovianno y los varios sentidos que la ciencia política le da es analógica, pero limitada; ciertamente, no puede extenderse más allá del hecho de que ambas tipos de constitución tratan de definir "la estructura básica

del proceso político", vease A. Gorlitz (ed.), (1972), Diccionario de Ciencia Política, "Constitución" (W. Euchner).

- (19) En concreto, la notación que utiliza Arrow procede de A. Tarski, (1941), Introduction to Logic; algunos comentarios del propio autor pueden verse en K. J. Arrow, (1951-1963²), II, y (1967b), "Values and Collective Decision-Making", en: F. Hahn - M. Hollis (eds.), (1979), I.1.
- (20) K. J. Arrow, (1951-1963²), pp. 51-3.
- (21) J Bentham, (1789), A Introduction to the Principles of Morals and Legislation, c. IV. Para un análisis crítico del "cálculo hedonista" en el marco de la teoría de la elección social, O. Höffe, (1975), Estrategias de lo humano, 4.3.2 y 4.3.3.
- (22) Puede encontrarse información sobre este punto en J. A. Schumpeter, (1954), pp. 1144-1165; y sobre el papel del cálculo de utilidades en la teoría neoclásica en M. H. Dobb, (1973), Teoría del valor y la distribución desde A. Smith, c. 7.
- (23) K. J. Arrow, (1951-1963²), p. 63.
- (24) M. H. Dobb, (1969), c. I.
- (25) L. Robbins, (1932), Ensayo sobre la naturaleza y significación de la ciencia económica, c. VI; y (1938), "Interpersonal Comparisons of Utility".
- (26) L. Robbins, (1932), p. 201.
- (27) L. Robbins, (1932), p. 207: "La economía nos procura una técnica para la acción racional".

- (28) N. Kaldor, (1939), "Welfare Propositions in Economics".
- (29) J. A. García Lopez, (1973), p. 119.
- (30) Como consecuencia de ello "el principio de compensación se pregunta solamente si los perdedores podrían ser compensados. (...). Si debe tener lugar la compensación, y en qué manera, y si es que tiene que darse, es una cuestión de distribución sobre la que tiene que consultarse a nuestro sentido moral. (...) el abierto propósito del principio de compensación es separar el problema de la "eficiencia" del de la distribución" (Subrayados y comillas del autor de la cita), E. J. Mishan, (1960), "Un repaso de la economía del bienestar durante el periodo 1939-1959", en J. E. Mishan, (1969), p. 59n.
- (31) J. R. Hicks, (1939), "The Foundations of Welfare Economics".
- (32) T. Scitovsky, (1941) "A Note on Welfare Propositions in Economics"; interesantes comentarios de J. E. Mishan, (1960), en: J. E. Mishan, (1969), pp. 35-7.
- (33) La historia de los principios de compensación se alarga en el tiempo más de lo que aquí exponemos. Un momento importante fue la crítica realizada por P. A. Samuelson, (1950), "Evaluation of Real National Income", donde señalaba la insuficiencia de tales principios demostrando que, desde el punto de vista de la distribución del ingreso, sus resultados podían no ser óptimos. Otro hito importante es I. N. D. Little, (1949), "The Foundations of Welfare Economics". Un resumen de

- todo esto puede hallarse en J. A. García Lopez, (1978), pp. 125-8.
- (34) K. J. Arrow, (1951-1963²), c. IV.
- (35) W. J. Baumol, (1946-7), "Community Indifference".
- (36) Había sido dentro de la teoría de la demanda donde se había estudiado la naturaleza de la elección individual. Obra puntera en aquella época fue la de J. R. Hicks, (1939), Valor y capital. Para una exposición sistemática del estado actual de la teoría, H. A. J. Green, (1971).
- (37) K. J. Arrow, (1951-1963²), pp. 53-54.
- (38) K. J. Arrow, (1967), en: F. Hahn - M. Hollis (eds.), (1979), pp. 111-112; D. J. Mayston, (1974), El concepto de opción social, c. 2.
- (39) A. K. Sen, (1970^b), Elección colectiva y bienestar social, p. 17.
- (40) K. J. Arrow, (1951-1963²), pp. 53-4.
- (41) K. J. Arrow, (1951-1963²), p. 55.
- (42) K. J. Arrow, (1951-1963²), pp. 75-7.
- (43) K. J. Arrow, (1951-1963²), p. 77.
- (44) K. J. Arrow, (1951-1963²), p. 78.
- (45) K. J. Arrow, (1951-1963²), p. 79.
- (46) La posición exacta de la nueva economía del bienestar, así como de otros autores como Samuelson, es más matizada y curiosa. Lo que estos autores entienden por estar "libre de ética" no es que no haya juicios de valor en los procesos económicos, sino que la teoría opera sólo con aquellos en los que hay unanimidad. Pare-

cen pensar que si no hay conflictos interpersonales sobre los valores, estos pueden admitirse "como si" fueran premisas factuales. En sí mismo esto es razonable, y sobre un supuesto similar trabajan las condiciones para las constituciones de Arrow. Ahora bien, lo que es raro es que, a pesar de aceptar los juicios de valor, siguieran proclamando la positividad de la economía. A. K. Sen, (1970^b), p. 79, señala esta paradoja: "los juicios de valor unánimes pueden proporcionar la base para una economía del bienestar muy satisfactoria, pero no porque no sean juicios de valor, sino porque son aceptables por todos". Esta perogrullada ha sido punto de litigio durante una veintena de años. Esto es lo que puede verse en I. M. D. Little, (1957²), c. IV, inmerso en la polémica de aquellos años.

- (47) K. J. Arrow, (1951-1963²), p. 80.
- (48) J. H. Blau, (1957), "The Existence of Social Welfare Functions".
- (49) K. J. Arrow, (1951-1963²), p. 90.
- (50) K. J. Arrow, (1951-1963²), p. 100.
- (51) K. J. Arrow, (1951-1963²), c. IV.
- (52) W. Vickrey, (1960⁵), "Utility, Strategy and Social Decision Rules".
- (53) D. C. Mueller, (1979), Elección pública, p. 199.
- (54) K. J. Arrow, (1951-1963²), III.6.
- (55) K. J. Arrow, (1951-1963²), c. IV, espc. pp. 115 y ss.
- (56) K. J. Arrow, (1951-1963²), p. 148.

- (57) K. J. Arrow, (1951-1963²), p. 151n; sobre los supuestos individualistas de la economía del bienestar, M. H. Dobb, (1969), Intr.; sobre el individualismo como fundamento en justicia económica, vease nuestra scc. IV.1.2.
- (58) K. J. Arrow, (1951-1963²), p. 151.
- (59) K. J. Arrow, (1951-1963²), pp. 152-4.
- (60) K. J. Arrow, (1951-1963²), p. 171.
- (61) Es importante señalar que, a pesar de que Arrow no es consistente en la expresión, la unanimidad de que se está hablando siempre es de valores, no de gustos. K. J. Arrow, (1951-1963²), p. 172.
- (62) D. Black, (1948), "On the Rationale of Group Decision-Making".
- (63) K. J. Arrow, (1951-1963²), pp. 182-3.
- (64) K. J. Arrow, (1951-1963²), pp. 184-6.
- (65) Los antecedentes son: F. H. Knight, (1931), The Ethics of Competition and other essays; G. J. Stigler, (1943), "The New Welfare Economics"; M. H. Dobb, (1933), "Economic Theory and the problems of a Socialist Economy"; A. P. Lerner, (1934), "Economic Theory and Socialist Economy"; M. H. Dobb, (1935), "A Reply"; A. P. Lerner, (1935), "A Rejoinder".
- (66) K. J. Arrow, (1951-1963²), pp. 188-9.
- (67) K. J. Arrow, (1951-1963²), p. 189.
- (68) K. J. Arrow, (1951-1963²), pp. 189-91.
- (69) K. J. Arrow, (1951-1963²), p. 191.

- (70) Como se pondrá de manifiesto en nuestros próximos capítulos, especialmente en III.3-4, el camino de salida del resultado de imposibilidad de Arrow se traza enriqueciendo la estructura informativa del modelo -y con ella la propia noción de racionalidad colectiva- que es muy limitada. A K. Sen, (1979), "Personal Utilities and Public Judgements: or What's Wrong with Welfare Economics", ha demostrado que es precisamente esa estructura la que excluye cualquier regla que no sea dictatorial.
- (71) K. J. Arrow, (1951-1963²), p. 198. Aunque la obra de J. M. Buchanan - G. Tullock, (1962), The Calculus of Consent, trata básicamente de los aspectos positivos de la elección social, en ella se encuentra una manera de seguir estas sugerencias de Arrow, aunque partiendo de supuestos diferentes. La regla de unanimidad que defienden para los procesos constitucionales implica debates, compromisos y enmiendas continuas hasta la obtención de un acuerdo.
- (72) Algunos de estos trabajos son bastante decisivos para orientaciones posteriores de la teoría de la elección social. Entre ellos destacan: A. Bergson, (1954), "On the Concept of Social Welfare"; M. C. Kemp, (1953-4) "Arrow's General Possibility Theorem"; J. M. Buchanan, (1954), "Individual Choice in Voting and the Market"; E. J. Mishan, (1957), "An Investigation into Some Alleged Contradictions in Welfare Economics". El artículo de Little al que dedicaremos nuestra atención es: I.

- M. D. Little, (1952), "Social Choice and Individual Values", en E. S. Phelps (ed.), (1973), pp. 137-52.
- (73) I. M. D. Little, (1952), en: E. S. Phelps (ed.), (1973), pp. 138 y ss.
- (74) I. M. D. Little, (1952), en: E. S. Phelps (ed.), (1973), p. 140.
- (75) K. J. Arrow, (1951-1963²), pp. 227 y ss.
- (76) K. J. Arrow, (1951-1963²), p. 228.
- (77) Th. Guibaud, (1952), "Les theories de l'intérêt general et le probleme logique de l'aggregation"; K. Inada, (1961), "On the Economic Welfare Function".
- (78) I. M. D. Little, (1952), en: E. S. Phelps (ed.), (1973), pp. 144 y ss.
- (79) I. M. D. Little, (1952), en: E. S. Phelps (ed.), (1973), p. 145.
- (80) K. J. Arrow, (1951-1963²), p. 221.
- (81) K. J. Arrow, (1951-1963²), p. 222.
- (82) K. J. Arrow, (1967b), en: H. Hahn - M. Hollis (eds.), (1979), p. 110.
- (83) K. J. Arrow, (1967b), en: F. Hahn - M. Hollis (eds.), (1979), p. 111.
- (84) K. J. Arrow, (1951-1963²), p. 222.
- (85) K. J. Arrow, (1967b), en: F. Hahn - M. Hollis (eds.), (1979), p. 119.
- (86) Este punto es aclarado totalmente cuando Arrow comenta: "estoy plenamente de acuerdo en que los juicios de valor sólo pueden ser sostenidos por individuos -y que, consecuentemente, la ordenación de preferencia

- construida mediante la función de elección social (basada en las preferencias de los individuos) es sólo un medio para tomar decisiones sociales y no es en absoluto una expresión de algún sistema ético interpersonal", K. J. Arrow, (1952), "The Principle of Rationality in Collective Decisions", en: K. J. Arrow, (1984), p. 50.
- (87) I. M. D. Little, (1952), en: E. S. Phelps (ed.), (1973), p. 150.
- (88) Ibidem.
- (89) I. M. D. Little, (1952), en: E. S. Phelps (ed.), (1973), p. 151.
- (90) K. J. Arrow, (1967a), "The Place of Moral Obligation in Preference Systems", en: K. J. Arrow, (1984), p. 79.
- (91) Vease P. Singer, (1973), Democracia y desobediencia.
- (92) Para nuestra exposición de la teoría de J. Rawls seguimos la presentación que este autor hace en J. Rawls, (1971), A Theory of Justice. Sólo de modo marginal nos referiremos a las obras anteriores y posteriores a ésta.
- (93) J. Rawls, (1971), p. 3.
- (94) J. Rawls, (1971), p. 7.
- (95) "Por instituciones más importantes entiendo la constitución política y los principales ordenamientos económicos y sociales", Ibidem.
- (96) J. Rawls, (1971), p. 6.
- (97) Rawls partía en J. Rawls, (1951), "Outline of Decision

Procedure for Ethics", de un modelo inductivista, tomado de las ciencias naturales tal y como las entendió el positivismo, para establecer una relación directa entre "hechos morales" y principios morales. Pero, en "A Theory of Justice" nos encontramos con un modelo constructivista que denota la influencia de Piaget y Quine. El cambio de paradigma se produce cuando aparece el concepto de "sentido de justicia" que vuelve más complicada la relación entre "hechos morales" y principios; vease J. Rawls, (1963), "The Sense of Justice". En cuanto a las falsas analogías, nos referimos a la idea que se hace Rawls de que la teoría moral opera de modo análogo a la lingüística; vease nota (103) infra.

(98) Esta teoría puede encontrarse en J. Rawls, (1971), c. VIII. En ella se pueden apreciar algunas influencias muy marcadas de Piaget y Kohlberg. Nos interesaría destacar sólo que en ella aparece el "sentido de la justicia" como un complejo de aspectos emotivos, volitivos y cognitivos.

(99) J. Rawls, (1971), p. 46.

(100) Esto, en la concepción de Rawls, significa más de lo que parece. Para que el método del equilibrio reflexivo tenga operatividad el sentido de justicia tiene que ser una capacidad natural que cada sociedad construye diferentemente. El modo particular en que Rawls lo ha construido ha dado pie a los críticos a señalar que su teoría de la justicia es, a pesar de sus propias

pretensiones de universalidad, una teoría sesgada ideológicamente por el liberalismo, y que, consecuentemente, no tiene por qué ser aceptada por todo el mundo. Que tal sesgo existe es declarado y admitido por Rawls en sus escritos posteriores, vease J. Rawls, (1982), "Social Unity and Primary Goods", en: A. K. Sen - B. A. Williams (eds.), (1982). Entre los primeros autores que señalaron este sesgo, B. Barry, (1973), The Liberal Theory of Justice.

- (101) J. Rawls, (1971), p. 47.
- (102) Estos dos tipos de argumentos han sido distinguidos por R. Dworkin, (1973), "The Original Position", en: N. Daniels (ed.), (1975), pp. 27-37.
- (103) A pesar de la opinión de R. M. Hare, (1973), "Rawls' Theory of Justice", en: N. Daniels (ed.), (1975), y a pesar de las propias palabras de Rawls sobre Quine y la analogía con la lingüística. Como ha demostrado R. Dworkin, (1973) y E. Nagel, (1973), "Rawls on Justice", en: N. Daniels (ed.), (1975), la teoría moral y las ciencias, incluida la lingüística, utilizan métodos distintos. Rawls acepta esta posición en J. Rawls, (1975), "The Independence of Moral Theory". R. Dworkin, en el artículo citado, señala que la analogía correcta sería con los procesos de deliberación legal sobre demandas nuevas no recogidas en la legislación, R. Dworkin, (1973), en: N. Daniels (ed.), (1975), pp. 33 y ss.
- (104) J. Rawls, (1971), p. 20.

- (105) D. Lyons, (1975), "Nature and Soundness of the Contract and Coherence Arguments", en: N. Daniels (ed.), (1975), pp. 146 y ss.
- (106) R. Dworkin, (1973), en: N. Daniels (ed.), (1975), pp. 36 y ss.
- (107) Puede verse en J. Rawls, (1975) y (1980), "Kantian constructivism in Moral Theory".
- (108) El que así suceda, sin embargo, propone un nuevo problema: ¿es la teoría de Rawls descriptiva y explicativa de las sociedades liberales occidentales, o más bien prescriptiva y justificativa, i. e., válida para cualquier sociedad? Este tipo de cuestiones han sido planteadas por R. M. Hare, (1973) y D. L. Scafer, (1979), Justice or Tyranny? A Critique of J. Rawl's Theory of Justice, quienes responden negándole a la teoría de Rawls cualquier capacidad prescriptiva. No obstante, y sin ánimo de resolver esta cuestión controvertida, podemos tener en cuenta los numerosos pasajes en que Rawls establece un continuo entre descripción y prescripción como un aspecto esencial a su concepción sustantiva de la justicia; vease J. Rawls, (1971), pp. 50-1.
- (109) Aunque en lo que sigue no consideraremos la posición original bajo la perspectiva de que responde a una teoría más profunda (en esencia Kantiana) tal y como la expresa Rawls en la s. 40 de su libro, sino sólo bajo el aspecto más limitado de expresión real de la reflexión moral sobre los principios de justicia, queremos dejar constancia de que compartimos la tesis

de R. Dworkin, (1973), según la cual Rawls no justifica sus principios desde, sino a través de la posición original.

(110) J. Rawls, (1971), p. 18.

(111) Nuestra argumentación no discurrirá por el camino de demostrar que la posición original no cumple estos objetivos. Sin embargo, subscribimos la formalización y discusión de la elección hipotética realizada por R. P. Wolff, (1977), Para comprender a Rawls, IV, así como sus conclusiones y las de aquellos que dudan de que los principios rawlsianos sean la única solución al juego de la elección inicial, como D. Lyons, (1975), en: N. Daniels (ed.), (1975), s. IV.

(112) J. Rawls, (1971), pp. 121-2.

(113) J. Rawls, (1971), p. 21.

(114) Las principales de estas condiciones son: las circunstancias de la justicia, las restricciones formales del concepto de lo correcto (right), el velo de ignorancia y la racionalidad de las partes. Vease cuadro general de la descripción de la posición original en J. Rawls, (1971), s. 25.

(115) D. Lyons, (1975), en: N. Daniels (ed.), (1975), s. III.

(116) J. Rawls, (1971), pp. 12 y 120.

(117) D. Lyons, (1975), en N. Daniels, (1975), p. 159.

(118) Recordemos que la definición de justicia procesal pura exige que no haya ningún criterio independiente al propio proceso de elección para dictaminar lo justo, y que sea el propio carácter del proceso el que

- transmita su fuerza moral a los resultados; J. Rawls, (1971), s. 14.
- (119) J. Rawls, (1971), p. 126.
- (120) De una situación a otra existe el abismo que señala R. Dworkin, (1973), en: N. Daniels.(ed.), (1975), pp. 19 y ss. entre interés antecedente e interés actual. Este mismo tipo de argumento es el que elaboran parcialmente autores como R. Nozick, (1974), Anarchy, State and Utopia, pp. 183 y ss., y J. R. Lucas, (1980), On Justice, c. X.
- (121) J. Rawls, (1971), p. 255.

CAPITULO SEGUNDO
TEORIA DE LA UTILIDAD Y ELECCION SOCIAL

Arrow había hecho la pregunta esencial, "¿Cómo son posibles los juicios de bienestar social?", y había sacado a la economía del bienestar de las discusiones "wertfrei" para enfrentarla al problema real: la elección social y los valores individuales. Sin embargo, el resultado de su trabajo era descorazonador: el teorema de (im)posibilidad de las constituciones aparece con una vigencia universal. El camino entre la subjetividad y la intersubjetividad se declara imposible; con él caía toda esperanza de justicia en el campo de las decisiones colectivas.

Sin embargo, ese resultado depende fuertemente del modelo de elección social diseñado por Arrow. Recordemos que ese modelo es ordinalista; esto es, reduce toda la información que aportan las preferencias individuales a relaciones de orden entre las alternativas. Ello, ya lo vimos, significaba el rechazo de las funciones de utilidad individuales y de las comparaciones interpersonales de utilidad. ¿Es, entonces, extraño que el modelo sea impotente para tratar las cuestiones de justicia? No parece que se pueda decir mucho sobre el valor comparativo de un estado social, si no podemos medir y comparar el

bienestar que reporta a los individuos afectados. Las condiciones de racionalidad que Arrow supone naturales y débiles reflejan una estrecha visión de la racionalidad implicada en los procesos de decisión social. Son ellas las que precipitan el resultado dictatorial y son, por tanto, ellas las que hacen imposible que el procedimiento concuerde con los valores democráticos.

Visto con perspectiva histórica el desarrollo de la teoría de la elección social, parece evidente que las nuevas investigaciones tenían que orientarse hacia la búsqueda de una estructura informativa más rica y otra noción de racionalidad colectiva. Sin embargo, la obra de Arrow no hacía esta dirección nada evidente. El recurso más obvio consistía rescatar el concepto tradicional de utilidad, tal y como ya había hecho la teoría de juegos en los años inmediatamente anteriores a la publicación de "Social Choice". Sobre ese intento, empero, ya se había pronunciado Arrow negándole -en buen paretiano- todo interés para los problemas relativos al bienestar social. Por este motivo, la recuperación del utilitarismo clásico por parte de la teoría de la elección social no parecía que fuera un camino favorable.

En las secciones en que hemos dividido este capítulo nos ocuparemos de cómo esta empresa que parecía improbable se llevó a cabo. Dedicaremos las dos primeras secciones a estudiar la viabilidad normativa de la teoría de la utilidad elaborada en la teoría de juegos -paso obligado para comprender la propia viabilidad del modelo

de elección social utilitarista. En la sección siguiente veremos cómo se puede pasar de la noción de utilidad individual a la de utilidad social -la cardinalización ética del bienestar-, condición imprescindible para hacer operativo un modelo de elección social alternativo al de Arrow. Finalmente, el resto del capítulo está consagrado a la exposición y análisis del modelo cardinalista más completo que la teoría nos ha ofrecido: el utilitarismo de preferencia de Harsanyi. La última sección es una comparación crítica entre el procedimiento contractual de Harsanyi y el de Rawls.

II.1 Las funciones de utilidad en la teoría de juegos.

En 1947 apareció una obra que revolucionó los estudios del análisis formal en su aplicación a las ciencias sociales, "Theory of Games and Economic Behaviour", de J. von Neumann y O. Morgenstern. (1). La teoría de juegos se presentó como una forma muy atractiva de proporcionar una base matemática al estudio de los comportamientos sociales y su ideal -precisar la noción de "comportamiento racional"- era el mismo que el de algunas de las ramas más importantes de la sociología y la economía. De ella, en general, se puede decir que se ocupa de encontrar una solución "óptima" para los tipos de casos en que existan conflictos interpersonales -en su terminología, conflictos entre "jugadores", pudiendo ser éstos personas, aso-

ciaciones, instituciones, etc. (2). La posibilidad de que este tipo de análisis llegue a aplicarse a problemas sociales, especialmente económicos, reside en que muchas de las situaciones conflictivas pueden plantearse como juegos en los que se requiere una estrategia por la que se obtenga una situación aceptable para todas las partes en pugna.

Por lo demás, la validez y utilidad de esta teoría del comportamiento racional proviene de la ductilidad -y, por tanto, de su realismo- para cuidarse de ese tipo de situaciones, ya que sus modelos tienen en cuenta datos que, en principio, parecería imposible formalizar: "(...) que los participantes en un juego poseen inteligencia e informaciones diferentes, que sus expectativas en punto al comportamiento de los demás jugadores son también diversas y que en principio pueden intentar lograr sus finalidades eligiendo entre varios caminos posibles" (3). El refinamiento puede llegar (y es prescriptivo que llegue) hasta considerar que "la posición de un jugador (o un sujeto económico o una empresa) se debilita considerablemente si el oponente descubre sus intenciones" (3). La teoría indica al jugador la forma de protegerse y "el significado que revestirán sus medidas de represalia para los otros jugadores" (3). En este sentido la teoría de juegos es muy atractiva para la economía, puesto que abandona el punto de vista irreal sobre el comportamiento del consumidor según el cual éste es una especie de autómatas que busca sólo la maximización de su cesta de bienes. En cam-

bio la teoría permite considerar a los sujetos económicos en cuanto abiertos a todo tipo de influencias, introduciendo en sus modelos no sólo estas interrelaciones, sino también las modificaciones de éstas durante los procesos de decisión que necesariamente se darán por la acción mediadora, entre el conflicto y el consenso, de la negociación. Por ello, las soluciones de los juegos no se dan para cada parte como un ente aislado, sino que son "aquella distribución o distribuciones de pagos que cada jugador debe hacer en función del comportamiento de los demás" (4).

Hay que dejar anotado que, dada la complejidad de las situaciones conflictivas que el mundo vive actualmente y dado que son precisamente las más complejas las que más interesa resolver, la teoría se enfrenta a un sinfín de problemas matemáticos para hacer llegar sus análisis a todos los fenómenos de la vida social. Esta es su limitación; pero, ello no disminuye en nada su valor que reside fundamentalmente en "su capacidad de poner claramente de manifiesto los distintos tipos de razonamiento que se aplican en los diversos tipos de conflicto" (5).

Si, con este planteamiento, la teoría de juegos parece ya que podría ser un instrumento útil para tender aquel puente y reconciliar lo colectivo y lo individual, la importancia concreta que ha tenido para el desarrollo de la teoría de la elección social radica en haber reformulado de una manera plausible la teoría de la utilidad, de forma que podría servir como ese procedimiento concreto que necesitamos para pasar de una orilla a otra. Para

que la teoría de juegos pueda tener capacidad operativa no sólo se requiere que disponga de modelos formales adecuados, sino también prioritariamente tiene que poder establecer con todo rigor los objetivos de los jugadores. Ello exige que la teoría disponga de un método por el que puedan establecerse las preferencias, fines o actitudes de las personas que intervienen y en forma que sean útiles al teórico. Este método es una renovación de la teoría de la utilidad del benthamismo y se forja en torno al concepto de "función de utilidad".

Una función de utilidad es una cuantificación de las preferencias de una persona con respecto a ciertos objetos. Su papel es asociar a cada situación alternativa un número que refleje la "atracción" de un individuo por cada situación. Pero, hace más, y esto es lo importante: asocia un número a las distribuciones alternativas de probabilidad. Esto es, no sólo dice que la situación x tiene una utilidad 4 para el individuo i , sino que dice la utilidad que tendría para i una situación en la que, con un determinado tanto por ciento de probabilidad, pudiera darse x (técnicamente: "rifa", "lotería", etc.). A esta función de utilidad se exige que sea construida de manera tal que el valor de la misma para cualquier distribución probabilística de ciertas alternativas sea la esperanza matemática de la utilidad, a fin de que se asegure la existencia de un índice único de utilidad (6).

Para poder construir una función de utilidad de von Neumann-Morgenstern las preferencias de los jugadores

han de cumplir ciertas condiciones de consistencia. Estas condiciones, en la enunciación de Luce-Raiffa, son (7):

- (1) Todo es comparable.
- (2) Preferencia e indiferencia son transitivas.
- (3) Un jugador es indiferente cuando premios equivalentes se sustituyen en una rifa.
- (4) Un jugador jugará siempre si las probabilidades son bastante buenas.
- (5) Cuanto más probable sea el premio preferido, tanto mejor será la rifa.

Diseñada bajo estas condiciones de racionalidad, la función de utilidad parece dar de nuevo pie a hablar de métodos que utilicen la agregación para obtener una preferencia social sobre las diferentes situaciones alternativas. Ya vimos que Arrow condena este tipo de métodos por que, en su opinión, no tienen significado alguno para la economía del bienestar. La agregación de utilidades es proscrita no sólo por no satisfacer la condición (C) de independencia de alternativas irrelevantes, sino por la propia definición de la regla de elección social. Esta hace de la ordenación social una función del conjunto de ordenaciones individuales. Si se toman los índices de utilidad como números para los mapas de indiferencia, bajo una regla arbitraria, la suma de los índices de utilidad es entonces una función de las ordenaciones individuales y puede servir para obtener una ordenación social. Pero, puesto que el tipo de ordenación variará dependiendo de la regla que se tome, vulneramos la propia defini-

ción de regla de elección social que exige que la ordenación social dependa única y exclusivamente de las ordenaciones individuales (8), y no de otro tipo de cosas como, por ejemplo, la regla por la que se asocian los números a los mapas de indiferencia. La cuestión es averiguar si esta crítica general es aplicable también a la función de utilidad de von Neumann-Morgenstern.

Arrow no duda en responder afirmativamente a esta cuestión. En primer lugar, al no darle ninguna significación "ética especial a la concreta escala de utilidad hallada", el método puede enunciar "las leyes de la conducta racional del consumidor (...), pero no sirve de nada en consideraciones de bienestar, particularmente si en lo que estamos interesados es en realizar elecciones en las que no entran elementos aleatorios. Lo contrario sería tanto como decir que la distribución del ingreso social tiene que estar dominada por las inclinaciones de los individuos por el juego" (9).

En segundo lugar, existe el mismo problema de la arbitrariedad de la escala de utilidad. Arrow lo explica con el ejemplo siguiente (10). Supóngase que para cada individuo hay siempre una alternativa que es preferible o indiferente a todas las demás alternativas y que hay otra respecto de la cual todas las demás son preferibles o indiferentes. Entonces, puede definirse de una manera única el índice de utilidad para esa clase asignando la utilidad 1 a la mejor alternativa y 0 a la peor alternativa. Supóngase tres individuos y tres alternativas, de

forma que:

$$A: 1(x) , 0,9(y) , 0(z)$$

$$B: 1(x) , 0,9(y) , 0(z)$$

$$C: 0,5(x) , 1(y) , 0(z)$$

Dado que la suma de utilidades arroja el siguiente saldo

$$2,5(x) , 2,8(y) , 0(z)$$

según el criterio de utilidad, y es preferible a todas las demás. La alternativa z es despreciable puesto que nadie la desea; si la eliminamos no se modificaría ese resultado. Pero, si la eliminamos, según la regla de asignación de utilidad a las preferencias, tendríamos otra distribución de utilidad:

$$A: 1(x) , 0(y)$$

$$B: 1(x) , 0(y)$$

$$C: 0(x) , 1(y)$$

Con lo que ahora obtendríamos la siguiente suma:

$$2(x) , 1(y)$$

y consiguientemente que x es preferida a y . La arbitrariedad de la regla hace fracasar todo el proyecto de hacer a la suma de utilidades una regla de elección social en el sentido de Arrow. Ello permite declarar a este autor que la idea de cardinalización de las preferencias no tiene ningún sentido.

Por último, aún siendo éste el mayor problema, si fuera resuelto, la teoría de la utilidad tendría aún que enfrentarse al problema inveterado de las comparaciones

interpersonales de utilidad. Sobre este tema Arrow se declara igualmente escéptico; no hay manera, en su opinión, de efectuar las comparaciones que son necesarias, a menos que se encuentre un "juicio de valor" concreto, no deducible de las sensaciones individuales, para hacer compatibles en el mismo plano las utilidades de diferentes individuos y otro juicio de valor más para hacer la agregación en cualquier fórmula matemática concreta" (11). A la resolución de esta problemática compleja en torno a la viabilidad de una función de utilidad social se entregaron los teóricos de la elección social que adoptaron el camino señalado por la teoría de juegos.

II.2 La deficiencia ética de la teoría de juegos.

Las críticas de Arrow coinciden en señalar la deficiencia normativa de la teoría de la utilidad y su incapacidad para hacerse cargo de los valores que son imprescindibles para los juicios de bienestar social. Sin embargo, sus defensores señalan la capacidad decisoria de la teoría desde el punto de vista prescriptivo y se ha hecho famosa la frase con que Braithwaite tituló una conferencia sobre el tema: la teoría de juegos como una herramienta para el filósofo moral (12). La cuestión es tratar de determinar qué hay de capacidad prescriptiva, si la tiene, en la teoría de juegos, ¿puede ser realmente un instrumento para el filósofo moral?

La pregunta que propone Braithwaite en las primeras páginas de su artículo y para cuya respuesta considera un instrumento adecuado la teoría de juegos es: "¿Puede el filósofo moral dar algún consejo a la gente que tiene objetivos diferentes con respecto a cómo pueden colaborar en tareas comunes de forma que obtengan la máxima satisfacción compatible con una distribución imparcial (fair)?"

(13) En esta pregunta tenemos enunciados todos los elementos del tema de la fundamentación de juicios de valor reunidos para situaciones de elección social. En primer lugar, lo que está en juego es la posibilidad básica y simple de hacer algo más que meros juicios técnicos; se trata de poder recomendar fundadamente cursos de acción sociales que inevitablemente presuponen, conllevan y producen juicios de valor. Mas, como en la conocida situación de Alicia -las recomendaciones varían según a dónde quiera uno conducirse-, tenemos a personas con objetivos diferentes y existe, en consecuencia, un conflicto no sólo de deseos, sino también de recomendaciones. Pero, se trata de encontrar una recomendación única para las partes en conflicto, puesto que la sociedad se entrelaza de tareas comunes. Entonces, la pregunta que se propone es: ¿qué condiciones ha de cumplir una recomendación -un juicio de valor- para que sea aceptada como única por las partes litigantes? Según el modelo que estudiamos, la recomendación tiene que cumplir una condición de racionalidad -máxima satisfacción para los litigantes- y una condición ética -imparcialidad en la distribución de las car-

gas y beneficios. Del contrapeso entre satisfacción personal y justicia social nace la recomendación bien fundada que estamos buscando. Pues bien, la teoría de juegos se ofrece como un instrumento adecuado para encontrar el punto de equilibrio entre racionalidad y moralidad que haga posible esa recomendación. Retraducida a estos términos, la cuestión inicial dice: ¿puede la teoría de juegos no sólo predecir la conducta racional de los jugadores, sino además prescribir la conducta justa de los jugadores?

Braithwaite está convencido de que ello es posible; que la teoría de juegos puede proporcionar imperativos hipotéticos que hagan posible la aplicación de la ética utilitarista: "(...) las recomendaciones constituirán lo que pueden llamarse principios morales de segundo orden que dan criterios de buen sentido, prudencia e imparcialidad (...)" (14), y ello con la ventaja suplementaria de que no se hacen comparaciones interpersonales a priori ni juicios de valor sobre los códigos morales de los jugadores.

Para explicar el modus operandi de la teoría, Braithwaite presenta un ejemplo sencillo. Lucas y Mateo viven en apartamentos contiguos y ambos tocan un instrumento, de forma que si ambos tocan al mismo tiempo, ninguno de los dos disfruta de su música. Por lo demás, ambos sólo pueden tocar de 9 a 10 de la mañana y no hay otro tipo de interferencias acústicas exteriores. El cuadro de la situación es el siguiente según el número de mañanas que

cada uno toca su instrumento:

		Mateo	
		Toca	No toca
Lucas	Toca	11	12
	No toca	21	22

El problema, entonces, es: "¿Puede idearse algún principio plausible que diga cómo habrían de dividir la proporción de días en que ambos tocan (11), Lucas toca solo (12), Mateo toca solo (21) y ninguno toca (22), de forma que se obtenga el máximo de satisfacción compatible con una distribución imparcial?" (15).

Para empezar hay que asociar números a la posibilidad de que resulte una cosa u otra. Para ello necesitamos una regla. Suponemos que Lucas es indiferente entre

- (a) ambos no tocan, y
- (b) que Mateo toque en $1/3$ de las mañanas y ambos toquen $2/3$ de las mañanas;

y, a su vez, es indiferente entre (a) y

- (c) que él toque en $1/6$ de las mañanas y que ambos toquen en $5/6$ de las mañanas.

Con ello se establece un conjunto de números -la preferencia del resultado 21 sobre 22 es doble que la del resultado 22 sobre 11, y la preferencia de 12 sobre 22 es cinco veces la de 22 sobre 11- que nos ofrece la escala de preferencias de Lucas sobre las alternativas y so-

bre su distribución probabilística. Si operamos del mismo modo con Mateo, y obtenemos su escala de preferencias, alcanzamos una base para hacer recomendaciones de estrategias de forma que se obtenga un punto de máxima satisfacción para ambos. Con una restricción, empero: siempre que los jugadores estén en una situación competitiva pura. El teorema minimax de von Neumann-Morgenstern demuestra que en todo conflicto de dos jugadores puramente competitivo hay un punto en el que ambos jugadores minimizan el daño máximo; por consecuencia, ambos jugadores han de seguir la estrategia que conduce a este resultado, si no quieren empeorar su situación y mejorar la de su adversario. Esta es una recomendación que cumple la condición de racionalidad antes señalada, y para los juegos clasificados como bi-personales y de suma cero (i. e., puramente competitivos), el cumplimiento de esa condición basta para hacer la recomendación oportuna.

Mas, ¿qué sucede si el juego es de suma no nula, i. e., ni puramente competitivo ni puramente cooperativo? Como señala Braithwaite, el teorema del minimax indica qué hacer con esa parte competitiva; pero, no dice nada sobre ese excedente de cooperación. Porque lo que hay que decir, apostillamos nosotros, no es cuál es el beneficio minimax, sino cómo ha de dividirse la cooperación para que pueda decirse de ella que es "imparcial" (16).

Braithwaite propone una representación parabólica para determinar el punto de imparcialidad. El método es más bien largo de explicar y no es relevante para nuestra

discusión. Gracias a él se obtiene, para el caso presentado, la siguiente recomendación: de 43 mañanas, 17 tocará Lucas solo, 23 tocará Mateo solo. La razón por la que Mateo tiene esta ventaja sobre Lucas proviene de sus escalas de preferencia; en ellas se refleja el que Mateo prefiere que ambos toquen a la vez a que ninguno toque, mientras que Lucas prefiere el silencio a la cacofonía. De este modo, en el cálculo entra como elemento decisivo la amenaza que Mateo puede ejercer sobre Lucas de tocar y producir la odiosa cacofonía, arma que no está a disposición de Lucas.

A este resultado es al que hay que plantear la pregunta inicial: ¿qué capacidad prescriptiva tiene la teoría de juegos? Braithwaite cree que se ha conseguido un resultado que tiene ese carácter especial de las recomendaciones justas de ser una distribución imparcial de los días de satisfacción. Y, ciertamente, dadas las preferencias de los jugadores, y la capacidad de amenaza de cada uno, es probable que aceptaran este resultado. Sin embargo, ¿es ello suficiente para decir que esa solución es "justa"? Queremos decir que la realidad en la que se plantean las situaciones como las que analiza Braithwaite incluyen como datos, quizás incambiables, las capacidades de maniobra, amenaza, poder, etc. de los jugadores, y que sobre esos datos el teórico apunta cuál sería la solución racional para ellos. Pero, ¿es esa solución, emanada de los datos incuestionables, la justa? ¿No hay aquí una subrepticia sumisión de lo justo a lo racional,

entendido éste como maximización del beneficio sin más?

En la tesis de Braithwaite hay una neta diferenciación de las dos condiciones, pero al final de su exposición parece que ha terminado haciéndolas equivalentes. Parece que los teóricos del juego sólo pueden defender el aspecto prescriptivo de la teoría al precio de subsumirlo bajo el aspecto predictivo. Quizás ello les sea inevitable desde el momento ^{cu que} -señala Sen- "la importancia asignada (...) a las ventajas de la amenaza y la evitación completa de comparaciones interpersonales parecen excluir toda una clase de juicios éticos que son relevantes para la elección colectiva" (17). A pesar de lo que parecen pensar estos teóricos, los juicios a los que se refiere Sen no son algo que se imponga desde fuera -y, por eso, deben pensar que no hay que tenerlos en cuenta- a la elección social, sino que desde siempre están allí en la medida en que están presentes en las escalas de preferencia de los individuos. Es un punto de vista poco interesante el considerar que nos han de ser indiferentes -a nosotros, teóricos- las ordenaciones de preferencias de los individuos distintas de aquellas que hacen por su mero interés personal. Arrow ya mostró que en la medida en que son indistinguibles las consecuencias de un curso de acción para mí y aquellas que lo son sólo para la sociedad, era más conveniente partir de las preferencias según juicios de valor que de las preferencias según gustos.

Por lo demás, hay que señalar que si la teoría de juegos no tiene capacidad prescriptiva también se debe

-como se desprende de los comentarios de Höffe (18)- a que sus modelos no permiten hacerse cargo de los valores sociales. Lo cual también es inevitable, puesto que si los tuviera en cuenta, las situaciones conflictivas que analiza ya no lo serían tanto o no lo serían en absoluto, haciendo sus análisis irrelevantes.

En definitiva, mientras la teoría de juegos parece un instrumento útil en la medida en que se mantiene dentro de los límites de las recomendaciones predictivas de conductas racionales, tratar de hacerlas pasar por recomendaciones normativas para las elecciones sociales es, por un lado, llevar a la teoría a un medio en el que le es imposible vivir; y, por otro, constituye un abuso teórico que explota las insuficiencias y ambivalencias del conocimiento actual de las condiciones de racionalidad y justicia.

II.3 Condiciones éticas para la cardinalización del bienestar.

A pesar de estos resultados negativos, la teoría de juegos, a través de las funciones de utilidad, abrió un camino que los teóricos de la elección social exploraron como una alternativa al resultado de imposibilidad de Arrow para la fundamentación de los juicios de bienestar social.

El artículo pionero de Flemming (1952) explora la

posibilidad de asignar números al bienestar social obtenible con las diferentes situaciones alternativas, así como a los diferentes bienestares que los individuos obtendrían en esas situaciones, "de tal forma que asegure que el bienestar social es la suma de los bienestares individuales" (19). Ahora bien, la posibilidad de obtener un concepto cardinal de bienestar depende de un sistema ético; i. e., es un sistema ético el que tiene que configurarse de tal manera que provea una forma de medir el bienestar como parte del proceso de evaluación ética que proponga. Para ello, es imprescindible que el sistema ético contenga una propiedad -bienestar, utilidad, felicidad, etc.- que sea cardinal, i. e., medible. ¿Cómo se sabe si una propiedad ética es cardinal? Si cumple las siguientes condiciones (20):

- (1) La propiedad ha de surgir en relaciones ordinarias de pares de objetos: A es más x que B.
- (2) La relación ha de ser asimétrica: A es más x que B; B es menos x que A.
- (3) La relación ha de ser transitiva: A es más x que B, B es más x que C, A es más x que C.
- (4) Si A es igual x que B, y A es más x que C, entonces B es más x que C; si A es igual x que B, y B es igual x que C, entonces A es igual x que C.
- (5) La relación ha de poder expresarse en una simple ecuación numérica o ley; esto conlleva la exigencia de reducir la cantidad de indicadores

numéricos a un mínimo, respetando sólo que se distingan por su origen y la proporción de la escala.

(6.1) Entre los objetos a los que pertenece la propiedad x tiene que haber alguno que, aunque sea por convención, pueda ser considerado como una o combinación de partes que son capaces de existir como objetos independientes.

(6.2) Un procedimiento de medición existe donde la propiedad x del complejo de objetos puede expresarse como la suma de la x -idad de sus partes.

El interés de estas condiciones de mensurabilidad de las propiedades éticas no sólo reside en el avance que suponen para el tratamiento cuantitativo del bienestar, sino también por que trazan los contornos del tipo de ética y del tipo de fundamentación que se está buscando.

Pero, sólo los contornos. Porque el estudio de Flemming lo que trata de establecer es "cómo habría de ser"; i. e., las condiciones formales que se requeriría que un sistema ético satisficiera para contener una propiedad mensurable útil para las cuestiones de la economía del bienestar. En este sentido, Flemming declara que él considera a la economía del bienestar como "wertfrei", por cuanto los sistemas de fines le están dados. Pero, sólo en ese sentido, puesto que reconoce -y es su supuesto de trabajo- que el bienestar es un valor, que la cardinalización que propone depende de presupuestos éticos y , consecuentemente, que él está diseñando el perfil de un sistema ético. Sobre este particular, quizás con

razón, observa el autor que si algunos sistemas éticos son más aptos que otros -entre estos estarían aquellos que subrayan la corrección moral-, se debe sólo a que se preocupan más por el bienestar que los segundos y, para ellos ha de ser naturalmente más útil el análisis económico (21).

La cuestión, por tanto, es encontrar las condiciones que un sistema ético tiene que cumplir para disponer de un concepto de bienestar que cumpla los requisitos anteriores de cardinalidad. Tales condiciones -que serán postulados éticos expresados en el lenguaje económico de las preferencias- son presentados por el autor no como condiciones mínimas, sino como el tipo de condiciones a las que un sistema ético se tendría que conformar para que su concepto de bienestar fuera medible. De todas formas, Flemming no guarda sus cartas y, en este punto, señala que está pensando, al diseñar esas condiciones, en "los sistemas éticos que personalmente considero más respetables intelectualmente" (21); estos parecen ser las diversas formas de utilitarismo, pero particularmente el utilitarismo ideal de G. E. Moore. Teniendo esto en mente, los postulados éticos básicos son los siguientes (22):

Postulado A (asimetría de la preferencia social):

Si desde un punto de vista social la situación \underline{x} es preferida a la situación \underline{y} , entonces \underline{y} no es preferida a \underline{x} .

Postulado B (transitividad de la preferencia social):

Si desde un punto de vista social, \underline{xPy} e \underline{yPz} , entonces \underline{xPz} .

Postulado C (transitividad de la indiferencia social): No hay indeterminación con respecto a la relativa deseabilidad de las situaciones: Ni xPy , ni yPz ; ni yPz , ni zPy ; entonces, ni xPz , ni zPx .

Con estos tres primeros postulados se asegura que la preferencia social determina un ordenamiento completo de las situaciones sociales que, como vimos, es una función de bienestar social de Bergson.

Postulado D (relación positiva de las preferencias sociales con las preferencias individuales): Si para un individuo i , $xP_i y$, entonces desde un punto de vista social, xPy . Este postulado expresa un juicio de valor individualista y es el que da todo el carácter al perfil del sistema ético.

Postulado E (independiente evaluación de la distribución de la utilidad entre cada par de individuos): "Sobre resultados con respecto a los cuales los intereses (preferencias) de dos individuos entran en conflicto, quedando inafectados los intereses de los otros individuos, la preferencia social habría de depender exclusivamente de la comparación de la importancia social relativa de los individuos concernidos" (23).

Este último postulado es una extensión del juicio de va-

lor individualista del Postulado D, y ambos establecen la exigencia de que la elección social dependa sólo de los intereses individuales que estén afectados. Por lo demás, las comparaciones interpersonales que en este postulado se introducen, se realizan bajo un juicio de valor que da la pauta única para la comparación y que establece ^{que} ponga en obra este procedimiento de elección social.

Al postulado E hay que defenderlo de una objeción que se basa en un malentendido. Al sostener que la justicia o injusticia de una distribución de renta depende sólo de los individuos afectados, puede surgir la sorpresa entre los que piensan -correctamente- que la justicia de una distribución sólo puede establecerse si se mira al conjunto de los miembros de la sociedad con los que, por indirectos sean los lazos, están unidos los sujetos afectados. Desde este punto de vista, sería correcto el rechazo del postulado E por no adecuarse a las exigencias naturales de la justicia.

Ahora bien, no es esto lo que exige el postulado E; éste no nos pide que separemos a los individuos y no tengamos en cuenta las alzas y las bajas de cada uno de ellos, de las economías y deseconomías externas. Lo que el postulado dice es que "una vez que estos efectos han sido tenidos en cuenta, la distribución de la renta en el resto de la sociedad no ha de tener ulterior influencia en nuestro juicio de valor" (24). Esto significa que hemos de trabajar sobre niveles dados e invariantes de

renta, los cuales se establecen una vez que se ha operado con todos los factores de inestabilidad. Por lo demás, esta exigencia no es incompatible con el tener en cuenta las distribuciones del resto de la sociedad, y el postulado E es consistente con ello, puesto que de lo que se trata es de no alterar las correspondencias interpersonales.

Después de la exposición de los postulados, a Fleming sólo le queda encontrar los indicadores de bienestar para un sistema ético que los satisfaga, de forma que con tales indicadores se puede proceder sumando los bienestares individuales para obtener el bienestar social (25). Concluye su trabajo con la afirmación de que dichos Postulados son en esencia similares a los de von Neumann-Morgenstern, pero más operativos que ellos en las situaciones de elección social (26).

En su primer artículo de 1953 (27), Harsanyi defiende la utilidad de la cardinalización del bienestar de críticas del tipo de las de Arrow formuladas en base a la mayor utilidad de la ordinalidad como medida. Su defensa parte de la utilización que se había hecho de la cardinalidad en la teoría de von Neumann-Morgenstern para elecciones bajo riesgo. La cuestión que trata de resolver Harsanyi en este artículo es si realmente son equivalentes los conceptos de utilidad de la economía del bienestar y los de la teoría de juegos. El punto era crucial, puesto que de su resolución dependía que los resultados de la teoría de juegos pudieran ser un punto de partida

útil para la teoría de la elección social

Algunas opiniones muy influyentes habían denegado tal equivalencia. Así Friedmann y Savage (28) señalaban las propiedades opuestas que la función de utilidad cardinal tiene en las dos teorías (29). Mientras la utilidad en economía del bienestar se rige por la ley de la utilidad marginal decreciente (i. e., con el aumento de la renta decrece su utilidad marginal), en el caso de las situaciones de riesgo en la teoría de juegos la utilidad sigue aumentando durante un periodo extenso, dado que los jugadores siempre están dispuestos a apostar más si hay a la vista una ganancia abundante. Para Harsanyi, sin embargo, el asunto no se concluye con ello, puesto que hay una manera de ver el bienestar que lo hace razonablemente apto para la cardinalidad.

La característica de los juicios de valor y, por tanto, de los juicios de bienestar social, es ser "juicios de preferencia impersonales y no egoistas" (30). Harsanyi explica que ésta es la característica relevante, puesto que si no fuera por ella los juicios de valor sólo serían aceptables por el que los enuncia; el auditorio ante juicios parciales jamás podría sentirse afectado por ellos y, en consecuencia, llegar a compartirlos. Esta característica -ya enunciada por la escuela sentimentalista escocesa del XVIII- es la que permite que las situaciones de elección en economía del bienestar sean situaciones de riesgo e incertidumbre.

Muy ligeramente esboza aquí Harsanyi su moderna elaboración de la teoría moral del observador imparcial.

Un individuo, explica, si tiene que hacer un juicio de bienestar tiene que situarse en una actitud de ignorancia con respecto a su posición en las distintas alternativas sociales; para ello ha de considerarse como teniendo una igual probabilidad de ocupar todas las posiciones sociales en cada una de las situaciones alternativas. Esto, claramente, es una elección bajo riesgo y, como vemos, viene exigida por la propia naturaleza -la impersonalidad- de los juicios de valor. De aquí que Harsanyi concluya que "la utilidad cardinal "maximizada" por los juicios de valor sobre el bienestar social y la maximizada en elecciones que implican riesgo pueden ser consideradas como fundamentalmente basadas en el mismo principio" (30). Reconociendo y teniendo en cuenta las diferencias de propiedades que tienen la utilidad las dos teorías, ello no es un obstáculo para que puedan considerarse equivalentes en lo fundamental: miden el valor que los individuos atribuyen a la posibilidad incierta de ganancias alternativas.

Vemos, con ello, que Harsanyi sigue la brecha abierta por Flemming, pero piensa que la cardinalización del bienestar recibe un apoyo más sólido si se realiza para perspectivas inciertas. Para ello, encuentra un modelo útil en los postulados de cardinalización de Marschak (31), más sencillos que los de von Neumann-Morgenstern y que además producen un postulado ético individualista más débil que el de Flemming. Este postulado es el que da carácter al modelo de Harsanyi y dice así: "Si dos perspecti-

vas P y Q son indiferentes desde el punto de vista de todos los individuos, también son indiferentes desde el punto de vista social" (32). Este postulado basta para determinar una función de bienestar social cardinal (que satisface las condiciones de racionalidad de Marschak) que es la media aritmética de las preferencias individuales (que, asimismo, satisfacen aquellas condiciones de racionalidad). Veamos a continuación los resortes de este logro.

II.4 El modelo de Harsanyi: impersonalidad y comparaciones interpersonales.

La sección 4 del artículo de Harsanyi de 1955 que vamos a comentar es muy interesante por cuanto muestra la distancia existente entre la teoría de la agregación de utilidades de los clásicos y lo que en aquel momento están tratando de elaborar los teóricos de la elección social. La distancia y la diferencia apreciables son consecuencia del impacto que la crítica de Arrow tuvo sobre la idea tradicional de cómo habría de diseñarse el procedimiento de agregación; el modelo de Harsanyi acusa recibo de esta crítica y adopta los elementos que ya eran ineludibles para el planteamiento correcto del problema.

El núcleo de esta diferencia radica en que la dicotomía entre lo individual -la utilidad- y lo social -el bienestar- ha dejado de ser clara o, mejor, no problemática. El concepto de suma que definía la preferencia so-

cial es problemático; el concepto introspectivo de utilidad individual es problemático. Este es el efecto de la crítica arroviana que ha acercado -borrado su precisión- los conceptos de bienestar social y de utilidad individual.

Por un lado, el bienestar social ha perdido su status de pura cantidad objetiva; por otro, la utilidad individual ya no es una noción tan subjetiva. Se podría decir que han nacido nuevos conceptos que llevan el mismo nombre que los viejos. Así, el bienestar social -antes uno y único para todos los individuos-, se concibe como un juicio que cada individuo puede formular y en el que, legitimamente, puede discrepar con sus conciudadanos; el bienestar social es ahora 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 ha pasado a ser un reflejo de la situación económica de todos los individuos afectados; aún más, depende para su formulación de esta interrelación. Así tenemos que el individuo puede tener legitimamente tanto funciones de bienestar social como funciones de utilidad individual, y tanto unas como otras reflejan un juicio personal de ese individuo que depende de las condiciones personales de todos los individuos. Los bloques monolíticos que formaban la objetividad del bienestar y la subjetividad de la utilidad se han fracturado y la confusión parece amenazar la capacidad de la teoría.

Harsanyi, sin embargo, está dispuesto a sacar provecho de ese acercamiento e interpenetración para realizar el objetivo de la teoría de la elección social: la construcción de la intersubjetividad. El primer paso en esta dirección consiste en no consentir que se borre toda diferencia entre lo social y lo individual; se trata de evitar que la preferencia social de un individuo se prescriba como la preferencia de la sociedad. Lo que las decisiones sociales democráticas requieren es que la preferencia social refleje las preferencias individuales de los miembros de la sociedad. Si esta condición en el modelo de Arrow aparecía estrictamente enunciada como una condición de racionalidad de las constituciones -la condición de no dictadura-, en el modelo de Harsanyi aparece como una condición previa de imparcialidad. Para que una función de bienestar social formulada por un individuo sea tal cosa -y no la expresión particular de los juicios de valor de ese individuo- ha de ser construida bajo un "compromiso imparcial" (fair), según el cual la función de bienestar social tiene por argumentos las funciones de utilidad de todos los individuos de la sociedad (34). La característica básica, pues, de este procedimiento es que implica una acción contractual constitutiva del mismo; los individuos contratan entre sí que en sus juicios sociales todos serán tenidos en cuenta y que nadie tendrá un peso superior al de los demás.

La exigencia que el compromiso de imparcialidad lanza sobre el procedimiento -que el juicio individual ten-

ga en cuenta igualmente a todos los miembros de la sociedad- obliga a operar una distinción entre los juicios que formula un individuo. La distinción separa dos ámbitos: lo personal y lo impersonal. La función de utilidad de un individuo se realiza en base a sus propias preferencias personales (que, para Harsanyi, incluyen tanto las hechas en base a gustos como a valores); refleja, pues, una "preferencia subjetiva" y son las preferencias reales del individuo, "mostrando una actitud egoísta en el caso de un altruista" (35). En cambio, las funciones de bienestar social reflejan lo que prefiere "sólo en aquellos posiblemente raros momentos cuando se fuerza a sí mismo a adoptar una actitud especial impersonal" (35). Estas son sus "preferencias éticas" y reflejan aquel compromiso de imparcialidad constitutivo de los juicios de valor social. Sólo las preferencias éticas cumplen los postulados éticos señalados y, estos, por tanto, definen exactamente el tipo de imparcialidad con el que el individuo se compromete. De este modo, se construye el puente entre la subjetividad y la intersubjetividad de los juicios de valor de un individuo; los postulados excluyen las preferencias subjetivas e indican la forma de la función de bienestar social que refleja la actitud de imparcialidad que la intersubjetividad de las decisiones sociales requiere (36).

La imparcialidad, por tanto, se cumple como impersonalidad. Pero, ¿qué es "preferir impersonalmente"? En la definición de esa actitud ética imparcial-impersonal

es donde se evidencia la ventaja de trabajar con perspectivas inciertas (consecuentemente, lo que le debe el modelo de elección social de Harsanyi a la teoría de juegos):

"(...) las preferencias de un individuo satisfacen este requisito de imparcialidad si ellas indican qué situación social elegiría si él no supiera cuál sería su posición personal en la nueva situación elegida (y en cualquiera de sus alternativas), sino que más bien tuviera igual posibilidad de obtener cualquiera de las posiciones sociales que existen en esta situación, desde la más alta hasta la más baja" (37).

Si la teoría de juegos no hubiera elaborado un modelo para abordar decisiones bajo incertidumbre, no hubiera sido posible pensar un procedimiento de este tipo para enlazar subjetividad e intersubjetividad. La actitud que genera el no saber-a-ciencia-cierta -la equiprobabilidad- es la que nos hace impersonales y, en consecuencia, la que configura nuestras preferencias de forma que puedan responder a aquel compromiso de imparcialidad que hace posible la existencia de las funciones de bienestar social.

La equiprobabilidad genera una actitud ética y hace posible que todo individuo tenga "un criterio objetivo para comparar las utilidades de sus semejantes entre sí y con la suya" (38). La equiprobabilidad es la piedra angular del modelo de Harsanyi porque es la que permite dar un origen común a la reconstrucción de las funciones

de utilidad de otros individuos y a la comparación de esas funciones de utilidad entre sí. Una vez obtenida así la escala de preferencias individuales, ya se puede construir la función de bienestar social que, por satisfacer los postulados de racionalidad de Marschak y el postulado individualista de Harsanyi, tiene una alta probabilidad de que sea la misma para todos los individuos. Así se alcanza un juicio de bienestar social que es válido para toda la sociedad.

Por último, la equiprobabilidad también define la forma de la función de bienestar social y, en último término la fórmula del criterio de justicia social. La utilización de las condiciones de racionalidad de Marschak -hoy ya condiciones de racionalidad bayesianas- ya implica que el bienestar social es una función cardinal aditiva de las utilidades individuales. La imparcialidad que define la equiprobabilidad hace innecesario que esas utilidades estén ponderadas, puesto que todas ellas se construyen desde un punto de origen común. Finalmente, la actitud incierta de la equiprobabilidad, la igual posibilidad de estar en la posición de cualquiera, termina de concretar esa función de bienestar social como "la media aritmética de las utilidades de todos los individuos de la sociedad" (38). Por consecuencia, entre dos estados sociales alternativos se recomendará la persecución de aquel que maximice la media aritmética de las utilidades de todos los individuos.

Ahora bien, ¿cómo son posibles las comparaciones interpersonales de utilidad? ¿Es realmente factible poseer un criterio objetivo para las comparaciones interpersonales que exige la equiprobabilidad? Harsanyi no tiene dudas acerca de que dichas comparaciones sean posibles; para él, el problema se reduce a saber "si hay alguna base lógica para tales comparaciones" (39). Cualquier individuo tiene a su disposición dos indicadores para saber la utilidad que los demás dan a las situaciones: las preferencias reveladas en las elecciones efectuadas y las expresiones de satisfacción (40). Las objeciones que se hacen a la teoría de la preferencia revelada señalan que esos indicadores no permiten comparar la utilidad que cada persona adscribe a cada situación. Las razones en las que se fundan estas objeciones suelen ser de dos tipos: metafísicas y psicológicas. Las primeras, en realidad, son fáciles de desbaratar. Objetan la diferencia radical de todos los individuos entre sí, de forma tal que, aunque manifiesten lo mismo, la expresión puede provenir de estados mentales completamente diferentes. Para Harsanyi, razonablemente, sostener esta postura es condenarse al solipsismo más estéril: "Si dos objetos o seres humanos muestran una conducta observable similar en todos sus aspectos relevantes, la presuposición de alguna diferencia entre ellos inobservable y oculta ha de ser considerada como una hipótesis completamente gratuita y contraria al verdadero método científico" (41). Ello, según nuestro autor, justifica que las preferencias reveladas sean uti-

lizadas como indicadores de la utilidad que personas diferentes adjudican a situaciones similares.

Las objeciones psicológicas son más serias: "Si un individuo prefiere la situación X a la situación Y, mientras otro prefiere Y a X, ¿sucede así por qué el primero le da una mayor utilidad a la situación X, o porque le da una menor utilidad a la situación Y, de lo que lo hace el último -o es quizás el resultado de ambos factores a la vez? Y, de nuevo, si en una situación dada un individuo da más signos de satisfacción o insatisfacción que otro, ¿sucede así porque el primero siente una satisfacción o insatisfacción más intensa, o sólo porque se inclina a dar mayor expresión a sus sentimientos?" (42). Nótese que el tipo de objeción es diferente a la metafísica. No se dice que similares preferencias reveladas pueden provenir de interioridades diferentes; sino que hay una multitud de preferencias reveladas en base a una multitud de diferencias psicológicas. La cuestión es cómo hacer para encontrar una medida común que nos permita apreciar estas diferencias como parte de un mismo sistema. El problema no es banal, pero admite una solución por contrastación empírica; se puede averiguar cómo crecen y decrecen estas satisfacciones y averiguar de ese modo a qué atenernos con las diferencias psicológicas. Aunque este método está abierto a muchas incertidumbres, dado que nuestros conocimientos de psicología son muy inseguros, lo que sí parece firme es que "las comparaciones interpersonales de utilidad no son juicios de valor basados

en algún postulado ético o político, sino más bien en proposiciones factuales basadas en ciertos principios de "lógica inductiva" (43). No es esto una desestimación de la crítica de Robbins a las comparaciones interpersonales. Según Harsanyi, esta crítica es válida. De lo que se trata más bien es de separar las comparaciones interpersonales que contienen algún postulado ético o político de aquellas que se basan en una mera información factual.

La palabra mágica es ésta: información factual. Las comparaciones hechas sobre preferencias reveladas han de estar basadas en una información empírica lo más fiable posible, de forma que se desaloje cualquier postulado convencional -que, por lo demás, sólo cumple la función de salvar las deficiencias de conocimiento real. Admitir estos postulados sería reducir las escalas de preferencias éticas a escalas subjetivas; asimismo, significaría un menoscabo a la imparcialidad y al individualismo débil admitido en la construcción de la función de bienestar social:

"Cuánto más completa sea nuestra información factual y más completamente individualista sea nuestra ética, más convergirán las funciones de bienestar social de los diferentes individuos hacia la misma cantidad objetiva, a saber la suma no ponderada (o, mejor, la media aritmética no ponderada) de todas las utilidades individuales" (44).

Estamos así, pues, ante un modelo que confía en la propia capacidad de los individuos de sobreponerse a sus parti-

cularismos y egoismos para salvar los problemas que presenta la construcción de un esquema que dé una meta a la comunidad y asegure el bienestar de todos.

II.5 El modelo de Harsanyi y el teorema de posibilidad de Arrow.

La cuestión que pende sobre la teoría de la utilidad como base para la elección social tal y como es formulada por Harsanyi es saber si es capaz de burlar la universalidad del resultado de imposibilidad de Arrow.

La función de bienestar social de Harsanyi es del mismo tipo que la de Bergson: consiste en una suma de utilidades o, mejor, una media aritmética -pues, sólo es suma si el número de individuos es constante- con la que se puede determinar el orden de preferencias entre estados sociales alternativos. Pero, a diferencia de la función de Bergson, en la medida en que constituye un juicio de valor imparcial, tiene que ser la misma para todo individuo que sea imparcial. De este modo son posibles los juicios de bienestar social según el procedimiento de Harsanyi.

Gracias a la imparcialidad obtenemos la unanimidad que negaba el resultado de imposibilidad de Arrow; pero, gracias a la imparcialidad deja de existir sobre todo el problema mismo que hacía necesaria la definición de una constitución. Arrow no veía como sería posible que los

juicios individuales pudieran ser juicios sociales; por ello su preocupación se centraba en el procedimiento por el que se podrían determinar -bajo qué regla- un ordenamiento social sobre la multiplicidad de los ordenamientos individuales. Harsanyi no tiene este problema porque él ha diseñado un modelo según el cual un cierto tipo de ordenamientos individuales de preferencias -los éticos- son siempre desde el primer momento ordenamientos sociales. Sin embargo, hemos de preguntarnos: ¿cumple realmente este propósito el modelo de Harsanyi? ¿Son realmente las escalas de bienestar cardinal intersubjetivas?

Pattanaik no lo ve tan fácil como nos lo hace suponer los artículos de Harsanyi. Su argumentación es muy sencilla: los ordenamientos éticos no pueden ser iguales para todos "por la simple razón de que las actitudes ante el riesgo de los diferentes individuos serán, en general, diferentes y que los índices de utilidad de von Neumann-Morgenstern son afectados por las actitudes ante el riesgo" (45). En la medida en que los índices de utilidad son variados, para cada individuo, según su actitud ante el riesgo -la posibilidad que suceda una cosa u otra-, ya no cabría pensar que los ordenamientos éticos, por más que cumplan todas las condiciones de racionalidad, de información factual, el presupuesto individualista, etc., podrían ser idénticos para todos; mucho menos que "todos los individuos acepten un juego imparcial como resultado del cual se puedan quedar en la posición de cualquier otro individuo" (46). La unanimidad creada por la imparciali-

dad se desvanece y con ella queda sin sentido la estrategia del observador imparcial. De nuevo aparece la escisión y el problema de Arrow cobra vigencia. Si las preferencias éticas ya no son eo ipso preferencias sociales, entonces habrá que ponerse a averiguar -la pregunta por el procedimiento- cómo se salva la separación entre las preferencias éticas y las sociales.

Para comprender exactamente cómo inciden los índices de utilidad en la función de bienestar social cardinal de Harsanyi veamos el siguiente caso (47). Sea un universo de dos individuos R, rico, y P, pobre, que tienen que elegir entre dos situaciones sociales: una en la que existe un impuesto progresivo, I, y otra en la que no existe, S. Supongamos que esto da la siguiente tabla, la cual da la renta en dólares para los dos miembros en las dos situaciones:

	Situación social	
	I	S
R	60	100
P	40	10

Los índices de utilidad correspondientes a R y P para cada nivel de renta, calculados de tal forma que resulten comparables interpersonalmente, y supuesto que para R la utilidad marginal de la renta es constante y para P decre-

ciente, son los siguientes:

	Situación social	
	I	S
R	0,6	1
P	0,4	0,2

La función de bienestar social de Harsanyi con estos datos sería para cada situación:

$$W_I = 0,5(0,6) + 0,5(0,4) = 0,5$$

$$W_S = 0,5(1) + 0,5(0,2) = 0,6$$

Los individuos seleccionarían la situación en que no hay impuesto como la situación que imparcialmente maximiza la utilidad esperada de la sociedad. Pero, claramente, P está peor en S que en I; en I su utilidad se dobla mientras que en esta situación R sólo pierde menos de 1/2. Por tanto, P tiene derecho a no querer aceptar el resultado. Pero, ni siquiera a aceptar el juego de imparcialidad de Harsanyi, si, como revelan los índices de utilidad sienta aversión hacia el riesgo. La función de bienestar social tiene en cuenta esta aversión tal y como la reflejan las funciones de utilidad individuales; pero, y esto es lo importante, no refleja la diferencia de actitudes entre los participantes: no las pondera. Sin embargo, como vemos por el ejemplo, son estas diferencias las que determinan las diferencias de las funciones de bienestar social

y las que rompen la unanimidad sobre las mismas (48).

Mueller ha mostrado que hay un procedimiento asequible al modelo de Harsanyi por el que se podría reducir el elemento de arbitrariedad que introducen las actitudes ante el riesgo. El observador imparcial que examina las alternativas puede operar bajo el supuesto de que "tiene una probabilidad igual de tener las preferencias sobre el riesgo de cualquier otro individuo" (49). En la siguiente tabla se muestra el valor esperado de los estados I y S, suponiendo que un individuo tiene la misma probabilidad de ser R o P y que es neutral con respecto al riesgo, N, o que siente aversión al riesgo, A:

	Situación social	
	I	S
N	0,5	0,6
A	0,44	0,42

y las funciones de bienestar social son:

$$W_I = 0,5(0,5) + 0,5(0,44) = 0,47$$

$$W_S = 0,5(0,6) + 0,5(0,42) = 0,51$$

De nuevo se preferirá el estado social en el que no existe el impuesto. No obstante, esta decisión puede ser objetada por P de la misma forma que la anterior, puesto que el estado S tenía desde el principio mayores probabilidades de salir elegido con preferencias neutrales ante

el riesgo que con preferencias de aversión al riesgo. Su objeción sería la misma de Arrow: un procedimiento que dependa de las actitudes individuales hacia el riesgo no tiene significado ético alguno en el ámbito del bienestar social. Mientras no se reduzca esa diferencia de actitudes, la función de bienestar social no podrá ser única para la sociedad. Mueller propone una técnica de sucesivas ponderaciones, bajo el supuesto de equiprobabilidad, hasta obtener un conjunto único de valores para cada situación social (50). Ahora bien, esto convierte al modelo -que, en principio, es bastante sencillo- en un artificio bastante inmanejable y sobre el que penden, a pesar de todos los refinamientos, dos preguntas cruciales:

"(1) ¿Podemos concebir que los individuos obtengan suficiente información sobre las posiciones y la psicología de otros individuos como para permitirles emprender las comparaciones interpersonales inherentes a la teoría?, (2) ¿pueden adoptar los individuos una actitud imparcial con relación a todos los individuos de la comunidad y, a partir de tal plataforma de imparcialidad, acordar un conjunto de ponderaciones (una actitud común respecto al riesgo) aplicables a las posiciones de cada individuo cuando efectúa la elección social?" (51).

La primera cuestión es descartable en la medida -pero, sólo en esa medida- en que los modelos normativos de la elección social no tratan de ser modelos empíricos y utilizables, sino modelos constructivos para la conceptuali-

zación de los problemas de las decisiones sociales en las sociedades democráticas y pluralistas occidentales. En esa misma medida, asimismo, podemos decir que el modelo de Harsanyi supera el resultado de imposibilidad de Arrow y es capaz de producir juicios sociales sobre el bienestar, puesto que fundamenta una norma de elección social: la media aritmética de las utilidades individuales. Sin embargo, del segundo problema no podemos decir lo mismo. En II.7 veremos la crítica que Rawls hace al procedimiento de Harsanyi en base precisamente a la desconfianza que provoca la posibilidad de una actitud imparcial en materia de elección de principios de justicia. En IV.2, a su vez, veremos que la crítica definitiva a la concepción de la fundamentación utilitarista de la justicia económica consiste en mostrar la imposibilidad de preferencias puras con relación a los otros miembros de la sociedad.

II.6 El utilitarismo de preferencia de Harsanyi.

Hemos visto que los autores hasta ahora tratados se declaran más o menos explícitamente en favor de la tradición ética del utilitarismo. Por otra parte, no es difícil ver que los modelos de la teoría de la elección social (como los de la teoría de juegos y la economía del bienestar) se adecúan sin demasiados problemas a un esquema general consecuencialista, y pueden considerarse herederos de las formulaciones clásicas del siglo XIX. Sin

embargo, hasta ahora no habíamos encontrado la forma específica en que afectan a estas formulaciones clásicas la utilización de métodos matemático de análisis y discusión. Harsanyi en la mayoría de sus artículos hace referencia a este tema y nos ofrece una alternativa viable de ética utilitarista basada en los conceptos nucleares de las teorías formales de la conducta humana.

Desde esos modelos, la ética aparece en la obra de Harsanyi como una de las tres teorías especiales de una teoría general de la conducta racional. Se incorpora con ello la ética a una matriz de racionalidad que tiene su modelo en el concepto de racionalidad que ha elaborado la economía, cuyos puntos clave son consistencia y elección entre fines. Frente a la concepción tradicional de la racionalidad, la economía trata de extender su campo de aplicación desde las estrictas situaciones de elección de medios para fines dados a las más generales de elección entre fines alternativos. Al mismo tiempo, se libera de la estructura consecuencialista causal -los medios que producen un fin- y concibe la conducta racional como aquella que es autoconsistente, i. e., que se conforma a unos criterios (52). Sobre esta base se define la conducta racional como aquella que se rige consistentemente por un código de reglas en la elección de fines; y, por tanto, una conducta ética racional es aquella que se rige por unas "reglas morales básicas de un código moral racionalmente elegido" (53).

Por la asimilación de todo tipo de conductas racio-

nales a un modelo formal, nos encontramos con que la ética está basada en un conjunto de axiomas y las reglas morales tienen la misma caracterización que los teoremas de la teoría de la utilidad. Las reglas morales son lógicamente necesarias, son imperativos hipotéticos y han de poder ser siempre revisables cuando su seguimiento produzca consecuencias inaceptables. La ética aparece, pues, en la clasificación de la teoría de la conducta racional como "la teoría de los juicios de valor morales racionales, i. e., de los juicios de preferencia racionales basados en criterios imparciales e impersonales" (54). Ya sabemos qué significa imparcialidad en el modelo de Harsanyi y, por tanto, podemos comprender que la idea de ética que tiene nuestro autor es básicamente la de una teoría que se ocupa de los intereses comunes a todos los miembros de la sociedad. De aquí que no sea extraño que la definición de juicios morales racionales se haga en términos de teoría de la utilidad; son racionales los juicios morales que maximizan el nivel de utilidad media de todos los individuos de la sociedad.

Otro efecto de esta concepción de la ética es que las reglas morales aparecen como imperativos hipotéticos. Consecuente con la idea de racionalidad como consistencia de acuerdo a criterios, los imperativos hipotéticos no son ya causales -como se concebía tradicionalmente-, sino formales: la prótasis especifica ciertos criterios deseados, mientras que la apódosis sugiere un modelo de conducta que satisface ese criterio. Combinada con la teoría

del observador imparcial, obtenemos imperativos morales con la siguiente forma: "Si tu quieres hacer lo que un observador imparcialmente "simpatético" recomendaría para su observancia general, haz X (o abstente de hacer Y)" (55). Como vimos, la teoría de la elección social cuenta con los procedimientos adecuados para formalizar los requisitos de imparcialidad y simpatía en axiomas, lo cual nos llevaría a reglas morales con una forma semejante a los teoremas de la teoría de la utilidad: "Si quieres que tu conducta satisfaga los axiomas A_1, A_2, \dots, A_n , entonces haz X (o abstente de hacer Y)" (56). Esta formulación supone que el campo de los códigos morales concretos queda librado a la elección personal de cada cual; aunque, naturalmente, sería tarea de la filosofía moral el tratar de averiguar qué código moral sería el adecuado para todo tipo de conducta moral racional.

Donde quizás sea más espectacular el impacto de las teorías formales en la ética es en el cambio que se opera en el concepto de "utilidad social". Las formulaciones tradicionales (aquellas que seguían más o menos fielmente a Bentham y hablaban de placeres, satisfacción o felicidad) se apoyaban en una tradición psicológica hedonista hoy totalmente obsoleta. Por otra parte, existían problemas muy serios para entender qué es lo que se pretendía maximizar al hablar del "placer de la mayoría" o "la mayor felicidad del mayor número". Correcciones como la del utilitarismo ideal de G. E. Moore que hablaba de "estados mentales de mérito intrínseco" no ayudaban mucho.

La precisión que los autores provenientes del campo de la economía y la estadística exigían para poder operar con sus modelos de decisión no se encontraba en aquellas formulaciones.

Para Harsanyi, la manera más fácil de hacer operativo, -y a la vez más realista- el concepto de "utilidad social" es definirlo en términos de las preferencias (por tanto, de las funciones de utilidad) de los miembros de la sociedad. Nos encontramos, así, ante lo que él denomina "utilitarismo de preferencia", el cual define como correcto aquel curso de acción que maximiza la media aritmética de las utilidades individuales, i. e., de las preferencias manifestadas por los individuos en sus funciones de utilidad. Este utilitarismo, dado los conceptos ya mencionados de racionalidad y de imperativo moral, es un tipo de utilitarismo de la regla. Las reglas morales son imperativos hipotéticos formulados por un observador imparcial y, por tanto, no prejuzgan qué son buenas consecuencias con independencia de lo que prefieren los sujetos interesados(57). Juzgar "simpatéticamente" qué son buenas consecuencias de un curso de acción sólo puede hacerse en términos de las preferencias de los implicados; se trata siempre de "ponerse en el lugar de los otros". De esta forma, la teoría de Harsanyi trata de conformarse y dar sentido a la regla de oro: tratar a los otros como queremos ser tratados es sólo "tratar a los demás de acuerdo con sus propios deseos (wants) y preferencias" (58). El concepto de preferencia que aquí se presenta es

el único adecuado para operar bajo este principio, puesto que permite respetar todos y cuales sean los deseos y objetivos de los demás. En este punto, no obstante, la teoría tiene que precisar que ese preferir no está sin cualificar. No es cualquier preferencia la que el observador imparcial habría de tener en cuenta, sino tan sólo las preferencias "verdaderas" de los sujetos, aquellas que un sujeto manifestaría "bajo la debida reflexión y en posesión de toda la información relevante" (59). Este concepto se contrapone al de preferencias "reales" que serían no sólo las irracionales, sino todas aquellas que se producen en condiciones de baja reflexividad e información.

Con este entramado teórico -imperativos hipotéticos, observador imparcial, preferencias "verdaderas"- se argumenta en favor de un utilitarismo que haría valer el criterio de las buenas consecuencias para las prácticas generales y no para los puros actos (60). A Harsanyi le interesa defender la no identificación entre utilitarismo del acto y utilitarismo de la regla -tesis de Brandt y Lyons (61)-, puesto que su modelo pretende cubrir las expectativas morales -de justicia y respeto a los derechos básicos- que el utilitarismo del acto no puede tener en consideración. Para ello el utilitarismo de preferencia ofrece un refinado modelo formal para la toma de decisiones morales. El modelo contiene los siguientes elementos:

- (1) Un número finito de agentes morales cada uno con su función de utilidad específica.

- (2) Un número finito de acciones posibles en un momento dado entre las cuales el agente tiene capacidad de decidir.
- (3) El agente puede encontrarse en cualquiera de un número finito de diferentes situaciones.
- (4) Las acciones de cada agente son regidas por una estrategia que es una función matemática que asigna una acción específica a cada situación posible -bajo el principio de los indiscernibles de Leibniz (62).

Hay que observar que el concepto de estrategia reemplaza al concepto de regla moral, lo cual no es meramente un cambio terminológico; contrariamente, el nuevo concepto tiene un plus de significado sobre el sustituido: indica no sólo que hay una regla para decidir, sino también que hay una alternativa viable entre todas.

En orden a especificar cuántos y quienes son los agentes envueltos en las situaciones de decisión y a los que afectan sus consecuencias -una objeción tradicional a las vagas formulaciones utilitaristas del pasado-, el modelo discrimina tres tipos de agentes:

- (1) Agentes primarios: los encargados de tomar las decisiones.
- (2) Agentes secundarios: los afectados directa o indirectamente por lo que se haga en las distintas situaciones.
- (3) Agentes terciarios: todo el resto de los agentes (62).

Asimismo para distinguir el tipo de conducta de los agentes, estos son considerados de dos tipos:

- (1) Agentes flexibles: agentes cuyas estrategias se determinan según el utilitarismo de preferencia.
- (2) Agentes rígidos: agentes cuyas estrategias están dadas (63).

Esta distinción es importante porque permite diferenciar no sólo al utilitarismo de otras teorías éticas, sino además al utilitarismo del acto y al utilitarismo de la regla. Para el primero, tan sólo el decisor utilitarista es un agente flexible, al cual se le pide que, dadas las estrategias de los demás individuos, elija la acción de mejores consecuencias. Por el contrario, para el utilitarismo de la regla todos los agentes son flexibles en la medida en que todos maximizan la utilidad social, si todos usan la misma regla (64).

Ahora bien, la diferencia crucial entre ambas formas de utilitarismo estriba en que el decisor que aplique el utilitarismo del acto tiene que centrarse en las consecuencias causales directas que resultarán de la adopción personal de una estrategia. En cambio, el utilitarista de la regla tiene que decidir bajo dos supuestos la bondad de las consecuencias de una estrategia:

- (a) Todos los agentes primarios usarán la estrategia S.
- (b) Todos los agentes secundarios conocerán, y actuarán sobre este conocimiento, que los agentes

primarios usarán la misma estrategia S.

De estos supuestos se sigue que un decisor que siga el utilitarismo de la regla no puede considerar sólo las consecuencias causales, sino también las "implicaciones lógicas no causales" de la adopción de una estrategia por todos los agentes primarios (65).

La adopción de una estructura consecuencialista de decisión moral provoca tres tipos de efectos que en su forma de darse distinguen netamente a las dos formas de utilitarismo consideradas:

- (1) Efectos de coordinación: Las teorías éticas presentan unas condiciones favorables o no para coordinar las estrategias de los agentes primarios de una manera socialmente deseable.
- (2) Efectos de expectación: Efectos de la adopción de cualquier estrategia sobre la capacidad de los agentes secundarios de formar expectativas y sobre sus sentimientos de confianza y seguridad.
- (3) Efectos de incentivo: Efectos sobre los incentivos de los agentes secundarios para comprometerse con los varios tipos de conducta socialmente beneficiosa (65).

El utilitarismo del acto se distingue claramente del utilitarismo de la regla en cómo produce los distintos efectos. Si atendemos a los efectos de coordinación, ambas teorías hacen recomendaciones diferentes y tienen diferentes implicaciones morales. Puesto que uno recomienda to-

mar como dadas las estrategias de los demás agentes y el otro no, el utilitarismo de la regla -p. ej., en una situación de voto- puede alcanzar el máximo de utilidad social -donde todos votan-, mientras que el utilitarismo del acto puede obtener el mínimo -donde nadie vota (66). Por otro lado, puesto que el utilitarismo de la regla impone la restricción de que todos han de seguir idénticas estrategias -no así el utilitarismo del acto-, donde sea imprescindible una coordinación de esfuerzos -p. ej., en situaciones de voto donde sea necesaria una mayoría de $2/3$, se puede conseguir fácilmente tal requisito; mientras que tal coordinación es dejada al azar por el utilitarismo del acto (67).

En cuanto a los efectos de expectación e incentivo es evidente que al utilitarismo del acto le está impedido el hacerse cargo de ellos; un decisor que lo siga sólo puede atenerse a consecuencias directas de su estrategia, no a posibles efectos indirectos en agentes secundarios. Por el contrario, se sigue de la aceptación de supuesto (b) que el utilitarismo de la regla tiene necesariamente que valorar las consecuencias de adoptar una estrategia sobre los sentimientos de seguridad e incentivo de los sujetos secundarios (68).

Por último, queremos señalar que la solidez y refinamiento con que Harsanyi dota al utilitarismo lo convierte en una ética social de una formidable capacidad operativa en las más variadas y complejas situaciones que se presentan en las sociedades democráticas actuales.

Por ello, es un enemigo del que no es fácil deshacerse

o tan siquiera desacreditarlo. Quizás el único flanco que presente más vulnerable sea precisamente el de sus supuestos últimos sobre el hombre y la sociedad; sobre todo, su concepción de lo que puede ser y es fundamento de juicios de valor social. Los fundamentos últimos -el individualismo- y los procedimientos de elección -las funciones de utilidad y las reglas utilitaristas- son los objetivos de nuestra investigación, porque han de ser las fuentes de plausibilidad de la teoría que sustentan. Si ellos pueden ser cuestionados, por muy sólida, viable, operativa que sea la teoría nunca podrá solicitar de nosotros la aquiescencia.

II.7 Elección de principios: Harsanyi versus Rawls.

Aunque la teoría de la elección social en general puede considerarse dependiente de una ética utilitarista -en el sentido, asimismo genérico de que identifica los hechos morales básicos con el bienestar de los individuos-, los diferentes modelos que en ella se han ido desarrollando muestran relaciones variables y matizadas con dicha ética. De entre ellos, como hemos visto, es el modelo de Harsanyi el que presenta una relación más fuerte y quizás sea también el único que se ha preocupado de fortalecerla reflexionando explícitamente sobre una nueva forma de utilitarismo más adecuada a los problemas actuales de la elección social.

Nos interesa destacar ésto porque creemos que sería útil para la comprensión de las cuestiones de fundamentación presentar una confrontación entre la ética utilitarista y su más directo oponente la ética deontológica. Rawls ha presentado no sólo la alternativa mejor argumentada al utilitarismo en los temas de justicia, sino que además la ha acompañado de una crítica general que el utilitarismo difícilmente puede obviar. No obstante, como fácilmente se puede comprobar, el utilitarismo es una tendencia que con el tiempo ha presentado muchas caras, perfeccionando y fortaleciendo su propia posición. En este sentido, nos parece interesante en orden a enunciar una reflexión distanciada sobre esta teoría comprobar qué valor tiene la crítica rawlsiana frente al utilitarismo de preferencia de Harsanyi (69).

Rawls se ocupa centralmente del tipo de utilitarismo que defiende Harsanyi en las secciones 27 y 28 de su libro, especialmente en ésta última -"Difficulties with Average Utility"- presenta sus críticas. Sin embargo, antes de pasar a su examen, debemos señalar dos cosas que facilitarán la comprensión de esta polémica. En primer lugar, Rawls presenta sus objeciones al principio de utilidad media sirviéndose de su peculiar método de considerar los principios de la justicia; es decir, examinando la validez de un principio por referencia a si sería elegido por los participantes en la posición original. En segundo lugar, ha de tenerse en cuenta que Harsanyi y otros autores (70) piensan que la posición original es un pro-

cedimiento del mismo tipo que el utilizado por él para alcanzar una función de bienestar social en base a los juicios éticos de un observador imparcial en una situación de equiprobabilidad. Esta similitud ha sido reconocida por Rawls (71) y, ciertamente, ambos procedimientos pueden considerarse formas de una teoría contractual que trata de alcanzar la constitución de una voluntad general en torno a unos principios de justicia (72).

Veamos, entonces, la argumentación de Rawls. Objeta este autor que el principio de utilidad media no sería elegido por los participantes porque en la posición original no existen bases para calcular las probabilidades de los resultados (indirectamente esto significa que no es utilizable el supuesto de equiprobabilidad), sino meras posibilidades de que un principio lleve a un tipo de consecuencias o no (73). Sobre esta base, si se quieren seguir adjudicando probabilidades, los participantes sólo podrán fijarlas operando con el principio de indiferencia de Laplace (principio de razón insuficiente): "Cuando no se dispone de ningún tipo de datos se estipula que todos los casos posibles han de ser igualmente probables" (74). No obstante, Rawls supone -según él, "plausiblemente"- que los participantes en la situación de elección inicial evitarían tener en cuenta todas las consecuencias que se obtuvieran con la utilización de ese principio, en virtud de la gran incertidumbre que lanzaría sobre los posibles resultados de la elección del principio de utilidad que apoya. Los participantes, sobre todo, evi-

tarían actuar bajo ese principio, al tener a la vista los resultados satisfactorios y seguros que ofrecen la elección de los dos principios de justicia rawlsianos (75).

Esta argumentación se basa en tres supuestos que el propio autor expone (76). En primer lugar, "lo esencial es no permitir que los principios escogidos dependan de actitudes especiales frente al riesgo" (77). Rawls considera -al igual que Arrow- que una elección influida por las actitudes ante el riesgo de los participantes carecería de cualquier significado moral. Pero, la situación de elección inicial se caracteriza por ser una elección incierta y arriesgada. ¿Cómo se resuelve, entonces, el problema de la ponderación del riesgo que aceptan los participantes?

"En la medida de lo posible, la elección de una concepción de la justicia habría de depender de una valoración racional de los riesgos que no se viera afectada por preferencias peculiares de los individuos respecto a la adopción de riesgos de un modo u otro" (77).

Finalmente, pues, la cuestión de la valoración individual de los riesgos se resuelve en contra de adoptar un modelo en el que las preferencias individuales se expresen bajo la forma de una función de utilidad del tipo de la de von Neumann-Morgenstern.

El segundo supuesto que señala Rawls consiste en la no admisión en su modelo de otros juicios de probabilidad que aquellos que tengan una base objetiva, i. e.,

probabilidades empíricas 78). Dado que el principio de Laplace sólo puede operar con probabilidades subjetivas, este principio es descartado en favor del principio maximin que opera con probabilidades objetivas. Rawls justifica esta elección del principio maximin en base al tipo especial de decisión que tiene que llevarse a cabo en la posición original: una decisión que, por su trascendencia, obliga a ser conservador (79).

El último supuesto se dirige en contra del núcleo mismo del modelo de Harsanyi: su función de decisión social la realiza un solo individuo qua observador ideal; de donde se sigue que "las utilidades representan la validez de las alternativas para este individuo, estimadas en función de su tabla de valores" (79). Rawls cree que esperar de ello un juicio intersubjetivo tiene tanto sentido como esperar que alguien tenga valores, fines, expectativas que no le sean propias. Se requeriría un esquema de preferencias previo para dar sentido social a una elección individual así concebida. ¿Para qué? Para que el observador imparcial -en base a ese esquema- pueda realizar comparaciones interpersonales.

Tenemos aquí de nuevo la problemática de la interpersonalidad. 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 sinsentido. Este procedimiento es adecuado cuando lo que se trata de comparar y evaluar son cosas como la posición social, la riqueza

o el bienestar; pero, ¿es igual de adecuado cuando lo que hay que comparar y evaluar son modos de vida y sistemas de vida ajenos? (80). Lo que el modelo de Harsanyi exige es que valoremos como observadores imparciales; pero, ello no es tarea fácil para quien le es inevitable tener su propio plan de vida y su sistema de fines. ¿A cuál le daremos prioridad en la comparación, si son irreconciliables, al suyo o al nuestro? La posición original es precisamente una situación diseñada pensando en que no hay una sola 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.

En resumen, si en la posición original no se admite que las actitudes ante el riesgo determinen la elección, ni que las probabilidades subjetivas sirvan de base a la misma, ni consiente comparaciones interpersonales, es imposible que los participantes tengan posibilidad de razonar de forma que prefieran el principio de utilidad media a los principios de justicia de Rawls. Porque, como Harsanyi apuntará, si se niega uno a admitir la utilización de probabilidades subjetivas, está obligado a utilizar algún tipo de regla maximin (81). La regla maximin, por lo demás, sólo puede favorecer la elección de principios tan conservadores como los de Rawls. En definitiva, la argumentación en contra del principio de utilidad es,

en realidad, una argumentación en favor de la utilización en la situación inicial de la regla maximin y de ninguna otra.

Harsanyi responde una por una las objeciones que Rawls presenta a su modelo en cada uno de esos tres supuestos. En primer lugar, explica que la teoría bayesiana de la elección bajo incertidumbre ha demostrado que cualquier decisor en una situación de ese tipo -y, ciertamente, de ese tipo es la posición original- "no puede evitar actuar como si intentara maximizar su utilidad esperada, calculada sobre la base de un determinado conjunto de probabilidades subjetivas" (82), a menos que se rechazasen los postulados de racionalidad de la teoría bayesiana; lo cual -opina Harsanyi- sería muy extraño dado el consenso actual sobre su atractivo.

Algunos ejemplos utilizados por este autor muestran intuitivamente la imposibilidad de evitar las probabilidades subjetivas. Supóngase la elección entre dos apuestas: "O le pago 100 \$ si el candidato X gana las próximas elecciones y si pierde no le pagaré nada. O le pagaré 100 \$ si no gana y si gana no le pagaré nada. ¿Cuál de las dos apuestas escogería?" (82). Si se acepta la primera apuesta -y sería irracional no escoger una de las dos-, se puede legítimamente suponer que se ha asignado al menos $1/2$ a la posibilidad de que el candidato gane; si se acepta la segunda apuesta se puede suponer que ha dado a lo más $1/2$ de probabilidad a que el candidato pier-

da. En cualquier caso, se puede suponer que ha dado una probabilidad subjetiva al resultado de las elecciones.

Harsanyi va más allá y señala que, incluso si el decisor se rige -en lugar de por la regla de maximización de la utilidad esperada- por la regla maximin, ello no le evita hacer una elección entre probabilidades subjetivas, "al menos implícitamente": "(...) el empleo del principio maximin en la posición original equivale a asignar una probabilidad de uno, o casi uno, a la posibilidad de que un individuo pueda terminar como el individuo menos favorecido de la sociedad (...)" (83). 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 conlleva los riesgos que el principio utilitarista inevitablemente supone (84). Harsanyi sostiene en contra que si bien la equiprobabilidad (bajo la cual se calcula la utilidad esperada) puede interpretarse -como hace Rawls- como una variante del principio de razón insuficiente, otra interpretación es posible; puede ser la expresión del principio moral de dar a los intereses de cada individuo el mismo peso. Si se acepta tal principio -y no aceptarlo es moralmente muy problemático (85)-, 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 -y contra la primera objeción de Rawls-, Harsanyi explica lo que ya sabemos sobre la intro-

ducción de las actitudes ante el riesgo en las elecciones sociales (86). Si las funciones individuales de utilidad de von Neumann-Morgenstern se construyen 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. Pero, construidas bajo el supuesto de equiprobabilidad, el factor de actitud ante el riesgo deja de ser arbitrario, quedando incorporado en las funciones de utilidad que expresan el concepto global de la "importancia subjetiva" que el individuo asigna a los diferentes resultados. Siendo así, la actitud ante el riesgo no tiene ningún efecto amoral y "las funciones de von Neumann-Morgenstern tienen un lugar totalmente legítimo en ética" (87).

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" (88). Pero, esta dificultad es salvable. Harsanyi y Rawls tienen dos conceptos diferentes del sujeto de los procesos sociales y aquí chocan frontalmente. Mientras que Rawls subraya 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 las funciones de utilidad de todos los seres humanos están gobernadas por las mismas leyes psicológicas básicas" (88).

Esto es lo que posibilita el funcionamiento del mecanismo del "ponerse-en-el-lugar-del-otro" cuyo resultado es que cualquier individuo es capaz de reconstruir la función de utilidad de cualquier otro individuo. Ello reduciría el problema de la interpersonalidad a un problema de comparación entre las funciones de una misma persona (intrapersonalidad): la propia y la que se tiene cuando uno se considera en el lugar del otro.

Esta argumentación en contra de los supuestos de Rawls tiene como objetivo -evidente y último- la defensa del principio de la maximización de la utilidad esperada en contra del principio maximin de Rawls, ofreciéndose así el modo por el que el principio de justicia elegido en la posición original fuera el principio de utilidad media. A este efecto sostiene que el diseño de la posición original es similar a su diseño de la situación de equiprobabilidad en la que el observador imparcial construye la función de bienestar social. Pero, el papel protagonista en esa construcción analítica lo tiene el principio de decisión que en ella se adopte y, en consecuencia, toda la plausibilidad de dicha construcción depende de la plausibilidad del principio que favorezca. En opinión de Harsanyi, Rawls ha arruinado su construcción, "potencialmente muy poderosa para clarificar el concepto de justicia y otros aspectos éticos", al darle ese papel a la regla maximin (89).

Entre los teóricos de la decisión el principio ma-

ximin ya tiene una historia contrastada. Durante los años cuarenta y la primera mitad de los cincuenta la opinión fue que el principio adecuado para las decisiones bajo incertidumbre tenía que pertenecer a la familia maximin. A partir de esa fecha empezaron a verse las consecuencias irracionales a las que llevaba su aplicación práctica. Ello trajo consigo el predominio actual de la tradición bayesiana que defiende el principio de maximización de la utilidad esperada como la regla canónica para ese tipo de situaciones (90).

¿Qué tipo de consecuencias origina la aplicación del principio maximin que han hecho que caiga en desgracia entre los matemáticos? Harsanyi lo explica con el siguiente caso (91). Suponga que le ofrecen un trabajo en Chicago muy satisfactorio para vd.; para ello tiene que abandonar un trabajo que le satisface poco en New York, y tomar inmediatamente un avión hacia Chicago. Vd. tiene terror a volar y teme que le suceda un accidente mortal. La situación de elección es la siguiente:

	Si el avión a Chicago	
	sufre un accidente	no sufre un accidente
Si escoge el empleo de N. Y.	Tendrá un empleo insatisfactorio y estará vivo.	Tendrá un empleo insatisfactorio y estará vivo.
Si escoge el empleo de Chicago.	Vd. morirá.	Tendrá un empleo satisfactorio y estará vivo.

La lógica del criterio maximin es evitar lo peor. Por ello, el razonamiento de la elección consistiría en lo siguiente:

- I. Si me quedo en N. Y., el peor resultado será:
estará vivo con un empleo insatisfactorio.
- II. Si parto para Chicago, el peor resultado será:
estará muerto.

La elección que recomienda la regla maximin es I, el menos malo de los dos resultados.

Lo que, por el contrario, recomienda el principio de la maximización de la utilidad es que vd. considere la probabilidad de que suceda un accidente -cosa que ni siquiera puede considerarse bajo el principio maximin- y si esa probabilidad es suficientemente pequeña, y su preferencia por el trabajo de Chicago es fuerte, se arriesgue y vuele a Chicago. !Quién no pensaría que esta es la solución racional! Harsanyi comenta sardónicamente: "Si tomara en serio el principio maximin, vd. ni siquiera podría cruzar la calle (al fin y al cabo un automóvil podría atropellarle). (...). Si alguien actuase realmente de esta manera, pronto terminaría en una institución mental" (92).

Siendo más serio, la razón por la que el principio maximin da lugar a este tipo de decisiones irracionales es precisamente por su negativa a utilizar probabilidades subjetivas. Si a cada situación posible, el sujeto le asigna una probabilidad, la incidencia de sucesos -que

en realidad son improbables- desfavorables queda reducida a un papel adecuado en el razonamiento de decisión.

No obstante, esto no decide todo con respecto al criterio maximin, tal y como es utilizado por Rawls. Harsanyi es sensible a la argumentación rawlsiana en su favor; la posición original, insiste Rawls, es una situación de elección especial en la que está justificado ser conservador. La pregunta que se hace Harsanyi es si puede realmente el principio maximin ser un fundamento ético: "(...) ¿obtendremos un concepto satisfactorio de justicia si nos imaginamos que los criterios de justicia los seleccionan personas en la posición original de acuerdo con el principio maximin?" (93). Estamos en un punto crucial de la confrontación en el que se va a poner de manifiesto con toda crudeza que, más allá de las argumentaciones técnicas, lo que está en ^{conflicto son} -si se nos permite utilizar este término- dos "concepciones del mundo" diferentes e irreductibles; consecuencia de ello será el que, por más razonamientos y refutaciones que se intercambien, ambos diálogos permanecerán impávidos (94).

Harsanyi argumenta en contra de la utilización del principio maximin en la posición original mostrando que el principio de diferencia que él selecciona produce decisiones con la misma característica de irracionalidad que produce la aplicación directa del maximin. De los casos que presenta Harsanyi para ilustrar esto, he elegido el más polémico (95). Considérese una sociedad de dos individuos. El primero, A, es un superdotado para las mate-

máticas y tiene un interés fuerte por esta disciplina; el segundo, B, es un individuo retrasado. La sociedad tiene un nivel suficiente de recursos para asegurar un buen nivel de vida a ambos; pero, además, tiene un excedente de recursos. Se proponen dos políticas alternativas para la inversión de estos recursos excedentes:

I: ofrecer al individuo A una enseñanza superior en matemáticas.

II: ofrecer al individuo B una enseñanza que elimine algunas de sus peores incapacidades (p. ej., podría conseguirse que aprendiera a atarse los zapatos y cosas por el estilo).

Se supone que no podrían repartirse los recursos entre ambas políticas y que, necesariamente, habría que optar por una de las dos. Ciertamente, el principio de diferencia de Rawls seleccionaría la política II, requiriendo que otorguemos "prioridad absoluta a los intereses del individuo peor situado, independientemente de cualquier otro hecho, incluso bajo las circunstancias más extremas" (96). El principio de maximización de la utilidad -y, en opinión de Harsanyi, el sentido común- recomendaría la política I (97).

Rawls había prevista esta objeción. Su contrareplica consiste en señalar la diferencia entre la regla maximin y el principio de diferencia: "(...) el problema de la justicia social no consiste en distribuir ad libitum entre ciertos individuos diversas cantidades, sea dinero, propiedades o cualquier otra cosa" (98). Es decir, el problema de la justicia no es un problema de distribución

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

El principio de diferencia, con sus diferentes especificaciones, es un principio complejo por sí mismo; pero, además, lo es porque está inserto en un sistema lexicográfico de principios, dentro del cual está indisolublemente unido con su prior el principio de libertad (99). Como tal participa de una concepción general de la justicia que ordena la realidad social de forma que, en la realidad, nunca se producirán casos del tipo descrito por Harsanyi. El efecto de su aplicación es siempre una mejora en todos los segmentos de la población, porque la característica de ese ordenamiento de la sociedad es la continuidad entre todos esos segmentos: "Porque al elevar las expectativas de los más aventajados, la situación de los peor situados es "continously" mejorada" (100). La estructura social que crean la aplicación de los principios no permite que se den casos en los que sólo un segmento mejore a costa de otros.

Harsanyi no contesta a esta objeción porque Arrow

ya lo había hecho de una manera a su juicio contundente. Comenta este autor que si la sociedad tuviera -o llegara a tener- esa propiedad de "continuidad" habría que reconocer que el principio de diferencia rawlsiano no tendría consecuencias análogas al maximin. Pero, ni la sociedad tiene actualmente esa propiedad ni parece que pudiera llegar a tenerla si se rigiese por los principios de justicia de Rawls. ¿Cómo podrían eliminar estos, o las instituciones que prescribieran, las desigualdades producidas por las herencias cultural y biológica o, simplemente, por sucesos fortuitos? Una vez que las desigualdades se presentaran, la continuidad se rompería al mismo tiempo -y por la misma causa- que la armonía de intereses. Por otra parte, observa Arrow, "si todas las satisfacciones van juntas, el conflicto entre el individuo y la sociedad desaparece" (101), borrándose la diferencia entre las implicaciones políticas del principio de diferencia y las del principio de la maximización de las utilidades.

Harsanyi, no obstante, es consciente de que lo que aquí está en litigio es algo que no es reductible con argumentos técnicos o contra-ejemplos que saquen a la luz consecuencias intuitivamente inaceptables. Lo dice del siguiente modo:

"Rawls es muy consciente de que el principio de diferencia tiene implicaciones de este tipo. Con todo cree que estas implicaciones son moralmente deseables debido a que en su opinión siguen el prin-

cipio kantiano de que las personas deberían 'trastarse entre sí no como medios, sino como fines en sí mismos' (102).

En el caso anterior ello significaría que asignar los recursos a A sería utilizar como medio a B. Harsanyi objeta que no entiende cómo se puede aplicar el fundamento kantiano a una situación de este tipo. Esto debe ser subrayado. No argumenta que la máxima kantiana sea inválida; dice que hay casos en que no es aplicable y, en definitiva, que es un error aplicar a esos casos el principio de diferencia en lugar del principio utilitarista.

Efectivamente, la situación es ésta: se niega a B unos recursos a los que no tiene derecho de propiedad previo, en favor de un individuo que derivará más utilidad de ellos. Por consiguiente, "está fuera de toda cuestión el utilizar la persona o propiedad de B en beneficio de otros individuos" (103); ¿se puede hablar, entonces, de que B está siendo utilizado como un medio? Como coda Harsanyi indica que si tuviera sentido aplicar la máxima kantiana a una situación de ese tipo, con la misma razón se podría hablar -en caso de elegir la política II recomendada por el principio de diferencia- de que A está siendo utilizado como un medio en beneficio de B.

En la confrontación de Rawls con Harsanyi encontramos sintéticamente las características de dos modos alternativos de considerar los temas de la justicia. El fundamento moral del utilitarismo -que rotulamos con el término "individualista"- se expresa en la máxima de dar igual

peso a los intereses de cada persona. El fundamento kantiano de la teoría de la justicia de Rawls se expresa en la máxima de dar a cada uno un "valor de diferencia" que exige tratar a las personas como fines y no como medios. Estos son los fundamentos que se articulan y a la vez diseñan diferenciadamente el mismo procedimiento de elección de los principios: un procedimiento contractual en su definición más genérica. El contrato utilitarista y el contrato deontológico son la expresión más nitida de dos maneras irreconciliables de concebir el esquema de construcción de lo social; más exactamente, cómo debe ser realizado lo social. Por eso, no es extraño que las normas -los principios de la justicia- que se seleccionan en ambos procedimientos sean asimismo irreconciliables; que cada una tenga una problemática propia y ajena a la de su oponente; que cada una de ellas tenga una plausibilidad de la que carece la otra y, finalmente, que cada una tenga unas impotencias que la otra sencillamente tiene superadas desde el principio.

NOTAS

- (1) J. von Neumann- O. Morgenstern, (1947), Theory of Games and Economic Behaviour.
- (2) El campo de aplicación de la teoría de juegos es muy amplio: desde los conflictos entre los miembros de una familia a conflictos militares, pasando por todo tipo de situaciones como negociaciones entre obreros y patronos, conflictos entre grupos políticos, raciales, etc.; incluso puede ofrecer una "filosofía de la historia" centrada en las estrategias del conflicto. Th Caplow, (1968), Dos contra uno: teoría de coaliciones en las triadas.
- (3) O. Morgenstern, (1968), "la teoría de juegos", en: D. M. Messick (ed.), (1968), p. 190.
- (4) O. Morgenstern, (1968), en: D. M. Messick (ed.), (1968), p. 192.
- (5) A. Rapoport, (1968), "Uso y abuso de la teoría de juegos", en: D. Messick (ed.), (1968), p. 229.
- (6) M. D. Davis, (1969), Teoría de juegos. Una introducción no técnica, p. 70.
- (7) R. D. Luce - H. Raiffa, (1957), Games and Decisions, pp. 23-31.
- (8) K. J. Arrow, (1951-1963²), Elección social y valores individuales, pp. 100-101.
- (9) K. J. Arrow, (1951-1963²), p. 65. Para una refutación de este punto de vista, P. K. Pattanaik, (1968), "Risk, Impersonality and the Social Welfare Function", en: E. S. Phelps (ed.), (1973), pp. 314 y ss.

- (10) K. J. Arrow, (1951-1963²), pp. 102-103.
- (11) K. J. Arrow, (1951-1963²), pp. 66-7.
- (12) R. B. Braithwaite, (1969), Theory of Games as a Tool for the Moral Philosopher.
- (13) R. B. Braithwaite, (1969), p. 4.
- (14) R. B. Braithwaite, (1969), pp. 5-6 (subrayado del autor de la cita).
- (15) R. B. Braithwaite, (1969), p. 9.
- (16) R. B. Braithwaite, (1969), p. 20.
- (17) A. K. Sen, (1970b), Elección colectiva y bienestar social, p. 153; y cc. 8 y 8" donde Sen analiza otros modelos de teoría de juegos a los que es aplicable esta misma conclusión, como el modelo de Nash: J. F. Nash, (1950), y (1953).
- (18) O. Höffe, (1975), Estrategias de lo humano, c. 3.3.
- (19) M. Flemming, (1952), "A Cardinal Concept of Welfare", en: E. S. Phelps (ed.), (1973), p. 245.
- (20) M. Flemming, (1952), en: E. S. Phelps (1973), pp. 246-7.
- (21) M. Flemming, (1952), en: E. S. Phelps (ed.), (1973), p. 248.
- (22) M. Flemming, (1952), en: E. S. Phelps (ed.), (1973), pp. 249-52. Nosostros presentamos su reenumeración más simple de J. C. Harsanyi, (1955), "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility", en: E. S. Phelps (ed.), (1973), pp. 267-71.
- (23) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.),

- (1973), p. 268.
- (24) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 271.
- (25) M. Flemming, (1952), en: E. S. Phelps (ed.), (1973), pp. 255-9.
- (26) M. Flemming, (1952), en: E. S. Phelps (ed.), (1973), p. 261.
- (27) J. C. Harsanyi, (1953), "Cardinal Utility in Welfare Economics and in the Theory of Risk Taking", en: J. C. Harsanyi, (1976).
- (28) M. Friedman - L. J. Savage, (1948), "The Utility Analysis of Choices Involving Risk".
- (29) M. Friedman - L. J. Savage, (1948), p. 283n.
- (30) J. C. Harsanyi, (1953), en: J. C. Harsanyi, (1976), p. 4.
- (31) J. Marschack, (1950), "Rational Behaviour, Uncertain Prospects and Measurable Utility".
- (32) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 272.
- (33) Los postulados de Marschak son:
1. Ordenamiento completo: La relación de preferencia establece un ordenamiento completo entre todas las perspectivas.
 2. Continuidad: Si la perspectiva P es preferida a la perspectiva R, mientras la perspectiva Q tiene una posición intermedia entre ellas (siendo preferida a R, pero no a P), entonces existe una mezcla de P y R, con las consecuentes proba-

bilidades, de forma que son indiferentes a Q.

3. Suficiente número de perspectivas no indiferentes: Hay al menos cuatro perspectivas mutuamente no indiferentes.
4. Equivalencia de mezcla de perspectivas no equivalentes: Si las perspectivas Q y Q' son indiferentes, entonces, para cualquier perspectiva P, una mezcla dada de P y Q es indiferente a una mezcla similar de P y Q'. Esta propiedad es la fundamental para la cardinalidad de la utilidad.

(34) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 275.

(35) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), p. 276.

(36) Es importante comparar y anotar la diferencia entre la distinción de Arrow de preferencias según "gustos" y según "valores" y la distinción de Harsanyi entre preferencias "subjetivas" y "éticas". Para Arrow, el individuo puede tomar en cuenta sólo los factores de su propio consumo, entonces hará un ordenamiento de preferencias según sus gustos; o puede tomar en cuenta relaciones de interdependencia -economías y deseconomías externas- y considerar también valores, entonces hará un ordenamiento de preferencias según valores. Son éstas últimas las que hay que tener en cuenta para formar la función de bienestar social.

La distinción de Harsanyi es más fina: lo que Arrow separa él lo clasifica como preferencias subjetivas;

sólo las preferencias hechas desde un punto de vista imparcial, simpatético y equiprobable son éticas, y sólo ellas constituyen el ordenamiento de la sociedad.

J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 276n.

- (37) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), pp. 276-7; y 277n: "O mejor, si él tuviera una posibilidad igual de ser "puesto en el lugar de" cualquier miembro de la sociedad, con relación no sólo a sus condiciones objetivas sociales (y económicas), sino también a sus actitudes y gustos subjetivos. En otras palabras, él debería juzgar la utilidad de la posición de otros individuos en términos no de sus propias actitudes y gustos, sino más bien en términos de las actitudes y gustos propios de los individuos que realmente están en esa posición".

(38) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 277.

(39) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 278.

(40) Sobre la teoría de la preferencia revelada, A. K. Sen, (1973b).

(41) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 279.

(42) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), pp. 279-80.

(43) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 282.

- (44) J. C. Harsanyi, (1955), en: E. S. Phelps (ed.), (1973), p. 283.
- (45) P. K. Pattanaik, (1968), en: E. S. Phelps (ed.), (1973), p. 306.
- (46) D. C. Mueller, (1979), Elección pública, p. 257.
- (47) El ejemplo es de D. C. Mueller, (1979), pp. 257 y ss.
- (48) P. K. Pattanaik, (1968), en: E. S. Phelps (ed.), (1973), p. 309.
- (49) D. C. Mueller, (1979), p. 258.
- (50) D. C. Mueller, (1979), p. 259.
- (51) D. C. Mueller, (1979), p. 261.
- (52) J. C. Harsanyi, (1975-6), "Advances in Understanding Rational Behaviour", en: J. C. Harsanyi, (1976), s. 3.
- (53) J. C. Harsanyi, (1978), "Rule Utilitarianism an Decision Theory", en: H. W. Gottinger - W. Leinfellner (ed.), (1978), p. 4. Una opinión similar es expresada por R. B. Brandt cuando afirma que la cuestión de la justificación de los códigos morales es correctamente enunciada cuando se pregunta: "¿Qué tipo de código moral social, si lo hay, tendería a apoyar en una sociedad en la que esperase vivir, si vd. fuera plenamente racional?", R. B. Brandt, (1979), A Theory of the Good and the Right, p. 185 (los subrayados son del autor de la cita).
- (54) J. C. Harsanyi, (1975-6), en J. C. Harsanyi, (1976), p. 97.

- (55) J. C. Harsanyi, (1958), "Ethics en Terms of Hypothetical Imperatives", en: J. C. Harsanyi, (1976), p. 29. Un desarrollo similar puede encontrarse en Ph. Foot, (1972), "Morality as a System of Hypothetical Imperatives".
- (56) J. C. Harsanyi, (1958), en: J. C. Harsanyi, (1976), s. V.
- (57) J. C. Harsanyi, (1958), en: J. C. Harsanyi, (1976), p. 31.
- (58) J. C. Harsanyi, (1978), en: H. W. Gottinger - W. Leinfellner (eds.), (1978), p. 4. Un utilitarismo muy semejante al expuesto por nuestro autor puede encontrarse en la mayoría de los representantes modernos de la teoría del observador imparcial; particularmente similar es la forma de utilitarismo defendida por R. M. Hare, (1976) y (1982).
- (59) J. C. Harsanyi, (1958), en: J. C. Harsanyi, (1976), p. 32.
- (60) La distinción entre "utilitarismo del acto" y "utilitarismo de la regla" es original de R. F. Harrod, (1936), "Utilitarianism Revised", y puede encontrarse elaborada en R. B. Brandt, (1959), Teoría Ética, pp. 408 y ss., y c. XV; J. J. C. Smart, (1956), "Utilitarismo extremo y restringido".
- (61) R. B. Brandt, (1959), y D. Lyons, (1965), Forms and Limits of Utilitarianism; para una tesis contraria, G. Ezorsky, (1968), y A. F. Gibbard, (1965).
- (62) J. C. Harsanyi, (1978), en: H. W. Gottinger - W.

- Leinfellner (eds.), (1978), p. 12.
- (63) J. C. Harsanyi, (1978), en H. W. Gottinger - W. Leinfellner (eds.), (1978), pp. 12-3.
- (64) J. C. Harsanyi, (1978), en: H. W. Gottinger - W. Leinfellner (eds.), (1978), p. 13.
- (65) J. C. Harsanyi, (1978), en: H. W. Gottinger - W. Leinfellner (eds.), (1978), p. 14.
- (66) J. C. Harsanyi, (1978), en: H. W. Gottinger - W. Leinfellner (eds.), (1978), pp. 16-7.
- (67) J. C. Harsanyi, (1978), en: H. W. Gottinger - W. Leinfellner (eds.), (1978), pp. 17-8.
- (68) J. C. Harsanyi, (1978), en: H. W. Gottinger - W. Leinfellner (eds.), (1978), pp. 15-6. Una tesis semejante es el nervio de la argumentación contra el utilitarismo del acto de D. H. Hodgson, (1955), Consequences of Utilitarianism, c. II.
- (69) No vamos, por tanto, a revisar toda la polémica general entre Rawls y el utilitarismo, sino sólo la parte que afecta al modelo de Harsanyi. Remitimos para esa tarea al excelente artículo de H. S. Goldman, (1980).
- (70) Esta similitud la afirma Harsanyi en (1975a), "Can the maximin principle serve as a basis for morality? A Critique of J. Rawls' Theory", en: J. C. Harsanyi, (1976); también K. J. Arrow, (1973), "Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice", en: K. J. Arrow, (1984); y R. M. Hare, (1976).
- (71) J. Rawls, (1974), "Some Reasons for the Maximin Criterion".

(72) Ambos autores coinciden en la idea básica de que un principio para estar fundamentado tiene que ser aceptado -o ser aceptable- unánimemente. Esta es la idea central de la noción de contrato: sobre los principios ha de llegarse a un acuerdo unánime. Pensar en la posibilidad de un acuerdo unánime y pensar en la imparcialidad es todo uno. Puesto que la única forma de considerar que un principio es aceptable unánimemente es pensarlo como no dependiente de las preferencias o posición social de nadie en particular. Aceptabilidad unánime, acuerdo e imparcialidad son los tres conceptos nucleares de un procedimiento contractual y en ellos coincide Harsanyi -y en general la teoría de la elección social- y Rawls. Véase T. M. Scanlon, (1982), "Contractualism and Utilitarianism".

(73) J. Rawls, (1971), A Theory of Justice, pp. 168-9.

(74) J. Rawls, (1971), p. 168.

(75) J. Rawls, (1971), p. 169.

(76) Los tres supuestos que a continuación exponemos son la forma concreta que adoptan las tres condiciones de la posición original que recomiendan la utilización del principio maximin para la efectiva discusión del principio de utilidad media. Estas tres condiciones son:

1. "(...) la situación tiene que ser tal que resulte imposible el conocimiento de las alternativas posibles, o que al menos sea altamente inseguro".
2. "(...) la persona que escoge tiene una concep-

ción 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".

J. Rawls, (1971), pp. 154, 155 y 156. Una apreciación crítica de estas condiciones generales en favor de la regla maximin puede verse en A. Buchanan, (1980), "A Critical Introduction to Rawls' Theory of Justice", en: H. G. Blocker - E. H. Smith (eds.), (1980), pp. 26-30; y una apreciación más técnica en R. P. Wolff, (1977), Para comprender a Rawls, c. XV.

(77) J. Rawls, (1971), p. 172.

(78) En teoría de la decisión se distinguen varios conceptos de probabilidad, cada uno de los cuales está ligado a un modelo diferente de la medición de la incertidumbre:

1. Probabilidad subjetiva (Savage): se deriva de la observación de la conducta del elector.
2. Probabilidad lógica (Carnap): un tipo de probabilidad subjetiva totalmente determinada por condiciones de simetría.
3. Probabilidad empírica (tradicional): estimada sobre la base de hechos observables -e. g., frecuencias estadísticas observadas.

D. J. White, (1969), Teoría de la decisión, s. 2.3.

(79) J. Rawls, (1971), p. 173.

- (80) J. Rawls, (1971), p. 174.
- (81) J.C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), pp. 46-7.
- (82) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 47.
- (83) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 90.
- (84) Este argumento puede verse en J. Rawls, (1971), p. 169.
- (85) Se enfrentan aquí dos posiciones irreductibles: el igualitarismo formal de la justicia, defendido tradicionalmente por el utilitarismo bajo la máxima: "cada uno cuenta como uno, y nadie como más de uno"; y el igualitarismo sustantivo de Rawls que se expresa en esa consideración preponderante que los menos favorecidos tienen en su teoría. Este enfrentamiento, dentro de la teoría de la elección social, será estudiado en nuestro c. IV.
- (86) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), pp. 48-9.
- (87) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 49. Así lo reconoce también K. J. Arrow, (1973), en: K. J. Arrow, (1984), p. 102., que había objetado lo mismo contra los índices de utilidad de von Neumann-Morgenstern.
- (88) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 50.
- (89) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 38.

- (90) Véase D. J. White, (1969), c. 2. Un texto canónico del modelo bayesiano: L. J. Savage, (1954), Foundations of Statistics. Algunos resultados interesantes de este modelo para las decisiones de grupo pueden verse en: H. Moskowitz, (1975), y G. Majone, (1975).
- (91) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 39.
- (92) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 40.
- (93) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 41.
- (94) Véase la respuesta de J. Rawls, (1974), al artículo de J. C. Harsanyi, (1975a), y la contrarespuesta de este en el "Post-Script" a dicho artículo.
- (95) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), pp. 41-2.
- (96) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 42 (subrayados del autor de la cita).
- (97) Este mismo caso es empleado por A. K. Sen, (1970b), para mostrar las consecuencias inaceptables del utilitarismo. Véase nuestro c. IV, s. 1.1.
- (98) J. Rawls, (1971), p. 157.
- (99) Recordemos que el principio de diferencia de Rawls no es la forma reducida que suelen manejar los teóricos de la elección social: "Entre los bienes de una determinada clase prioritaria, las desigualdades habrían de ser permitidas si incrementan la parte de los menos bien situados", K. J. Arrow, (1973), en:

K. J. Arrow, (1984), p. 99. Véase el enunciado completo de los dos principios de justicia, las dos normas de prioridad y la concepción general de la justicia en J. Rawls, (1971), pp. 302-3.

(100) J. Rawls, (1971), p. 158.

(101) K. J. Arrow, (1973), en: K. J. Arrow, (1984), p. 103.

(102) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 42.

(103) J. C. Harsanyi, (1975a), en: J. C. Harsanyi, (1976), p. 43.

CAPITULO TERCERO

RACIONALIDAD COLECTIVA

En el capítulo anterior expusimos un modelo de elección social alternativo al de Arrow y a través de él tratamos de mostrar de qué manera y en qué condiciones el utilitarismo estaba vigente en la teoría de la elección social. Del mismo modo, en el presente capítulo nos ocuparemos de otro modelo alternativo y con él trataremos de mostrar la problemática genérica que la definición de la racionalidad colectiva propone a la teoría de la elección social.

Este modelo es el de Sen y constituye una de las aportaciones más interesantes de la economía para la comprensión de las cuestiones básicas de la filosofía política. Nosotros expondremos las investigaciones de este autor en torno a las condiciones de racionalidad de las elecciones sociales (una parte tan sólo de su muy abundante obra). El hilo conductor es, pues, único, pero multifásico. Así, en la sección primera, al analizar las condiciones formales de racionalidad -la transitividad, básicamente-, el objetivo perseguido por Sen es superar el resultado de imposibilidad de Arrow; el nuestro será ca-

librar si las condiciones de racionalidad propuestas por él pueden sustituir ventajosamente las funciones imprescindibles para la elección social que realiza la transitividad. En cambio, al seguir en la sección segunda la investigación de las condiciones normativas de racionalidad, encontraremos que, aunque el tema de fondo siga siendo la superación del resultado arroviario, éste se presenta de otra manera. Se trata aquí no tanto de superarlo como de demostrar la universalidad del conflicto entre condiciones normativas de racionalidad dentro de una estructura informativa ordinal. Por eso no parte del resultado de imposibilidad arroviario, sino de un modelo más manejable para hacer evidente esta imposibilidad: la paradoja del liberal paretiano.

A través de su análisis y de los intentos de superación, se mostrará la imposibilidad de adecuación entre racionalidad colectiva y valores individuales no como fruto de la naturaleza de los términos en sí mismos, sino como producto de cómo se intentan relacionar; i. e., de cómo se conciben los procedimientos de elección social. Esto es lo que veremos en la sección tercera. Esta crítica de los procedimientos nos llevará insensiblemente a considerar algunos de los supuestos tradicionales y más básicos de la economía. Sen -lo veremos en la sección cuarta- ha desatado una crítica feroz contra ese cúmulo de hipótesis que se encierran en la noción de "hombre económico racional", y ha mostrado que la única manera de superar la problemática que plantea el conflicto entre

lo colectivo y lo individual es dar a los sujetos de las elecciones sociales la capacidad para decidir cómo y con qué materiales se va a realizar la elección social.

Esta idea es el núcleo del modelo de Sen. Para desarrollarla ha sido necesario presentar la crítica que conlleva -explícitamente- de la noción de racionalidad tradicional en economía: la consecuencialista. Pero, también, el modelo de racionalidad que -implícitamente- conlleva y que tiene más que ver con Rawls, Richards y, en general, con la tradición kantiana de nuestros días. Ambas tareas las llevamos a cabo en las secciones quinta y sexta respectivamente.

III.1 Las condiciones formales: crítica de la transitividad.

Arrow exigía que los individuos, si sus preferencias han de ser relevantes para la función de bienestar social, han de tener ordenamientos de preferencias; esto es, las relaciones de preferencia o indiferencia que se establezcan entre las alternativas han de cumplir las propiedades reflexiva, de orden completo y transitiva. Como ya señalamos en nuestro primer capítulo, las dos primeras son propiedades de mero sentido común y sólo la transitividad es una condición sensu stricto de racionalidad. En buena medida el resultado de imposibilidad de Arrow viene determinado por esta exigencia de que las preferencias

se presenten como ordenamientos, puesto que obliga en todos los casos en que se dan ciclos de preferencia a elegir una alternativa, y el único método disponible es haciendo decisivo a un individuo sobre los otros para todas las alternativas (1).

La finalidad con la que Arrow introduce la transitividad y la exigencia de que las preferencias sean ordenamientos es doble. Por una parte, se trata de asegurar que siempre haya una mejor alternativa y, por tanto, de que en cualquier situación haya una elección. El que ningún proceso de elección quede indecيدido es vital para el funcionamiento de los sistemas democráticos; porque, como comenta Mueller, "en la democracia las "vias muertas" constituyen una franca invitación a la dictadura" (2). Por otra parte, se trata de asegurar que, sea cual sea el resultado de la elección, éste no depende de la "ruta" seguida para su adopción; esto es, no ha de depender del proceso por el que se han ido seleccionando las alternativas.

Ahora bien, Sen ha estudiado la posibilidad de mantener estas dos funciones fundamentales para la elección social que cumple la transitividad, pero sin requerir que las preferencias sean transitivas. Para que exista siempre una mejor alternativa no son necesarios los ordenamientos, basta con que se den lo que Sen define como "funciones de elección". La función de elección existe siempre que haya una mejor alternativa dentro de cada subconjunto del conjunto total de las alternativas (3). Esta

exigencia se cumple siempre que dicho subconjunto no sea vacío; es decir, no requiere que haya una preferencia estricta, sino simplemente una preferencia débil (la relación ordinal "al menos tan buena como"). ¿Cuáles son las condiciones para que exista una preferencia débil en cada subconjunto no vacío? Como para los ordenamientos son condiciones necesarias la completitud,

$$\forall x, y \in S: (x \neq y) \rightarrow (x R_y \vee y R_x)$$

puesto que si no se diera esta condición tendríamos que para el conjunto de elección del par (x, y) , $C(S, R)$,

$$-(x R_y) \wedge -(y R_x)$$

y, por tanto, sería vacío. El mismo resultado se obtendría si violara la condición de reflexividad,

$$\forall x \in S: x R_x$$

puesto que habría una alternativa, \underline{x} , que no sería tan buena como ella misma.

Sin embargo, para que exista una preferencia débil entre las alternativas no necesariamente han de cumplir la condición de transitividad,

$$\forall x, y, z \in S: (x R_y \wedge y R_z) \rightarrow x R_z$$

Siendo una condición suficiente -por la que se obtiene un ordenamiento de Arrow-, no es una condición necesaria para la existencia de una función de elección. Con condiciones más débiles se puede obtener una preferencia débil. De este tipo son la cuasi-transitividad,

$$\forall x, y, z \in S: (x P_y \wedge y P_z) \rightarrow x P_z$$

que exige que la preferencia sea transitiva, pero no la indiferencia; y la aciclicidad,