

JOHN C. HARSANYI

FUNCIONES NO LINEALES DE BIENESTAR SOCIAL:
¿TIENEN LOS ECONOMISTAS DEL BIENESTAR UNA
EXENCIÓN ESPECIAL
DE LA RACIONALIDAD BAYESIANA?¹

ABSTRACT. Se argumenta que la teoría de la decisión bayesiana supone una solución a un importante problema filosófico, esto es, el problema de cómo definir el comportamiento racional bajo riesgo e incertidumbre. El autor ha mostrado en artículos anteriores que si tomamos en serio los postulados bayesianos de racionalidad y adoptamos un punto de vista individualista acerca del bienestar social, entonces nuestra función de bienestar social debe ser una función lineal de las utilidades individuales: de hecho, debe ser su media aritmética. El presente artículo critica la opinión de Diamond y Sen de que uno de los postulados bayesianos (a saber, el principio de lo seguro) no se aplica a las decisiones sociales, por más que pueda aplicarse a decisiones individuales. También se critica la propuesta de Sen de convertir el bienestar social en una función no lineal de las utilidades individuales. La función de bienestar social propuesta por el autor depende de comparaciones de utilidad interpersonales. Se defiende el uso de tales comparaciones. Se ha comentado también que cualquiera que piense que la forma utilitarista (es decir, lineal) de la función de bienestar social no es lo suficientemente igualitaria, debería rechazar el axioma del individualismo del autor, en vez de intentar rechazar los axiomas de racionalidad bayesiana. Sin embargo, esto sería equivalente a concederle prioridad a consideraciones igualitarias por encima, en muchos casos, de consideraciones humanitarias. Finalmente, el artículo discute las razones por las que un completo

¹ El autor desea agradecer a la National Science Foundation por apoyar esta investigación mediante la beca GS-3222 al Centro para la Investigación de la Ciencia de la Dirección, Universidad de California, Berkeley. Gracias también al Instituto de Economía Matemática, Universidad de Bielefeld, donde el autor fue Profesor Visitante en 1973-74; así como al profesor Reinhard Selten, de la Universidad de Bielefeld, al profesor Amartya K. Sen, de la London School of Economics, y a los profesores Louis Gevers y Robert Deschamps, ambos de la Universidad de Namur, por sus discusiones tan provechosas.

acuerdo en la fórmula matemática de la función de bienestar social no ocasionaría un estado utópico de consenso moral: todavía existirían controversias morales que surgirían de desacuerdos acerca de qué predicciones hacer sobre hechos empíricos futuros.

1. INTRODUCCIÓN

Además de problemas sobre los hechos empíricos y de validez formal (lógica y matemática), un tercer tipo importante de problemas teóricos son los de buscar un concepto matemático riguroso (generalmente, pero no necesariamente, un concepto definido por axiomas formales) como posible sustituto de un concepto vago y poco claro del discurso precientífico de sentido común (Carnap, 1950, pp. 3-8). Describiré los problemas de este tipo como problemas *conceptuales* o *filosóficos*. Estos juegan un papel principal en las discusiones filosóficas. Pero, también, surgen en economía y otras ciencias sociales. Son ejemplo de esto los problemas de cómo definir el comportamiento racional bajo certeza, riesgo e incertidumbre, así como en situaciones de juego; cómo definir el concepto de interés público o, más exactamente, cómo definir una función de bienestar social que formalice adecuadamente nuestra noción intuitiva del interés público; cómo definir el poder social, o el nivel social, etc. (Como todo el que esté familiarizado con la física moderna sabe, los problemas filosóficos de este tipo también surgen en las ciencias naturales, aunque el acento esencial será obviamente diferente).

La teoría bayesiana del comportamiento racional bajo riesgo e incertidumbre (la teoría de la decisión bayesiana) es uno de los pocos casos en los que un problema tan esencialmente filosófico ha encontrado una solución específica y concreta, aceptada como correcta por una creciente convergencia de la opinión experta. (El problema de cómo definir el comportamiento racional bajo certeza ya ha sido solucionado por la teoría económica clásica, mientras que el problema de cuál debería ser la definición precisa del comportamiento racional en situaciones de juego es todavía una cuestión abierta, aunque quizás ahora se estén haciendo progresos hacia su solución (ver, p. ej. Harsanyi, 1975b; cf. también Harsanyi, 1966).)

Propongo aquí resumir brevemente los argumentos a favor de la posición bayesiana: (1) los axiomas (postulados racionales) de la teoría bayesiana son intelectualmente obligatorios para cualquiera que se haya tomado la molestia de entenderlos adecuadamen-

te; y ningún argumento se ha propuesto, todavía, que arroje serias dudas sobre estos axiomas como criterios sensatos para el comportamiento racional². (2) La teoría bayesiana no parece conducir a implicaciones contrarias a la intuición. (3) Se sabe que las principales definiciones rivales del comportamiento racional bajo riesgo e incertidumbre tienen implicaciones altamente contrarias a la intuición (Radner y Marschack, 1954).

En dos artículos anteriores (Harsanyi, 1953 y 1955) he argumentado que si tomamos los postulados racionales de la teoría bayesiana seriamente, entonces podemos obtener una solución clara y concreta también para el tradicional problema filosófico de definir una función de bienestar social adecuada. De hecho, puede demostrarse que la función de bienestar social debe ser una función lineal de todas las utilidades individuales —o, más exactamente, debe ser definida como la media matemática de los niveles de utilidad de todos los individuos de la sociedad.

Para establecer este resultado, debemos distinguir entre dos clases de preferencias expresadas por cualquier individuo dado. Una clase incluye sus preferencias *personales*, basadas en su gusto y sus intereses personales (así como en los intereses de aquellos individuos, si los hay, cuyo bienestar es un asunto de interés personal para él). La otra clase comprende sus preferencias *morales* o *sociales*, que expresan sus puntos de vista sobre los intereses generales de la sociedad, definidos en términos de criterios impersonales e imparciales. Suponiendo que ambas clases de preferencias satisfagan ciertos requisitos de consistencia (y continuidad), las preferencias personales de cada individuo pueden ser representadas por su *función de utilidad*, mientras que sus preferencias morales pueden representarse por su *función de bienestar*

² Los postulados de la teoría bayesiana tienen un significado muy claro en todas aquellas situaciones en donde la utilidad de cada 'premio' es independiente de la probabilidad de conseguirlo. (Virtualmente todas las situaciones importantes para la economía del bienestar pertenecen a esta categoría.) Las dificultades de interpretación surgen sólo cuando éste no es el caso, por ejemplo, cuando parte de la atracción por la escalada reside en el peligro inherente; es decir, en el hecho de que la probabilidad de llegar a salvo a la cima es menor de uno (cf. Marschak, 1950). Pero incluso en tales casos los postulados conservan su validez formal, si los 'premios' se definen adecuadamente (por ejemplo, la escalada debe definirse de tal manera que incluya la presencia de peligro, o, más exactamente, la satisfacción del peligro vencido con éxito). Sin embargo, esto significa que en tales casos las implicaciones de comportamiento de los postulados bayesianos se volverán algo ambiguas a menos que estos postulados puedan complementarse con una teoría psicológica que prediga cómo dependerán cuantitativamente las utilidades de los premios de sus probabilidades.

social. Por lo tanto, bajo este modelo, cada individuo tendrá una función de bienestar social propia, posiblemente diferente de las funciones de bienestar social de otros individuos. (Aunque, como he argumentado (Harsanyi, 1955, p. 320), con información suficiente acerca de las funciones de utilidad de otras personas, las funciones de bienestar social de individuos diferentes llegarán a ser idénticas.)

Para representar las relaciones matemáticas entre la función de bienestar social de un individuo dado (el individuo *j*) y las funciones de utilidad de todos los individuos de la sociedad, propuse los siguientes postulados (Harsanyi, 1955, pp. 312-314). *Postulado a (racionalidad individual)*: las preferencias individuales de todos los individuos satisfacen los axiomas de racionalidad bayesiana. *Postulado b (racionalidad social)*: las preferencias morales (o sociales) del individuo *j* satisfacen los axiomas de racionalidad bayesiana. *Postulado c (individualismo)*: si todos los individuos son *personalmente* indiferentes entre dos situaciones sociales, entonces el individuo *j* será *moralmente* indiferente entre ambas.

Obviamente, el postulado *a* implica que las preferencias personales de cada individuo pueden ser representadas por una función de utilidad de von Neumann-Morgenstern (es decir, por una función de utilidad cardinal que mide la utilidad de cualquier lotería por su utilidad esperada). De la misma forma, el postulado *b* implica que la función de bienestar social que representa las preferencias morales del individuo *j* tendrá la naturaleza de una función de utilidad de von Neumann-Morgenstern. Finalmente, los tres postulados juntos implican que la función de bienestar social de *j* será una *combinación lineal* de utilidades individuales.

El mismo resultado, o más bien un resultado algo más fuerte, puede obtenerse mediante un análisis conceptual de la naturaleza de las preferencias morales (los juicios de valor moral). Como ya he indicado, las preferencias morales difieren de las preferencias meramente personales en que se basan en criterios impersonales e imparciales. Ahora supongamos que un individuo dado *j* eligiera entre dos situaciones sociales en completa *ignorancia* —o al menos con una *despreocupación* voluntaria completa— sobre cuál sería su posición personal en cada situación. Especificando aún más, supongamos que él haría su elección bajo la suposición de que tendría exactamente la *misma probabilidad* de ocupar cualquiera de las posiciones sociales disponibles y, efectivamente, de *adoptar* la identidad personal de cualquier individuo en la sociedad. Claramente, bajo esta disposición de cosas, su elección satisfaría necesariamente los requisitos de impersonalidad e

imparcialidad en el mayor grado posible y, por lo tanto, se calificaría plenamente como un enunciado de preferencia moral, es decir, como un juicio de valor moral.

Al mismo tiempo, su elección sería una elección que supondría *riesgo* (ya que sólo podría hacer predicciones probables sobre su posición personal en cada situación). Por lo tanto, según la teoría de la decisión bayesiana, él tendría que elegir la situación social que le proporcionase una mayor utilidad esperada, que en este caso significaría elegir la situación que concediese un *nivel medio de utilidad* más alto a los individuos de la sociedad. Por lo tanto, nuestro modelo nos lleva a la conclusión de que enunciar una preferencia moral (un juicio de valor moral) es equivalente a expresar una preferencia basada en la media aritmética de todas las utilidades individuales como la función de bienestar social de uno (Harsanyi, 1953).

En años anteriores, mi teoría ha sido citada con aprobación por algunos distinguidos economistas (ver, p. ej., Theil, 1968, p. 336). Pero ha sido también criticada (Diamond, 1967; Rawls, 1971; Sen, 1970 y 1973)³. La crítica más específica fue la de Diamond, quien al menos reconoció claramente que mi teoría puede ser rebatida sólo si uno rechaza uno o más de sus axiomas. Por el contrario, Rawls nunca ha aclarado qué axioma(s) en concreto de mi teoría, si hay alguno, desea rechazar. Diamond ha elegido rechazar el principio de lo *seguro* cuando se aplica a la decisión social (lo cual supone rechazar mi postulado *b*). Sen (1970, p. 145) parece adoptar una posición similar. Más adelante, consideraré el argumento de Diamond más detalladamente. Intentaré demostrar que los economistas del bienestar social no están más autorizados que las personas con profesiones de menos categoría a rechazar el principio de lo *seguro* u otros axiomas bayesianos de racionalidad; e intentaré perfilar alguna de las curiosas implicaciones del punto de vista de Diamond (y de Sen).

También argumentaré que la propuesta de Sen (1973, p. 18) de hacer nuestra función de bienestar social dependiente no sólo del valor medio de las utilidades individuales, sino también de alguna medida de desigualdad (dispersión) de estos niveles de utilidad, está abierta a las mismas objeciones que la opinión de que la utilidad de un billete de lotería debería depender no sólo de su utilidad (media) esperada, sino también de alguna medida

³ En este artículo discutiré principalmente las críticas de los profesores Diamond y Sen, dado que ya he considerado en otra parte la teoría del profesor Rawls (Harsanyi, 1975a).

de riesgo (de alguna medida de dispersión de los resultados de utilidad posibles).

Está suficientemente claro por qué algunos economistas y filósofos no están contentos con la teoría utilitarista que se deriva del uso de una función lineal de bienestar social. Mientras que Robbins (1938) se opuso una vez al utilitarismo porque, aparentemente, pensaba que tendría implicaciones *demasiado* igualitarias, en nuestra época la mayoría de las objeciones vienen probablemente de personas que encuentran que el utilitarismo *no es lo suficientemente* igualitario. Argumentaré que cualquier persona que adopte la última posición no debería intentar amañar los postulados de racionalidad de la teoría bayesiana; esto podría llevar a resultados altamente contrarios a la intuición. Por el contrario, debería rechazar el *postulado del individualismo* (postulado c) de mi teoría (véase arriba; y también Harsanyi, 1955, p. 313), que hace que las preferencias sociales estén completamente determinadas por las preferencias individuales.

Esto significa que cualquiera que quiera adoptar una posición moral más igualitaria de lo que ya es la posición utilitarista, debe admitir que el bienestar de los miembros individuales de la sociedad *no* es el valor moral fundamental y que *está dispuesto*, en ciertos casos, a sacrificar consideraciones humanitarias a objetivos igualitarios cuando hay un conflicto entre ambos. Esta es la verdadera cuestión moral aquí; y esta importante cuestión moral no debería obscurecerse por objeciones superficialmente atractivas —pero en realidad bastante insostenibles— al principio de lo seguro o a cualquier otro axioma de racionalidad.

2. LA ECONOMÍA DEL BIENESTAR Y EL PRINCIPIO DE LO SEGURO

Para nuestros propósitos es suficiente considerar el principio de lo seguro tal y como se enuncia para situaciones de *riesgo* (donde todas las probabilidades son conocidas por el sujeto que decide). En este caso, el principio sostiene que, siendo lo demás igual, es siempre preferible tener una oportunidad de ganar un *premio de más alto valor* con una probabilidad positiva dada a tener la oportunidad de ganar un *premio de menor valor* con la misma probabilidad. Una variante de este principio es el principio de sustitución: no existe ninguna diferencia entre ganar un premio u otro con una probabilidad dada, si uno considera estos dos premios como *igualmente valiosos*. Ambas formas del principio de lo seguro son requisitos intelectualmente obligatorios del comportamiento racional y

es difícil imaginar una situación en la que un individuo racional encontrara justificado violar cualquiera de ellos.

Diamond admite que el principio de lo seguro es una regla sensata para la elección individual, pero le niega cualquier validez para elecciones sociales. Incluso a primera vista, sería sorprendente que esta opinión fuera correcta. Seguramente, cuando actuamos en nombre de otras personas, por no hablar de cuando actuamos en nombre de la sociedad, estamos bajo la obligación de seguir criterios de racionalidad *más altos* si cabe que cuando tratamos nuestros propios asuntos. Si la prudencia común requiere que los individuos cumplan el principio de lo seguro, entonces a los funcionarios gubernamentales que se supone que cuidan de los intereses comunes no se les puede perdonar el que no hagan lo mismo. Como tampoco los economistas del bienestar pueden aconsejar razonablemente a estos funcionarios que hagan lo contrario. Por supuesto, esto no es un argumento definitivo: después de todo, *podría* haber razones muy especiales por las que los funcionarios públicos y los economistas del bienestar no se sintieran obligados por estos criterios de racionalidad bayesianos. Para llegar a conclusiones más firmes, tenemos que considerar el argumento de Diamond en términos más específicos.

Diamond imagina una sociedad hipotética que consta de dos individuos, donde el gobierno puede elegir entre dos políticas alternativas. Una política produciría el vector de utilidad (1,0) con certeza. La otra produciría los dos vectores de utilidad (1,0) y (0,1) con probabilidades iguales. Suponiendo que esta sociedad le concediera igual importancia a los intereses de los dos individuos, es fácil ver que, según el principio de lo seguro (principio de sustitución), a las dos políticas se les debería asignar el mismo valor desde un punto de vista social. Pero, según la propia opinión de Diamond, la segunda política es, de hecho, socialmente preferible a la primera sólo porque concedería a *ambos* individuos una 'oportunidad justa', mientras que la primera no lo haría. Según Diamond, esto demuestra que el principio de lo seguro no tiene validez para decisiones sociales.

Para llegar a un mejor entendimiento del argumento de Diamond, propongo aplicarlo a un par de situaciones hipotéticas más concretas. Por ejemplo, imaginemos dos sociedades, *A* y *B*. La sociedad *A* tiene una distribución de ingresos extremadamente injusta, tan extrema que, de hecho, incluso observadores políticamente muy conservadores la encuentran repugnante. Además, la movilidad social es casi imposible y esta movilidad no se basa, en absoluto, en lo que podríamos llamar mérito individual. La socie-

dad *B* es exactamente igual a la sociedad *A*, excepto por la siguiente diferencia. Debido a una antigua costumbre social, todos los niños nacidos en *B* durante un determinado mes del calendario son aleatoriamente distribuidos por funcionarios entre todas las familias que tuvieron un hijo durante ese período, así que cada niño nacido en ese mes tendrá la misma oportunidad de acabar en una familia determinada. (Supondré que todas las familias aceptan por completo esta costumbre social y tratan a los niños que les son fortuitamente asignados como si fueran suyos.)

¿Deberíamos entonces decir que la sociedad *B* sería moralmente menos censurable que la sociedad *A*, porque en *B* todos los individuos tendrían una 'oportunidad justa' de acabar en una familia rica y, por lo tanto, en una posición social y económica privilegiada? Por hipótesis, *B* es una sociedad con una distribución de ingresos tan injusta como *A*. En ambas sociedades, la posición social y económica de cualquier individuo no tiene nada que ver con el mérito personal, sino que es más bien una cuestión de 'suerte'. En *A* depende completamente del accidente de nacer —de la 'gran lotería de la vida' que decide quién nace en una familia determinada. Por el contrario, en *B* depende completamente de una lotería dirigida por el gobierno. ¿Por qué deberíamos asignarle a una lotería organizada por burócratas del gobierno una dignidad moral más elevada que a la 'gran lotería de la vida' que elige una familia para cada uno de nosotros sin la ayuda de la intervención gubernamental? ¿Por qué debería una lotería burocrática ser considerada como un mecanismo 'más justo' de distribución que la gran lotería biológica que produce la naturaleza?

De hecho, supongamos que conseguimos información fiable especificando que las familias en las que nacemos son siempre literalmente escogidas por una enorme lotería celestial. ¿Puede alguien afirmar seriamente que esta información metafísica variaría de alguna manera nuestra condena social de las injusticias sociales y económicas?

Ahora, consideremos otro ejemplo. Supongamos que el gobierno puede elegir entre dos políticas. La primera política consistiría en abolir una tarifa protectora obsoleta y, en estos momentos, económicamente perjudicial. Esto beneficiaría a todos los ciudadanos (a muchos de ellos muy sustancialmente), excepto a un grupo de obreros y empleados de la hasta ahora protegida industria, los cuales sufrirían moderadas pérdidas económicas. La segunda política nos llevaría al mismo vector de utilidades individuales que la primera política, excepto que los componentes individuales de este vector serían cambiados al azar, porque ahora los ganadores y per-

dedores serían elegidos por una lotería dirigida por el gobierno. Para precisar nuestras ideas, supondremos que la segunda política consistiría realmente en poner en práctica la primera política (es decir, eliminar la tarifa protectora), seguida de una redistribución arbitraria de ingresos basada en una lotería.

Una vez más, ¿tendría sentido afirmar que la segunda política sería moralmente preferible a la primera? Bajo la primera política, los perdedores serían los miembros de una industria en particular, que presumiblemente habrían entrado a formar parte de esta industria por asociación familiar o por otros accidentes de la vida personal. Por lo tanto, ser un miembro del grupo perdedor sería tanto una cuestión de 'mala suerte' personal como lo sería en la segunda política, donde los perdedores serían elegidos literalmente por lotería. De nuevo, ¿qué haría de esta lotería un mecanismo de distribución moralmente superior a aquellos accidentes de la historia que hacen que las personas formen parte de una industria en particular?

Sería fácil aducir muchos más ejemplos para corroborar la misma conclusión. La sugerencia de Diamond de que los privilegios sociales y económicos distribuidos por políticas gubernamentales con componentes arbitrarios son moralmente más aceptables que los privilegios sociales y económicos distribuidos por accidentes de nacimiento e historia personal —es decir, la sugerencia de que el primer tipo de 'suerte' personal es moralmente superior al segundo tipo— no merece ni ser considerada. Por lo tanto, su afirmación de que las elecciones sociales no están sujetas al principio de lo seguro se cae por su propio peso.

Déjenme añadir que la misma conclusión se aplica a la distribución de cualidades biológicas tales como la inteligencia, el talento artístico y científico, la salud, la belleza, la fortaleza física, etc.. Bajo garantías apropiadas, las técnicas de ingeniería genética podrían, sin duda, beneficiar a la humanidad mejorando la herencia biológica del individuo medio de generaciones futuras. Pero una mera redistribución de las mismas cualidades biológicas sin desarrollar por medio de una lotería oficial (¡qué idea tan horrible!), incluso si esto fuera tecnológicamente factible, ciertamente no representaría desde un punto de vista moral ninguna mejora de la situación existente.

3. UTILIDAD MEDIA VERSUS DISPERSIÓN DE LA UTILIDAD

Mientras que mi propia teoría haría de la función de bienestar social la media aritmética de las utilidades individuales, Rawls (1958 y 1971) ha propuesto una función de bienestar social basada en el principio maximin que siempre mide el nivel de bienestar de la sociedad por el nivel de utilidad del individuo peor parado. Esto significa matemáticamente que la función de bienestar social de Rawls siempre asignará un *peso infinitamente mayor* a los intereses de los más pobres —o a los de los miembros de la sociedad de cualquier otra manera menos afortunados— del que le asigna a los de los miembros más ricos —o a los de los miembros de cualquier otra manera más afortunados. Por el contrario, mi propia función de bienestar social siempre asignaría el *mismo* peso a las necesidades *igualmente apremiantes* de diferentes individuos independientemente de sus posiciones sociales o económicas. Pero, por supuesto, dado que típicamente las personas pobres tienen muchas más necesidades apremiantes insatisfechas que las personas ricas, en la práctica y en la mayoría de los casos mi propia función de bienestar social conducirá a decisiones sobre políticas similares a las de Rawls, porque dará mayor prioridad a las necesidades de los pobres. Sólo en aquellos casos, bastante especiales, en los que algunas personas ricas puedan tener necesidades más apremiantes que las de los pobres, mi función de bienestar social llevará a decisiones opuestas sobre políticas. Sin embargo, precisamente en estos casos, pienso que mi función de bienestar social conduciría a decisiones moralmente justas mientras que la de Rawls conduciría a decisiones injustas.

Por ejemplo, si un filántropo tiene que decidir si dar \$100 en efectivo, o el mismo valor en comida, o en ropa, etc., a un hombre pobre o a un millonario, está claro que él debería escoger al hombre pobre ya que éste tendrá una mayor necesidad de dinero extra, o comida extra, o ropa extra, etc. Pero si tiene que elegir entre dar una medicina escasa para salvar la vida a un hombre pobre o a un millonario, entonces la única consideración relevante debe ser quién la necesita más, es decir, quién podría obtener un mayor beneficio médico de ella. Seguramente, sería una discriminación bastante inmoral negarle al rico la medicina, cuando hay mejores razones desde el punto de vista médico a que él la tenga, meramente porque da la casualidad de que él es rico. Sin embargo, la función social de Rawls nos forzaría, precisamente, a comprometer-

ternos en tales prácticas inmorales de discriminación en algunos casos bastante especiales, pero quizás no del todo raros⁴.

En su interesante libro, *On Economic Inequality*, Sen (1973) ha adoptado una postura intermedia entre la de Rawls y la mía. Está de acuerdo con Rawls en que la función de bienestar social debería dar *más peso* a las funciones de utilidad de las personas pobres que a las de los ricos. Pero, bajo la teoría de Sen, a las primeras se les daría sólo un número *finito* de veces más peso que a las segundas. Mi propia teoría igualaría el bienestar social al valor medio de los niveles de utilidad de los diferentes individuos, de esta manera el bienestar social sería una función lineal de las utilidades individuales. Por el contrario, la teoría de Sen haría depender el bienestar social no sólo de la media, sino también de cierta medida de *desigualdad* —es decir de *dispersión*— entre los niveles de utilidad de diferentes individuos (Sen, 1973, p. 18)⁵. Esto significa que el bienestar social tendría que ser una función no lineal de las utilidades individuales⁶. Para una referencia más fácil, describiré este argumento en favor de una función de bienestar social no lineal como argumento de la dispersión de la utilidad en relación al bienestar social.

Desafortunadamente, la teoría de Sen está abierta a la misma objeción que la teoría de Rawls: daría lugar a discriminaciones injustas en contra de las personas que disfrutasen de niveles de utilidad relativamente altos, de la misma manera —aunque no tan frecuentemente— que la de Rawls. Además, el argumento de dispersión de la utilidad en relación al bienestar social —que subraya la teoría de Sen— muestra un parecido formal estrecho con lo que yo llamaré el argumento de dispersión de la utilidad en relación a las loterías, a saber, con el punto de vista de que la utilidad de un billete de lotería debería depender no sólo de su *utilidad (media) esperada*, sino también de cierta medida de *riesgo*, es decir, de cierta medida de *dispersión* de los resultados de utilidad posibles.

⁴ Para una discusión más detallada del punto de vista de Rawls, se refiere al lector a Harsanyi (1975a).

⁵ Sen también considera medidas de desigualdad que no establecen una ordenación completa entre distribuciones de ingresos alternativas (distribuciones de utilidad), sino que sólo establecen una ordenación parcial (Sen, 1973, pp. 55-65).

⁶ Cualquier función de bienestar social lineal y simétrica de las utilidades individuales es esencialmente la misma que su media aritmética, aparte una posible constante multiplicativa o aditiva. Por lo tanto, si queremos que nuestra función de bienestar social dependa de cierta medida de desigualdad (de dispersión) entre los niveles de utilidad individuales, tenemos que hacerla no lineal sobre las utilidades individuales.

Por lo tanto, según va el argumento, la utilidad de un billete de lotería debe ser una función no lineal de estos resultados de utilidad posibles.

Sin embargo, se sabe que el argumento de dispersión de la utilidad en relación a las loterías es erróneo (véase Luce y Raiffa, 1957, p. 32; o véase cualquier libro de texto sobre la teoría de la decisión). Ciertamente, un argumento similar sería perfectamente válido si todas las referencias a las *utilidades* de los posibles resultados (premios) se reemplazaran por referencias a su *valor monetario*. Es completamente cierto que, para la mayoría de los que toman decisiones, el valor de un billete de lotería no sólo está determinado por el *ingreso de dinero (medio) esperado* que se asocia a él. Más bien, depende fuertemente del *riesgo* que implica, es decir, de la *dispersión* de los posibles ingresos monetarios.

Este argumento, sin embargo, no puede extenderse de los posibles ingresos monetarios a sus utilidades. La razón básica reside en el hecho de que la utilidad de cualquier posible ingreso de dinero se mide por la función de utilidad de von Neumann-Morgenstern del que toma la decisión, la cual ya ha tenido en cuenta su actitud hacia el riesgo. Por ejemplo, si se tiene una actitud negativa hacia el riesgo, entonces la función de utilidad desplegará una utilidad marginal decreciente para el dinero (es decir, será estrictamente cóncava en dinero). Por lo tanto, la aversión hacia el riesgo se verá reflejada completamente en las utilidades que se asignen a los varios posibles ingresos y, por lo tanto, también en la utilidad esperada asociada al billete de lotería. Por lo tanto, sería *innecesario e inadmisibile* contarla dos veces si tuviéramos en cuenta por segunda vez la aversión al riesgo del que toma la decisión y así se hiciera depender la utilidad del billete de lotería no sólo de su utilidad esperada, sino también de la dispersión de las utilidades alcanzables.

De hecho, tal procedimiento no sólo sería innecesario; sería *inconsistente desde el punto de vista lógico*. Esto es así porque se puede demostrar que, si el que toma la decisión sigue los postulados de racionalidad de la teoría bayesiana, entonces *debe* medir la utilidad de un billete de lotería por su utilidad esperada (suponiendo que todas las utilidades se expresan en las unidades de utilidad de von Neumann-Morgenstern). Esto significa, simplemente, que *no puede* hacer que su utilidad dependa de la dispersión de los posibles resultados de utilidad, y que no puede hacerla una función no lineal de la última.

Ahora me propongo demostrar que el argumento de la dispersión de la utilidad de Sen en relación al bienestar social sucumbe,

esencialmente, ante la misma objeción que el argumento de la dispersión de la utilidad en relación a las loterías: es una transferencia ilegítima de una relación matemática para una cantidad de dinero, para la que es válida, a niveles de utilidad, para la que no es válida. Es completamente cierto que el bienestar social no puede equipararse al ingreso medio por persona, y no puede, ni siquiera, ser una función de esta *única* variable. Por el contrario, si el ingreso por persona se mantiene constante, entonces, como la mayoría de nosotros seguramente aceptaríamos, el bienestar social aumentaría con una distribución más igualitaria de ingresos. Esto significa que el bienestar social es una función no lineal (de hecho, presumiblemente cóncava) de los ingresos individuales.

Esto es así porque, si nosotros podemos redistribuir los ingresos sin costes y podemos transferir \$100 de un hombre rico a uno pobre, entonces, por regla general, el perjuicio (la no utilidad) que se inflige al primero será considerablemente menor que el beneficio (la utilidad) que le corresponde al último. Esto se deriva de la ley de utilidad marginal decreciente del dinero para cada individuo⁷ (junto con la suposición de que individuos diferentes tendrán funciones de utilidad razonablemente similares para el dinero, de manera que la ley de utilidad marginal decreciente del dinero les afectará de una manera similar). Por supuesto, este argumento tiene sentido sólo si son *comparables* las ganancias y pérdidas de utilidad de diferentes individuos. En Harsanyi (1955, pp. 331-321), he explicado mis razones para creer en la posibilidad de comparaciones de utilidad interpersonales. (Creo que Sen está de acuerdo conmigo en este punto (Sen, 1973, p. 14).) Volveré a este problema más adelante (Sección 4).

Sin embargo, si incluso el bienestar social es una función no lineal de los ingresos individuales, de aquí no se deriva en absoluto que sea también una función no lineal de las *utilidades* individuales. Si disminuimos el nivel de utilidad de un hombre rico en 100 unidades de utilidad y simultáneamente aumentamos el nivel de utilidad de un hombre pobre en 100 unidades de utilidad, entonces (suponiendo que hemos medido la utilidad en las mismas unidades en ambos casos) la pérdida de utilidad sufrida por el primero será exactamente la *misma* que la ganancia de utilidad que

⁷ La fuerza de este argumento no disminuirá significativamente incluso si aceptamos la opinión de Friedman y Savage (1948) de que las funciones de utilidad de las personas pueden tener algunos niveles —relativamente cortos— en que aumente la utilidad marginal del dinero. Globalmente, estas funciones todavía mostrarán una tendencia de conjunto a la disminución en la utilidad marginal del dinero.

le corresponde al segundo. Tiene bastante sentido asumir una ley de utilidad marginal decreciente para el *dinero* (o para las comodidades); pero no tendrá sentido, en absoluto, suponer una ley de utilidad marginal decreciente para la *utilidad*. (No tendría sentido afirmar que un aumento de utilidad de 1.000.000 de unidades a 1.000.100 de unidades es un incremento de utilidad más pequeño que otro de 200 unidades a 300 unidades. Describiendo ambos como incrementos de 100 unidades, nos comprometemos a suponer su igualdad).

Lo que está en juego aquí, por supuesto, no es una simple cuestión (más bien trivial) de matemáticas. Por el contrario, como ya he argumentado al discutir la teoría de Rawls, estamos tratando una cuestión *moral* importante. Cuando le asignamos la misma medida cuantificadora a cambios de utilidad que afectan a dos individuos diferentes (por ejemplo, cuando describimos ambos como cambios de 100 unidades), entonces estamos implícitamente afirmando que estos cambios de utilidad para ambos individuos conllevan necesidades humanas igualmente apremiantes. Pero, siendo esto así, sería una discriminación bastante injusta —y, en muchos casos, bastante inhumana— sostener que, por principio, la satisfacción de las necesidades de un hombre debería tener una prioridad social menor que la satisfacción de las necesidades de otro hombre. Aunque he ilustrado este punto con un ejemplo (relativo a una medicina que salva la vida), propondré ahora un segundo ejemplo porque éste me dará la oportunidad de discutir algunos otros aspectos del problema.

Supongamos que hay dos chicos de cinco años en mi vecindario. Uno de ellos, *A*, es un chico de temperamento afortunado, que parece ser feliz la mayoría del tiempo y que encuentra gran placer en los pequeños regalos. El otro chico, *B*, tiene un temperamento bastante desafortunado. Parece infeliz la mayor parte del tiempo y los pequeños regalos le causan poca satisfacción. Da la casualidad de que yo tengo un pequeño regalo en mi bolsillo. ¿A qué chico debería dárselo?

La teoría utilitarista tiene una respuesta clara a esta pregunta: el regalo debería ir a aquel niño que mayor utilidad sacara de él. Presumiblemente, esto significa que debería ir a *A*, del que se espera que consiga un disfrute más inmediato de él. (Pero esta conclusión tendría que invertirse, en el caso de que yo pensara que hay una esperanza razonable de que el recibir regalos y otras muestras de atención pudiera tener a la larga un efecto suficientemente beneficioso en la personalidad desfavorable de *B*.)

Por el contrario, la teoría de Rawls *siempre* estaría a favor de dar el regalo a *B*, que es obviamente el chico menos afortunado de los dos. Finalmente, la teoría de Sen sugeriría que debería dársele a *A* o a *B*, dependiendo de la distancia real entre los niveles de utilidad de los dos chicos: si el nivel de utilidad de *B* *no* está muy por debajo del de *A*, entonces la mayor utilidad marginal para *A* del presente debería ser el factor decisivo y el regalo debería ir a *A*; mientras que si el nivel de utilidad de *B* *está* muy por debajo del de *A*, entonces esta diferencia en sus niveles de utilidad será el factor decisivo y el pequeño regalo debería ir a *B*.

Una vez más, la cuestión es ésta. En caso de que llegara a la conclusión de que, una vez considerado todo, yo podría crear más felicidad dándole el regalo a *A*, ¿se me permite hacerlo? O, ¿se me pide que bajo ciertas circunstancias discrimine a *A* y le dé el regalo a *B*, incluso aunque sepa que *A* haría un mejor uso de él —meramente porque *A* es ya un muchacho bastante feliz?

Tal y como están las cosas, *A* tiene un nivel de utilidad bastante alto, mientras que *B* lo tiene bastante bajo. Esta situación no la he creado yo: ya existía cuando yo aparecí en escena. La cuestión es qué obligaciones me impone a mí tal estado de cosas. Sen y yo coincidimos en que una obligación que tengo es la de crear tanta felicidad humana como pueda en esta situación. Pero Sen parece sostener que estoy también bajo una segunda obligación (que en algunos casos puede anular a la primera) de *compensar a B* por un nivel de utilidad tan bajo como el suyo. En mi opinión, la cuestión de compensar a *B* no se plantea en absoluto. Por supuesto, yo ciertamente le debería a *B* una compensación justa si su bajo nivel de utilidad fuera el resultado de acciones culpables mías en el pasado. Pero, bajo nuestras suposiciones, éste no es el caso. Por lo tanto, los sentimientos de *culpabilidad* injustificados acerca del bajo nivel de utilidad de *B* no deberían tener influencia alguna en mi comportamiento. (Ni tampoco debería mi comportamiento verse influido, por supuesto, por ningún *resentimiento* irracional contra las personas que, como *A*, son lo suficientemente afortunadas como para disfrutar de niveles altos de utilidad, logrados sin recurrir a actividades moralmente reprobables.) Más bien, mi única obligación en esta situación es conseguir la meta básica de todas las acciones morales buenas, a saber, crear tanta felicidad como sea posible en este mundo.

Para concluir, hemos hallado que el argumento de la dispersión de la utilidad en relación al bienestar social de Sen representa una extensión ilegítima de una relación matemática propia de los ingresos a las utilidades correspondientes, del mismo tipo que

el argumento de la dispersión de la utilidad en relación a las loterías. También hemos hallado que las razones por las que esta extensión es ilegítima son muy similares en ambos casos. La razón básica es que, dada la no linealidad –concavidad o cuasi-concavidad– de las funciones de utilidad individuales en términos de dinero y comodidades (es decir, su tendencia a mostrar una utilidad marginal decreciente), las necesidades de los pobres tendrán normalmente una prioridad social mayor que las de los ricos, incluso si usamos una función lineal del bienestar social. Por lo tanto, es bastante *innecesario* y, de hecho, es *moralmente inadmisibile*, hacer que la función de bienestar social sea no lineal sobre las utilidades individuales –como es innecesario e inadmisibile hacer que la utilidad de un billete de lotería sea no lineal sobre los posibles resultados de utilidad.

De hecho, tal función de bienestar social no lineal sería también *inconsistente desde el punto de vista lógico*, al menos para las personas que creen en los postulados de racionalidad bayesianos y en la filosofía moral humanista e individualista. Ya que, como he mostrado (Harsanyi, 1995, pp. 312-314), los postulados de racionalidad bayesianos, así como un axioma de individualismo (postulado *c*), implican lógicamente el uso de una función lineal del bienestar social.

4. COMPARACIONES INTERPERSONALES DE UTILIDAD

Los postulados de racionalidad bayesianos y el axioma del individualismo sólo implican que nuestra función de bienestar social debe ser lineal sobre las utilidades individuales: pero no dicen nada acerca del peso que deberíamos darle a las funciones de utilidad de los diferentes individuos. Sin embargo, es natural complementar esta teoría añadiendo un axioma de simetría, que requiere que la función de bienestar social trate las funciones de utilidad de diferentes individuos de una manera similar y asigne el mismo peso a cada una de ellas⁸.

Este requerimiento, a su vez, implica que nuestra función de bienestar social debe basarse en *comparaciones interpersonales de utilidad*. Ya que, para asignar el mismo peso a las funciones de utilidad de diferentes individuos, debemos poder expresar todas

⁸ Alternativamente, para establecer este requisito de pesos iguales, podríamos utilizar mi modelo de equiprobabilidad para las preferencias morales (juicios morales de valor). Ver sección 1 anterior; y también Harsanyi (1953).

ellas en la misma unidad de utilidad. Por supuesto, cuando definimos por primera vez una función de utilidad de von Neumann-Morgenstern para cada individuo a la manera usual, elegiremos normalmente una unidad de utilidad independiente para cada individuo. Pero, entonces, debemos emprender comparaciones interpersonales de utilidad para poder estimar las *ratios de conversión* entre las unidades de utilidad de los diferentes individuos. (Por ejemplo, primero podemos elegir como unidad de utilidad para cada individuo su utilidad para \$1 extra. Pero entonces debemos intentar estimar cómo se comparan unas con otras las utilidades marginales de los diferentes individuos para \$1 extra.)

Esta dependencia de las funciones lineales de bienestar social –y de la teoría utilitarista en general– de las comparaciones interpersonales de utilidad ha dado lugar a muchos malentendidos. Gran parte de estos malentendidos podrían haberse evitado si se hubiera prestado más atención al parecido entre el papel que las probabilidades subjetivas juegan en la teoría de decisión bayesiana y el papel que las comparaciones interpersonales de utilidad (las ratios de conversión de la utilidad interpersonal) juegan en la teoría moral utilitarista.

Más específicamente, dado que en situaciones inciertas la utilidad esperada se define en términos de probabilidades subjetivas, la teoría de la decisión bayesiana nos exige asignar probabilidades subjetivas a contingencias alternativas. Obviamente, este requerimiento no se basa en la suposición de que los que toman decisiones sean necesariamente muy hábiles en la valoración de estas probabilidades. De hecho, en situaciones en las que se tiene poca información para guiar los juicios de probabilidad, estos juicios serán con seguridad de una calidad bastante pobre. Sin embargo, la teoría bayesiana sugiere que deberíamos tomar nuestras decisiones a partir de tales valoraciones de probabilidad, incluso en situaciones de información insuficiente, porque:

(1) Si nuestro comportamiento es consistente con los postulados de la teoría de la racionalidad bayesiana, entonces simplemente *no podemos evitar* hacer tales valoraciones de probabilidad, al menos implícitamente. Por ejemplo, supongamos que tengo que elegir entre dos apuestas. La apuesta *A* me haría ganar \$100 si el candidato *X* gana una cierta elección y me haría perder \$100 si él pierde; mientras que la apuesta *B* invertiría estas dos contingencias. Entonces, al elegir la apuesta, yo implícitamente haré un juicio de probabilidad acerca de las oportunidades de que él gane o pierda la elección. Simplemente no puedo evitar hacer tal juicio.

(2) Dado que tenemos que hacer tales juicios de probabilidad al menos implícitamente, nos iría mejor si los hiciéramos explícitamente: esto nos permitiría evitar inconsistencias perjudiciales en nuestros juicios de probabilidad y también nos permitiría hacer el mejor uso posible de la información de la que realmente dispongamos —por mucha o poca que sea esta información en cualquier caso dado.

De la misma manera, la teoría utilitarista requiere que usemos una función de bienestar social basada en comparaciones interpersonales de utilidad. Sin embargo, este requerimiento no presupone de ninguna manera que los que tomen las decisiones sean particularmente hábiles en hacer tales comparaciones. Presumiblemente, cuando conocemos a dos individuos (o dos grupos de individuos) razonablemente bien y tenemos un buen conocimiento de las situaciones en las que se hallan, entonces podemos comparar las utilidades que ellos obtendrían de conjuntos de bienes dados (o incluso de beneficios menos tangibles), con una precisión tolerable. (Por ejemplo, creo que puedo afirmar con cierta seguridad qué amigo mío obtendría más utilidad de una opera de Mozart, o de una buena cena en un restaurante francés, etc.) Por el contrario, las comparaciones de utilidad entre dos individuos con ambientes sociales y culturales poco conocidos para nosotros estarán sujetas a amplios márgenes de error. Pero, en cualquier caso, la teoría utilitarista sugiere que deberíamos hacer nuestras elecciones morales a partir de tales comparaciones de utilidad interpersonal, porque:

(1) Si seguimos los axiomas de la teoría utilitarista (y, cada día, todos nosotros nos comportamos como utilitaristas, al menos algunas veces), entonces simplemente *no podemos evitar* hacer comparaciones de utilidad interpersonal, al menos implícitamente. Por ejemplo, cuando tenemos que decidir si queremos regalarle algo a *A* o a *B*, entonces una de las consideraciones importantes (aunque quizás no la única consideración importante) será cuál de los dos, si *A* o *B*, puede conseguir una utilidad mayor de ello.

(2) Ya que tenemos que hacer, de cualquier forma, tales comparaciones de utilidad interpersonal, estamos bajo la obligación moral de hacerlas con el máximo cuidado posible y con el mejor uso de toda la información relevante disponible —al menos cuando estas comparaciones sirvan como base para una decisión moral importante. Cualquier intento de evitar tales comparaciones interpersonales, debido a algún prejuicio moral en contra de ellas, sólo puede conducir a decisiones morales irresponsables e irreflexivas.

Como he argumentado en otra parte (Harsanyi, 1955, pp. 316-321; y 1975a, sección 8), no hay nada misterioso en nuestra innegable habilidad para hacer tales comparaciones con diferentes grados de precisión dependiendo de la situación. Cualquier comparación tal es lógicamente equivalente a la predicción de cuáles serían nuestro comportamiento y nuestras reacciones emocionales en ciertas situaciones hipotéticas (posiblemente en situaciones muy diferentes de cualesquiera que hayamos experimentado hasta ahora). Intentar valorar las utilidades que otro individuo de una personalidad y ambiente social muy diferentes obtendría de diferentes conjuntos de bienes equivale a lo mismo que valorar las utilidades que yo obtendría de varios conjuntos de bienes si mis ingresos, posición social, situación personal, o actitudes emocionales fueran exactamente como las tuyas. Por lo tanto, equivale a lo mismo que valorar las utilidades que yo obtendría de varias conjuntos de bienes, si mis características y mi situación personales sufrieran un cambio específico.

Nadie pretende que tales valoraciones sean siempre fiables. Todo lo que sostenemos es que en muchos casos estamos bajo la obligación moral de hacer tales valoraciones y, de hecho, de hacerlas de forma tan cuidadosa y concienzuda como podamos —así como, en muchos casos, estamos bajo una obligación prudencial (pragmática) de hacer juicios de probabilidad y de hacerlos tan cuidadosa y concienzudamente como nos sea posible.

5. FUNCIONES DE BIENESTAR SOCIAL SUPER-IGUALITARIAS

Debido a la gran fuerza de los postulados de racionalidad bayesiana, me parece que es una tarea bastante desesperada construir una teoría moral a partir del rechazo de estos postulados para las elecciones individuales o las sociales (es decir, a partir del rechazo de mis postulados *a* o *b*). Tal teoría sólo podría dar lugar a implicaciones altamente contrarias a la intuición. Más bien, cualquiera que quisiera construir una teoría *super-igualitaria* —es decir, una teoría más igualitaria de lo que es la teoría utilitarista— tendría que negar o, al menos debilitar sustancialmente, mi axioma del individualismo (postulado *c*).

Un posible enfoque podría ser éste. Supongamos que el individuo *j* quiere construir una función de bienestar social para la sociedad *X*. Supondré aquí que el propio *j* es un observador externo y no un miembro de la sociedad *X*. (El modelo puede extenderse fácilmente al caso en el que *j* es realmente un miembro.) Enton-

ces, j podría empezar especificando sus preferencias políticas acerca de la distribución de utilidades e ingresos que a él le gustaría ver en la sociedad X . Para una referencia más clara, las describiré como *preferencias igualitarias de j* . (Estas preferencias pueden depender de la variedad de utilidades individuales y/o de los ingresos individuales de la sociedad X , o pueden depender de alguna complicada función de las utilidades individuales y/o los ingresos.) En vista del argumento anterior, será deseable suponer que las preferencias igualitarias de j satisfacen los axiomas de racionalidad bayesianos (por ejemplo, los postulados de Marschak (1950)). Entonces, j puede reemplazar mi postulado c por un postulado menos individualista c^* : "Si dos alternativas son indiferentes desde el punto de vista de cada individuo i en la sociedad y son también indiferentes desde el punto de vista de mis preferencias igualitarias (las de j), entonces son también indiferentes desde un punto de vista social".

Mediante una leve modificación de la prueba aportada en mi artículo (Harsanyi, 1955, pp. 312-314), es fácil verificar que estos axiomas implican el siguiente teorema:

TEOREMA. La función de bienestar social W de un individuo j debe tener la siguiente forma matemática:

$$W = t \sum_{i=1}^n a_i U_i + (1 - t) V_j, \text{ con } 0 \leq t \leq 1,$$

donde $U_1, \dots, U_i, \dots, U_n$ son las funciones de utilidad de los distintos miembros individuales de la sociedad X ; V_j es una función de utilidad de von Neumann-Morgenstern definida en términos de las preferencias igualitarias de j ; mientras que t y a_1, \dots, a_n son constantes.

Si j así lo desea, entonces puede elegir $t=0$, y puede escoger hacer V_j bastante independiente de las funciones de utilidad U_1, \dots, U_n : en este caso, por supuesto, W perderá toda conexión con las preferencias individuales. Pero, incluso si no va tan lejos, la forma matemática de W indicará claramente que el bienestar de los miembros individuales de la sociedad X no es su consideración primordial y que pretende sacrificar el bienestar de ellos —al menos en algunos casos— a sus propias preferencias igualitarias, cuando aquél entra en conflicto con éstas.

Personalmente, hallaría tal teoría ética o de economía del bienestar bastante censurable desde un punto de vista *moral*. Para mí, las consideraciones humanitarias nunca deberían ocupar un

segundo lugar frente a otras consideraciones, incluyendo las igualitarias. De hecho, a un nivel más profundo de análisis filosófico, creo que tal teoría sería bastante *irracional*, porque el individuo *j* no puede tener ningún motivo racional para desear imponer sus propias preferencias políticas a los miembros de la sociedad *X* (sea o no *j* un miembro de esta sociedad). Pero la racionalidad es un concepto de muchas dimensiones: la teoría que he descrito sería consistente, al menos, con los requerimientos más patentes de la racionalidad, al conformarse a los postulados de la racionalidad bayesiana —incluso si violara algunos criterios de racionalidad a niveles más profundos.

6. LAS FUNCIONES DE BIENESTAR SOCIAL Y EL PROBLEMA DEL CONSENSO MORAL

He sostenido que hay razones bastante importantes para usar funciones de bienestar social sobre utilidades individuales, al menos para los que creen en una filosofía moral humanista e individualista. Entre las personas que aceptan este punto de vista, desaparecería una fuente de desacuerdo moral: estarán de acuerdo, al menos, en la fórmula matemática apropiada de la función de bienestar social.

Por supuesto, quedaría otra fuente de desacuerdo bastante obvia: tales personas podrían aún diferir en cómo comparar las utilidades de diferentes individuos. Sin embargo, no se debe exagerar la probable importancia práctica de este problema. A este respecto, los tratamientos convencionales de la economía del bienestar dan una impresión bastante engañosa. Una pequeña reflexión mostrará que muy pocas de las controversias políticas de la vida real surgen de desacuerdos sobre comparaciones interpersonales de utilidad.

Esto, por supuesto, no significa que un acuerdo sobre la fórmula matemática de la función de bienestar social (incluso si éste fuera acompañado de un acuerdo —completo en la medida de lo razonable— sobre las comparaciones interpersonales de utilidad en todos los casos importantes) nos acercaría más a un estado utópico de consenso moral —así como una aceptación total de la teoría bayesiana tampoco nos acercaría más a un consenso sobre problemas de política puramente pragmáticos (es decir, sobre aquéllos que no implicaran asuntos morales controvertidos importantes).

En mi opinión, las fuentes más importantes de desacuerdo moral son los desacuerdos sobre qué predicciones condicionales o

incondicionales —si determinadas o probabilísticas— hacer acerca de hechos empíricos futuros. Por ejemplo, no podemos esperar un consenso moral en nuestra sociedad mientras que discrepemos sobre cuáles serán los probables efectos de las políticas anti-inflacionistas convencionales en situaciones económicas algo inusuales, para las que no existen precedentes históricos reales; o sobre cuáles serán los efectos a largo plazo de la inflación continuada en nuestras instituciones sociales; o sobre cómo se verán afectados nuestros hijos a la larga por el énfasis cada vez mayor que se pone en nuestras escuelas en la ‘creatividad’ y el énfasis cada vez menor en la ‘excelencia intelectual’; o sobre si una mayor presión sobre ciertos regímenes despóticos agilizará o ralentizará su esperanza de democratización y liberación, etc. Sin duda, un mayor progreso en la psicología humana, en las ciencias sociales, y, en algunos casos, en las ciencias naturales, dará respuestas más claras a algunos de estos problemas que ahora se nos plantean. Pero este tipo de problemas sin resolver siempre nos acompañarán, en parte porque la solución misma de estos problemas nos llevará a problemas nuevos. Al resolver algunos de estos problemas se crea una situación histórica radicalmente nueva, e incluso si hemos aprendido a hacer predicciones en la situación anterior, éstas quizás no nos ayuden mucho en la situación nueva. La economía keynesiana nos ha dado la posibilidad de hacer predicciones mucho mejores sobre los efectos de varias políticas económicas en condiciones de desempleo masivo. Esto mismo nos ha dado la posibilidad de eliminar esas mismas condiciones y crear una situación económica completamente nueva de empleo alto sostenido en la que puede que las predicciones keynesianas dejen de funcionar.^{9, 10}

⁹ En análisis más profundos, los desacuerdos acerca de las comparaciones interpersonales de utilidad son también desacuerdos acerca de las predicciones en ciertas situaciones hipotéticas —como desacuerdos sobre cuáles serían las preferencias y las reacciones emocionales de ciertos individuos si sus ingresos, posiciones sociales, niveles de educación, actitudes culturales e incluso sus personalidades, fueran diferentes de cómo son en realidad.

¹⁰ Los filósofos contemporáneos dividen los enunciados declaratorios en dos clases principales, esto es, enunciados lógico-matemáticos y enunciados empíricos. Por supuesto, esto priva a los enunciados filosóficos (conceptuales) de un lugar apropiado, ya que en realidad no pueden ser incluidos en ninguna de esas categorías (véase sección 1 anterior). Además, esta clasificación tampoco presta la suficiente atención a los problemas especiales relacionados con los enunciados acerca de hechos empíricos del futuro que obviamente no pueden ser verificados de la misma manera que los enunciados acerca de hechos empíricos del presente y del pasado, pero que son ingredientes esenciales de algunos de nuestros juicios morales de valor más importantes.

Sin embargo, incluso si un acuerdo sobre la fórmula matemática de nuestra función de bienestar social no eliminara nuestros desacuerdos morales más importantes —mientras seguimos discrepando en nuestras predicciones sobre los hechos empíricos futuros—, tal acuerdo es un objetivo por el que bien vale la pena luchar. Lo que una buena teoría formal del bienestar social puede hacer por nosotros en el campo de las decisiones políticas, económicas y morales es lo mismo que la teoría de la decisión bayesiana puede hacer por nosotros en el campo de las decisiones puramente pragmáticas. Puede ayudarnos a organizar nuestro análisis de la situación, a clarificar lo que sabemos y lo que no sabemos y cuáles son las suposiciones implícitas que estamos haciendo; y, lo más importante de todo, nos puede ayudar a reunir una gran cantidad de información —a menudo bastante heterogénea— en un proceso coherente y sistemático de toma de decisiones.

De hecho, incluso si no consideramos sus posibles usos prácticos, parece ser una proposición filosóficamente bastante interesante, en caso de que fuera cierta, el que nuestros criterios básicos de racionalidad, junto con una filosofía moral humanista e individualista, no nos dejan otra opción que definir nuestra función de bienestar social como una combinación lineal de utilidades individuales y, de hecho, como su media aritmética.

BIBLIOGRAFÍA

- CARNAP, Rudolf, *Logical Foundations of Probability*, University of Chicago Press, Chicago, 1950.
- DIAMOND, Peter, "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility: A Comment", *Journal of Political Economy*, 75 (1967), 765-766.
- FRIEDMAN, Milton and SAVAGE, Leonard J., "The utility Analysis of Choices Involving Risks", *Journal of Political Economy*, 56 (1948), 279-304.
- HARSANYI, John C., "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking", *Journal of Political Economy*, 61 (1953), 434-435.
- HARSANYI, John C., "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility", *Journal of Political Economy*, 63 (1955), 309-321.
- HARSANYI, John C., "A General Theory of Rational Behavior in Game Situations", *Econometrica*, 34 (1966), 613-634.

- HARSANYI, John C., "Can the Maximin Principle Serve as a Basis for Morality: A Critique of John Rawls' Theory", *American Political Science Review*, 69 (1975a), 594-606 [trd. cast. "¿Puede el principio maximin servir como base ética? Una crítica a la teoría de J. Rawls", *Hacienda Pública Española*, 48 (1977), 305-321].
- HARSANYI, John C., "The Tracing Procedure: A Bayesian Approach to Defining a Solution for n-Person Noncooperative Games", *International Journal of Game Theory*, 4 (1975b), 61-94.
- LUCE, R. Duncan and RAIFFA, Howard, *Games and Decisions*, John Wiley, New York, 1957.
- MARSCHACK, Jacob, "Rational behavior, Uncertain Prospects and Measurable Utility", *Econometrica*, 18 (1950), 111-141.
- RADNER, Roy and MARSCHACK, Jacob, "Notes on Some Proposed Decision Criteria", en Robert M. Thrall et al. (eds.), *Decisions Processes*, John Wiley, New York, 1954.
- RAWLS, John, "Justice as Fairness", *Philosophical Review*, 67 (1958), 164-194 [trd. cast. "Justicia como equidad", en: M. A. Rodilla (comp.), *J. Rawls. Justicia como equidad*, Madrid: Tecnos, 1986, c. 2].
- RAWLS, John, *A Theory of Justice*, Harvard University Press, Cambridge, Mass., 1971 [trd. cast. *Teoría de la Justicia*, México: F.C.E., 1978].
- ROBBINS, Lionell, "Interpersonal Comparisons of Utility", *Economic Journal*, 48 (1938), 635-641.
- SEN, Amartya K., *Collective Choice and Social Welfare*, Holden-Day, San Francisco, 1970 [trd. cast. *Elección colectiva y bienestar social*, Madrid: Alianza, 1976].
- SEN, Amartya K., *On Economic Inequality*, Clarendon Press, Oxford, 1973 [trd. cast. *Sobre la desigualdad económica*, Barcelona: Crítica, 1979].
- THEIL, Henri, *Optimal Decision Rules for Government and Industry*, North-Holland and Rand McNally, Amsterdam and Chicago, 1968.