

EL ANALISIS DE INCENTIVOS EN ECONOMIA NORMATIVA

Salvador Barberá Sáñez

Departamento de Teoría Económica
e Instituto de Economía Pública
Universidad del País Vasco

SEPARATA DE LA *REVISTA ESPAÑOLA DE ECONOMIA*,
II EPOCA, VOLUMEN I, NUMERO 1, 1984

EL ANALISIS DE INCENTIVOS EN ECONOMIA NORMATIVA

Salvador Barberá Sández (*)
Departamento de Teoría Económica
e Instituto de Economía Pública
Universidad del País Vasco

Resumen Este ensayo interpreta la evolución reciente de la Economía Normativa. Los problemas de incentivos han pasado a considerarse centrales para la evaluación y el diseño de mecanismos de asignación de recursos. Sólo aquellas instituciones que generen incentivos adecuados pueden verse como soluciones viables a los fallos de mercado. El análisis de tales cuestiones utilizando el instrumental de la teoría de juegos exige una reelaboración de toda la economía del bienestar.

Abstract This is an essay on recent changes in normative economics. Incentive issues have become a fundamental concern in the evaluation and the design of resource allocation mechanisms. Only those institutions with good incentive properties constitute acceptable solutions to market failures. The analysis of such properties with the tools of game theory requires a rewriting of welfare economics.

1. INTRODUCCION

La economía del bienestar ha experimentado cambios muy profundos en los últimos diez años, y seguirá cambiando en el futuro próximo, porque las amplias perspectivas que se le han abierto revelan aún mucho terreno por conquistar. Motor importante de estos cambios es la introducción sistemática de consideraciones estratégicas, de incentivos, como parte integrante de la problemática de la economía normativa. Tales consideraciones no sólo amplían el marco de la economía del bienestar, sino que permiten y estimulan un cambio sustancial de énfasis y de perspectiva, incluso en relación a otras cuestiones más tradicionales.

Buena parte de la motivación inicial para el estudio de la estructura de incentivos en mecanismos de asignación de recursos surge del estudio de economías con bienes públicos, en las que se hace probable la aparición de *free riders* bajo distintas formas institucionales. Nuestro propósito en estas páginas es describir las bases y direcciones generales de esta ampliación de la economía normativa, utilizando como hilo conductor, allí donde sea útil una interpretación concreta, la referencia a problemas de incentivos en economías con bienes públicos.

El trabajo gira en torno a unos pocos resultados —seleccionados algunos por su especial importancia, otros porque vienen directamente de la investigación del

(*) El autor agradece el apoyo financiero del Instituto de Estudios Fiscales y los detallados comentarios de un evaluador anónimo.

presente autor— y no pretende en modo alguno ser una revisión de literatura. Sin embargo, contiene referencias suficientes para que quien se interese por los sesgos y direcciones que aquí se destacan pueda encontrar puntos de arranque para lecturas posteriores.

2. NUEVOS ENFASIS Y PERSPECTIVAS DE LA ECONOMÍA NORMATIVA

Para colocarnos en perspectiva, empecemos por describir en términos amplios un enfoque tradicional en economía normativa. Punto de referencia obligada son los llamados Teoremas Fundamentales de la Economía del Bienestar (1). Dichos teoremas, en su formulación rigurosa, establecen condiciones bajo las cuales todo equilibrio walrasiano de una economía será, de existir, eficiente en el sentido de Pareto; y otras bajo las que cualquier asignación eficiente en una economía constituirá, junto con precios adecuados, un equilibrio walrasiano para determinadas especificaciones de los parámetros distributivos (dotaciones iniciales de recursos y participaciones en beneficios).

Uno de los papeles de estos teoremas ha sido el de dar expresión rigurosa a algunas intuiciones que, desde A. Smith, han compartido (o en otros casos combatido) los economistas acerca de las propiedades ventajosas del «mercado» y de la «competencia» como instituciones que promueven la eficacia y permiten aumentar el bienestar. Más específicamente, los teoremas ponen de relieve las ventajas de que todos los intercambios, directos o indirectos, entre dos bienes se lleven a cabo a una misma tasa o precio relativo; y las de que todos los agentes tomen sus decisiones considerando tales tasas de intercambio comunes como parámetros, sin considerar el valor estratégico de sus acciones, sin intentar influir sobre los precios a través de ellas (2).

El segundo papel de los Teoremas Fundamentales ha sido el de señalar posibles direcciones en las que ampliar el debate en torno al alcance, tanto teórico como práctico, de sus conclusiones. En un formato tradicional, al análisis cuidadoso de su contenido le siguen dos tipos de ampliaciones. Una primera es el estudio de los llamados «fallos de mercado», es decir, de aquellos fenómenos reales cuyas contrapartidas teóricas (i) no pueden incorporarse al modelo para el que son válidos

(1) Formuladas por primera vez con todo rigor por K. Arrow (1951). Lo que queremos señalar aquí es que, de hecho, marcan una forma de pensar no sólo para investigadores, sino a cualquier nivel de docencia: basta con pensar en la difusión y permanencia de los trabajos expositivos de Bator (1957, 1958).

(2) Esta forma de apuntar a lo esencial de los Teoremas Fundamentales pretende separar desde ahora dos aspectos que, en este artículo, se argumenta son distintos: las consecuencias de ciertos comportamientos, por un lado, y su plausibilidad, por otro. La competencia podría ser causa de que los individuos fueran precio-aceptantes, y de que todas las transacciones de un mismo bien se llevaran a cabo al mismo precio. Pero probarlo rigurosamente exigiría trabajo adicional, y en todo caso igual comportamiento podría en principio predicarse en una economía totalmente centralizada a la Lange.

aquellos teoremas, o en cuya presencia no resultan plausibles (ii) los supuestos de comportamiento que subyacen a la noción de equilibrio walrasiano, o (iii) las hipótesis técnicas necesarias para demostrarlos (3).

A la detección de posibles fallos de mercado le sigue el estudio de las posibles «soluciones» ante aquellos fallos, proponiendo extensiones o modificaciones de la noción de equilibrio walrasiano, estudiando la eficiencia de las asignaciones que satisfacen las nuevas condiciones de equilibrio, y discutiendo las consecuencias institucionales que encierra cada una de estas modificaciones, así como su plausibilidad. Ejemplo de este tipo de ampliación, en el caso del estudio de economías con bienes públicos, lo constituye el desarrollo de nociones como la de equilibrio de Lindahl, o la de equilibrio político-económico, o la de equilibrio con suscripción, el estudio de la eficiencia (o falta de ella) de las asignaciones asociadas con cada uno, y el análisis de las posibilidades y de las dificultades específicas a cada una de las instituciones que, implícitamente, presupone cada equilibrio: precios personalizados, asignación mediante voto político o cooperación espontánea, respectivamente (4).

Esta secuencia —teoremas fundamentales, fallos de mercado, soluciones (de mercado o no, según los casos) a dichos fallos— ha sido el esquema seguido por la economía normativa durante muchos años, y aún hoy sigue siendo una forma importante de enfocar los difíciles problemas, teóricos y prácticos, del diseño de instituciones. Sin embargo, lo que queremos argumentar en lo que sigue, a nivel general primero, y con ejemplos más concretos después, es que el enfoque tradicional nos lleva imperceptiblemente a poner énfasis indebidos en ciertas distinciones, mientras que no agudiza suficientemente otras que son mucho más importantes, tanto para el debate teórico como para el debate más general sobre el papel relativo de los mercados y del Estado en la economía.

En primer lugar, la secuencia tradicional pone excesivo énfasis en un concepto teórico de equilibrio, el walrasiano, y en una interpretación concreta de su significado, como formalización de la competencia en mercados impersonales. Gracias a trabajos que arrancan desde los años 50, y que han estado permanentemente ligados a figuras como K. Arrow y L. Hurwicz (5), se va imponiendo una formulación más amplia de la economía normativa como estudio de las posibilidades de

(3) Ejemplos de i) serían, por ejemplo, la presencia de bienes públicos, para los cuales no es válida la especificación habitual de que la suma de lo consumido por cada individuo debe igualar a la cantidad total disponible del bien; o la ausencia de mercados para algún bien relevante, en cuyo caso no podríamos hablar del precio de éste, algo que el modelo da por supuesto. Ejemplos de ii) son todas aquellas situaciones en las que, ya sea por barreras de entrada u otras razones, existen posibilidades de obtener ventajas influenciando estratégicamente los precios a través de las cantidades. Ejemplos de iii) son aquellos fenómenos reales, indivisibilidades, efectos externos, rendimientos crecientes [asociados en parte a fenómenos del tipo i)] cuya traducción en modelos teóricos hace implausible las hipótesis de convexidad.

(4) No hay posibilidad de cubrir bibliográficamente un campo tan amplio, ni hace falta aquí. Mencionemos solamente, para el caso con bienes públicos, dos artículos de revisión, fuertemente complementarios entre sí: los de Milleron (1972) y Tulkens (1978). Para una introducción en castellano, véase Barberá (1979). En todos estos trabajos, y aun con algunas referencias a cuestiones de incentivos, el énfasis sigue siendo esencialmente mecanicista.

(5) Trabajos fundamentales como Arrow y Hurwicz (1960), o los referenciados en Hurwicz (1973).

diseño y comparación de procedimientos para la asignación de recursos, en cuyo seno aquel concepto teórico y aquella interpretación aparecen en perspectiva como casos particulares, aunque muy importantes por su valor paradigmático.

Además de esta ampliación de perspectivas, que ya viene de lejos, pero cuya extensión ha sido gradual, la activación del estudio de cuestiones de incentivos desde un punto de vista de teoría de juegos, a lo largo de los 70, permite un cambio de énfasis cuyas consecuencias nos parecen aún no del todo desarrolladas. Si volvemos a considerar las listas de «fallos de mercado», podemos apreciar que entre éstos se encuentran objetos de índole diversa. Una cosa es decir que si todos los agentes pagan lo mismo por un bien público que valoran marginalmente de modo distinto no habrá eficiencia —y que, por tanto, en presencia de bienes públicos los precios de eficiencia deben ser personalizados—, y otra es discutir la mayor o menor plausibilidad de que, en un contexto donde tales precios sean en principio posibles, los agentes actúen estratégicamente como *free riders*. La primera es una observación sobre las propiedades de un mecanismo —donde el término mecanismo se utiliza para significar que los individuos se tratan como meros transmisores de información, sin iniciativa estatéctica— y la segunda se refiere a las razones por las que resulta más o menos pertinente la referencia a dicho mecanismo como formalización de instituciones cuya implantación vaya a recomendarse —es una observación sobre la estructura del juego que se abre ante los agentes económicos cuando éstos se ven constreñidos a actuar en el seno de un mecanismo, pero operan estratégicamente dentro de sus posibilidades—. Ambas observaciones son importantes, y las dos han sido siempre objeto de atención. Sin embargo, mientras que la primera ha recibido tratamiento formal desde hace tiempo, la segunda habría quedado relegada al contexto interpretativo hasta entrados los 70, cuando en realidad resulta posible darlas a ambas un tratamiento teórico de igual nivel, aunque distinto (6), (7).

Lo que hemos querido enfatizar con este ejemplo es que la irrupción de la teoría de juegos en muchos terrenos de la economía normativa permite reclamar para el debate teórico gran parte de los temas que hasta hace pocos años hubiesen sido sólo objeto de especulación y materia de opinión. Y esto no sólo amplía el campo de trabajo teórico, sino que hace también recomendable e induce una reestructuración en los énfasis de la economía normativa. No parece demasiado fructífero pensar en un mundo «clásico», restringido, donde funciona bien la competencia, y un mundo con «fallos de mercado» donde ésta deja de funcionar y acaso ciertas correcciones restituyan la eficiencia, ni tampoco un debate interpreta-

(6) Véase el propio trabajo de Samuelson (1954), y el distinto grado de detalle y precisión con que aborda una y otra problemática.

(7) Al empuje de los setenta hacia un tratamiento formal de temas de incentivos contribuyen además de la revitalización de la teoría de juegos, novedades en por lo menos tres corrientes próximas pero distintas de la economía: teoría de bienes públicos, con los trabajos pioneros de Clarke (1971) Groves (1973), Groves y Loeb (1975), Groves y Ledyard (1977) y Green y Laffont (1979); teoría de elección social —con aportaciones iniciales de Gibbard (1973) y Satterthwaite (1975), y teoría de mecanismos de asignación de recursos, en cuyo seno los trabajos de Hurwicz van sesgándose hacia esta temática a partir de 1972. Para una visión del desarrollo de la teoría de juegos en las direcciones más relevantes para lo aquí tratado, véase Moulin (1981).

tivo sobre la mayor o menor cercanía de uno u otro al mundo real, como base para tomas de posición sobre los papeles respectivos del mercado y del estado, como se sigue casi inmediatamente del esquema tradicional descrito anteriormente. Más bien existen dos problemas distintos, ambos importantes y ambos susceptibles de tratamiento teórico, y que merecen atención tanto en el contexto de los teoremas fundamentales como en el de economías más complejas: el de las propiedades de cada mecanismo desde una perspectiva no estratégica, y el de las estructuras de juego y de los problemas de incentivos que surgen de imponer cada mecanismo sobre agentes racionales. La contrapartida interpretativa del primero sería: ¿qué consecuencias son de esperar del funcionamiento de un mecanismo, si éste se implanta y funciona?; la del segundo sería: ¿qué posibilidades de éxito resulta plausible atribuirle a un sistema, dadas las posibilidades estratégicas que deja abiertas a los agentes?

Con estos énfasis, podemos ver cómo, de hecho, muchas «soluciones» a los «fallos de mercado» no son sino versiones más amplias o modificaciones de los teoremas fundamentales, aplicables a otros contextos —con precios personalizados, con suficientes mercados, con impuestos, etc.—. Su *status*, en relación a cuestiones del primer tipo, es decir, como estudio de propiedades de mecanismos, es análogo al de los teoremas fundamentales. Viceversa, y del mismo modo que a nivel interpretativo al menos se ha cuestionado siempre la viabilidad de las soluciones propuestas ante aquellos fallos, también conviene cuestionar la de la solución walrasiana en el contexto donde son válidos los teoremas fundamentales, y tanto en uno como en otro caso someter esta segunda cuestión a un análisis teórico, cuando sea posible (8).

En lo que sigue, esta exposición se hace más detallada y parcial, aunque el propósito general seguirá siendo la insistencia en que el nivel de análisis estratégico conduce a cambios de óptica y de énfasis importantes. Para ello, empezaremos por describir dos resultados que, en sus respectivos contextos, pueden verse como «teoremas fundamentales» generalizados y demuestran que el debate normativo esencial no puede situarse al nivel de la existencia de mecanismos «de mercado» que promuevan la eficiencia, sino al nivel de su plausibilidad, dadas las estructuras de incentivos inducidos por cada uno en cada contexto (sección 3).

Seguiremos con una descripción de dos marcos alternativos, bajo uno de los cuales existen mecanismos con buenas propiedades estratégicas, mientras que en el otro, más general, no son posibles, y pondremos énfasis sobre el papel que en dicho problema de diseño juega la capacidad de cada individuo para alterar la asignación social de recursos (sección 4).

(8) Diversos tipos de trabajos han cuestionado la viabilidad de la solución walrasiana poniendo de relieve su manipulabilidad, ya sea por medio de ocultación de recursos [Mas Colell (1976), Postlewaite (1979)], ya sea por ocultación de las verdaderas preferencias [Hurwicz (1972)], incluso si las economías cuentan con muchos agentes [Roberts y Postlewaite (1976)]. Un enfoque alternativo parte de postular comportamientos no competitivos, y estudia condiciones bajo las que, en condiciones límite, la hipótesis competitiva es una buena aproximación de la conducta estratégica de equilibrio; ejemplos en esta dirección se encuentran en el número monográfico del *Journal of Economic Theory* (1980).

Finalmente, trataremos de ofrecer un marco general en el que se sugieren muchos caminos para la plena incorporación de los incentivos en el tratamiento sistemático de cuestiones normativas, y a través del cual queda clara la amplitud de los temas y tratamientos que quedan aún por desarrollar (sección 5).

3. PRECIOS Y EFICIENCIA EN CONTEXTOS CON BIENES PUBLICOS Y OTROS «FALLOS DE MERCADO»

En este apartado queremos destacar que, desde el punto de vista formal, las asignaciones de equilibrio en que, bajo restricciones presupuestarias adecuadamente definidas, cada consumidor maximiza su utilidad y cada productor sus beneficios, son eficientes. Lo que varía en presencia de bienes públicos y, más en general, de otras complicaciones, es la naturaleza de los mercados y de los sistemas de precios a través de los que se coordinan las acciones individuales. De ahí que, una vez hayamos ilustrado en este apartado la posibilidad de teoremas muy generales sobre las relaciones entre precios y eficiencia, pasemos a dedicar el resto del trabajo a la cuestión, menos definitivamente explorada, del alcance de tales resultados a efectos normativos.

Un primer ejemplo en esta dirección nos lo proporciona el trabajo de Andreu Mas Colell (9) sobre eficiencia y descentralización en economías con bienes públicos puros. En un contexto limitado, pero muy relevante porque se ajusta bien al caso de decisiones entre unos pocos proyectos públicos alternativos cuyo volumen en relación con el conjunto de la actividad económica no es excesivo, demuestra que toda asignación es eficiente si y sólo si puede obtenerse como un equilibrio en valores, donde dicho concepto de equilibrio admite la posibilidad de precios personalizados e impuestos y subsidios de capitación.

Un segundo ejemplo viene dado por algunos resultados aún sin publicar, obtenidos por el presente autor en colaboración con H. Sonnenschein, y que intentan dar expresión precisa a una idea muy generalizada: que «si hubiese mercados para todas las interacciones relevantes» las asignaciones de equilibrio con precios personalizados serían eficientes (10).

Tanto este último resultado como el anterior permiten, efectivamente, establecer una correspondencia entre equilibrios y asignaciones eficientes en presencia de bienes públicos, demostrando así el poder de los precios como orientadores de decisiones colectivamente eficientes tomadas descentralizadamente por individuos maximizadores. Lo crucial, sin embargo, como ya se desprende de los trabajos iniciales de Samuelson, no es si ciertos equilibrios son eficientes, sino el punto hasta el que resulta plausible que surjan las instituciones asociadas implícitamente

(9) Este trabajo de Mas Colell (1980) contiene mucho más que lo que aquí se cita, ya que procede también, y fundamentalmente, a discutir cuestiones estratégicas.

(10) Este resultado se demuestra bajo la hipótesis de rendimientos constantes a escala, expresión a su vez de otra creencia ampliamente difundida: la de que ésta sería una característica común a todos los procesos productivos, con tal que, en su descripción, «se tuviesen en cuenta todos los factores de producción relevantes». Véase McKenzie (1981).

con dichos equilibrios, o que los individuos se comporten dentro de ellas aceptando los parámetros impuestos sobre ellos (precios, restricciones directas, ...), sin intentar sacar ventajas estratégicas que les queden abiertas.

Empezaremos por describir el resultado de A. Mas Colell en el que se relacionan eficiencia y descentralización en economías con bienes públicos puros (11).

Existe un número finito de *proyectos*, K , y un número también finito de agentes $I = \{1, \dots, n\}$. Los estados económicos vendrán descritos mediante la especificación de un proyecto en K y n números no negativos, en la forma $(x, m) \equiv (x, m_1, m_2, \dots, m_n) \in K \times [0, \infty)^n$; se entiende que en el estado (x, m) cada individuo i además de disfrutar de los servicios que proporciona el proyecto x , puede disponer de la cantidad m_i de un bien privado, cuya más fácil interpretación es como «dinero».

Cada individuo viene descrito por dotaciones iniciales del bien privado, en cantidad w_i , y por preferencias \geq_i sobre $K \times [0, \infty]$ que satisfacen las siguientes propiedades para todo i :

- A.1. \geq_i es un preorden continuo, reflexivo, completo y transitivo.
- A.2. a) \geq_i es continuo y estrictamente monótono en la dimensión del dinero, es decir, que para todo $x \in K$, si $m' > m$ esto implica que $(x, m') >_i (x, m)$.
- b) Si $m > 0$, $(x', m) >_i (x, 0)$ para todo $x, x' \in K$.

Finalmente, se postula una función de coste $c: K \rightarrow (-\infty, \infty)$ que se supone continua.

En este contexto, un estado económico (x, m) es *factible* si

$$c(x) \leq \sum_{i \in I} w_i - \sum_{i \in I} m_i,$$

es decir, si el coste del proyecto x más las cantidades de dinero en el estado no exceden las disponibilidades iniciales; y es *eficiente* en el sentido de Pareto si es factible y no existe ningún otro estado también factible (x', m') tal que $(x', m'_i) \geq_i (x, m_i)$ para todo $i \in I$, con preferencia estricta para algún $j \in I$.

El concepto de equilibrio propuesto por Mas Colell para este modelo es el de un equilibrio en valores.

Un estado (\bar{x}, \bar{m}) es un equilibrio en valores si existen:

- a) un vector $v = (\bar{v}_1, \dots, \bar{v}_n)$ de funciones

(11) La presentación que sigue, así como la de algunos otros resultados a lo largo del texto, puede no ser lo suficientemente detallada para el lector poco inclinado a la formalización, ni lo bastante rigurosa para el habituado. Con todo, nos arriesgamos a introducir algunos pasajes en este tono porque creemos firmemente que debe hacerse un verdadero esfuerzo de aproximación entre economía formalizada y economía «literaria», a base de escribir resultados formales e interpretar su intención y alcance.

$$v_i: K \rightarrow [-\infty, \infty) \quad (i=1, \dots, n)$$

a las que llamaremos *valoraciones*;

b) un vector $\pi = \pi_1, \dots, \pi_n \in \mathbb{R}^n$ de *transferencias netas* para el cual

$$\sum_{i \in I} \pi_i = \sum_{i \in I} v_i(\bar{x}) - c(\bar{x})$$

tales que:

1) para cada i , (\bar{x}, \bar{m}) maximiza \geq_i sobre el conjunto

$$\{(x, m_i) | v_i(x) + m_i = w_i + \pi_i\},$$

2) \bar{x} maximiza

$$\sum_{i \in I} v_i(x) - c(x)$$

sobre todos los $x \in K$.

Finalmente, se establece el siguiente

Teorema 1.—El estado (\bar{x}, \bar{m}) es un equilibrio en valores si y sólo si (\bar{x}, \bar{m}) es eficiente en el sentido de Pareto.

Interpretemos brevemente este resultado: cada valoración $v_i(\cdot)$ establece la cantidad $v_i(x)$ de la que el individuo i debe desprenderse si el proyecto x se pone en práctica, para cada x . Un equilibrio en valores se obtiene en una situación en la que, dadas sus dotaciones iniciales y las transferencias netas que ha recibido (o pagado, si π_i es negativo), cada individuo se encuentra en el mejor de los estados compatibles con la valoración que se le atribuye, y el excedente «social» —pagos de individuos menos coste del proyecto de equilibrio— se ve también maximizado. Una valoración es, pues, un «precio» personalizado en una versión general, ya que al ser en principio los proyectos un objeto abstracto, y no estar definido el concepto de «cantidad de proyecto», no tendría sentido tampoco hablar de precio «por unidad». En este aspecto, el modelo es muy general y demuestra que, esencialmente, todo equilibrio es eficiente y cualquier estado eficiente puede obtenerse como equilibrio (para transferencias adecuadas), análogamente a lo postulado por los teoremas fundamentales. Lo que varía es la interpretación del resultado: ¿hasta qué punto puede confiarse en que estos precios personalizados se formen espontáneamente?; ¿en qué medida aceptarán los individuos tales precios como parámetros? Hay razones para pensar que en ambos casos la respuesta sea, por lo general, poco favorable. Pero lo importante es destacar que estas cuestiones pueden formularse y tratarse formalmente, en un marco de teoría de juegos.

Por otra parte, observemos que, si bien general en ciertos aspectos (no hace falta suponer nada acerca de la naturaleza de los proyectos), el modelo contiene algunos supuestos restrictivos, que lo acercan a un tratamiento de equilibrio parcial: los derivados del tratamiento agregado de todos los bienes privados en forma de

un bien con las características de «dinero». Como veremos más adelante, en el marco de economías para las que este modelo sea adecuado, y gracias a la posibilidad de transferencias, es posible diseñar procedimientos de asignación con propiedades atractivas desde el punto de vista de los incentivos. Y aunque sus limitaciones sean evidentes, conviene señalar que, en muchos contextos, la hipótesis de un bien transferible universalmente deseado y en relación al cual las preferencias son «separables» posee un gran atractivo (12).

El segundo resultado que queremos utilizar como reflejo de la amplitud de posibilidades que encierran las extensiones a nivel formal de teoremas sobre las relaciones entre precios y equilibrios es parte de una investigación conjunta entre el autor y H. Sonnenschein. A diferencia de la anterior formulación, la que sigue supone que el espacio de bienes es el espacio euclídeo, lo cual la distancia de ciertas formulaciones propias de la teoría de la elección social y supone cierta pérdida de generalidad. Se gana en generalidad, sin embargo, en lo referente al número de bienes, al tipo de preferencias y a las clases de «fallos de mercado» considerados.

Como ya hemos mencionado, la extensión y profundidad de las dificultades para lograr asignaciones eficientes mediante precios y mercados en presencia de «fallos de mercado» es materia de debate y opinión. Un argumento habitual para minimizar tales dificultades atribuye una buena medida de ellas a los economistas, y a su modo de modelar la realidad económica, más que a la economía misma. Así, por ejemplo, la presencia de rendimientos a escala no constantes, puede verse como consecuencia de algún «factor de producción» cuyo papel en la producción no ha sido recogido en nuestra forma de modelar la economía. Si esto es así por olvido del modelo, corrija-se éste; si es porque no existen realmente mercados para dicho bien, la posibilidad y consecuencias de creación de dichos mercados podrá ser objeto de estudio, pero ello no es obstáculo para incorporar al bien en el modelo, y de ello debería seguirse la constancia de los rendimientos. De un modo similar habría que tratar, según esta concepción, los efectos externos tecnológicos: una descomposición adecuada de cada bien —a efectos de darle precio— debería distinguir su capacidad de dar servicios a quien lo utiliza directamente así como la de generarlos para otros agentes. En resumen: según una visión bastante generalizada, pero pocas veces concretada con precisión, los fallos de mercado se derivan exclusivamente de una falta de instrumentos —mercados y precios— para tener en cuenta toda la riqueza de bienes y servicios que fluyen entre los agentes económicos: los rendimientos crecientes y los efectos externos serían fenómenos poco sustantivos, no básicos, capaces de ser absorbidos al igual que toda dificultad asociada con bienes públicos, mediante estructuras de precios y de mercados suficientemente amplias.

Nuestro modelo proporciona una formulación precisa de estas ideas, y confirma su amplitud de alcance a un nivel formal (13). Pero, a la vez, debe servir para

(12) Esencialmente, se trata de las hipótesis bajo las cuales es razonable un análisis de equilibrio parcial.

(13) El procedimiento habitual es el inverso al seguido aquí: se suele empezar demostrando los teoremas fundamentales para el caso particular de economías «clásicas», y se argumenta después que,

poner de relieve la improcedencia de cualquier salto lógico que, desde una formulación mecanicista, pretenda derivar conclusiones normativas sin pasar por un análisis de incentivos que valore las posibilidades de que surjan y funcionen adecuadamente los precios y los mercados previstos formalmente —y dados por supuesto— en el teorema.

Veamos ahora concretamente un modo de modelar lo que acabamos de exponer.

Suponemos m consumidores, designados por los índices i, j, k, \dots

l bienes de consumo, con índice h

n factores de producción, con índice q .

Un estado de consumo viene dado por un vector en R_+^{ml} (es decir, $m \cdot l$ números reales no negativos). Un estado de consumo vendrá indicado genéricamente por x, y, \dots , donde $x = (x_1, x_2, \dots, x_m)$ tiene por componentes x_i el consumo del i -ésimo individuo, y para cada individuo i , $x_i = (x_{i1}, \dots, x_{il})$ tiene por componentes x_{ih} la cantidad del bien h consumida por i en x . Un estado de factores viene dado por un vector n -dimensional $w = (w_1, w_2, \dots, w_n) \in R_+^n$, donde cada w_g representa las cantidades utilizadas del factor g .

Las posibilidades de transformación de la economía vienen dadas por un cono convexo T en R_+^{n+ml} con vértice en el origen. Genéricamente $(w, x) \in T$ indica que los consumos en x pueden conseguirse si se dispone de los factores w . Obsérvese que la hipótesis de que T es un cono convexo formaliza la idea de que todos los factores y productos relevantes han sido tenidos en cuenta en el modelo, y de que, en estas condiciones, resulta legítima la hipótesis de rendimientos a escala constantes.

Las preferencias de los individuos se consideran definidas sobre R_+^{ml} , se denotan por \geq_i , y se suponen preórdenes continuos completos, reflexivos y transitivos. Obsérvese que, al suponer que las preferencias de cada individuo se definen sobre sus propios consumos y el de todos los demás, admitimos todo tipo de efectos externos en el consumo.

Un sistema de precios para los bienes de consumo vendrá dado especificando, para cada individuo i , un $p^i \in R^{ml}$, cuyos elementos se denotan más explícitamente por $p^i = (p_1^i, \dots, p_m^i)$ y a su vez $p_j^i = (p_{j1}^i, \dots, p_{jl}^i)$. Interpretaremos que p_{jh}^i es lo que i paga al sector de producción (o recibe de éste, si el valor es negativo), por el hecho de que el individuo j consuma una unidad del bien h . Así, pues, admitimos la posibilidad de pagos (y cobros) a tipos personalizados, y no sólo como consecuencia de las propias acciones, sino también de las ajenas.

Los precios de factores se modelan, más tradicionalmente, como puntos en R^n (un precio para cada factor) que admitiremos puede ser negativo o positivo, para dejar abierta la posibilidad de factores perniciosos. Conjuntamente, un sistema de precios será un vector no nulo $(q, p^1, p^2, \dots, p^m)$ en $R^{n+(m \times m \times l)}$.

bajo reinterpretaciones adecuadas de lo que son bienes, mercados y precios, aquellos resultados pueden cubrir mundos más complejos. Nuestro enfoque tiende a colocar lo general por delante, y permite después ver el caso «clásico» y otros como casos particulares.

Una economía quedará completamente especificada cuando lo hayan sido

$$\{l, n, \bar{w}, T, m, (\geq_i)_{i=1}^m\},$$

es decir, el número de bienes de consumo y factores, las cantidades de estos últimos, las posibilidades de transformación, el número de individuos y sus preferencias.

Un estado de consumo x es eficiente en el sentido de Pareto para esta economía si

- 1) $(\bar{w}, x) \in T$, y
- 2) $\nexists (\bar{w}, x') \in T$ tal que $x' \geq_i x$ para todo i , con $x' \geq_j x$ para al menos un j .

Diremos que un sistema de precios (q, p) y un estado de consumo x constituyen un equilibrio con aceptación de precios (14) si

- 1) $\forall_i, x' > x \rightarrow p^i \cdot x' > p^i \cdot x$, y
- 2) $(\bar{w}, x) \in T$ y, $\forall (\bar{w}, x') \in T$
 $q \cdot \bar{w} + \sum p^i \cdot x \geq q \cdot \bar{w} + \sum p^i \cdot x'$

La condición 1) refleja la hipótesis de que, en equilibrio, cada individuo maximiza su utilidad bajo una restricción de balance, aunque aquí se expresa ésta implícitamente, exigiendo que cualquier asignación preferida a la de equilibrio sea más costosa que ésta. Recordemos que x no representa sólo el consumo del individuo i , sino también el de todos los demás, y que los precios en p^i se refieren también a los consumos de todos los individuos, por lo que $p^i x$ supone una valoración de lo que, dado el sistema de precios, el individuo i debe pagar por x . Asimismo, 2) es una condición global de maximización de beneficio, para un sector productivo que sólo hemos descrito de forma agregada, y que suponemos recibe íntegramente como ingresos los pagos de los consumidores, a la vez que paga por los factores (15).

Finalmente, nótese que x , el objeto que consumidores y productores eligen en esta formulación, no es el consumo de cada agente, sino, conjuntamente, el de todos ellos, por lo que un equilibrio puede verse aquí —y por extensión, en contextos más habituales que son casos particulares de éste— como una situación que, cada uno desde sus respectivas preferencias y restricciones, todos los agentes coinciden unánimemente en escoger.

Estamos ahora en condiciones de enunciar los siguientes teoremas:

Teorema 2.—Si (q, p, x) es un equilibrio con aceptación de precios para una

(14) Siguiendo a Debreu (1959) (cap. 6), podría también decirse que x es un equilibrio relativo al sistema de precios (q, p) .

(15) Obsérvese que, gracias a la hipótesis de rendimientos constantes, podremos despreocuparnos del modo de reparto de los beneficios, ya que éstos serán nulos en equilibrio.

economía y si \geq_i , cumple la hipótesis de no saciación local para todo i , el estado x es eficiente en el sentido de Pareto.

Teorema 3.—Si x es un estado eficiente en el sentido de Pareto, con todos los consumos estrictamente positivos, si \geq_i es convexo para cada i , y si $\overline{\geq_i(x)} = \geq_i(x)$ para al menos un i , existe un sistema de precios (q, p) tal que (q, p, x) es un equilibrio con aceptación de precios (16).

Ya hemos insistido en una de las conclusiones de estos resultados: que, dados precios y mercados suficientes, es concebible lograr la eficiencia en ellos si se consigue que operen y que los agentes los acepten. Otra conclusión de esta formulación amplia es que permite colocar lo general por delante de lo particular: así, un conjunto de proposiciones que no detallaremos, pero que se desprenden de lo anterior nos permite ver cómo la eficiencia de los equilibrios de Lindahl en presencia de bienes públicos, o la de los equilibrios walrasianos en ausencia de éstos, se derivan de nuestros teoremas, reforzando así nuestro argumento de que lo crucial en estos tipos de análisis no es la posibilidad de generar modelos cuyos equilibrios sean eficientes, sino hallar modos de profundizar teóricamente en la cuestión de las posibilidades de derivar de ellos prescripciones normativas compatibles con los incentivos de los agentes.

4. MECANISMOS NO MANIPULABLES

Si, en muchos casos, un concepto de equilibrio adecuadamente general permite relacionar estados de equilibrio con estados eficientes, ¿cómo podemos verificar la plausibilidad de que las instituciones implícitas bajo aquel concepto de equilibrio vayan a ser operativas?

En una primera respuesta, podríamos decir que tal sería el caso si las propias instituciones generasen incentivos tales que lo mejor para quienes operan en su seno fuera siempre acatar las reglas. Una segunda posibilidad sería que, aun cuando pudiera haber incentivos para que los agentes intentaran manipular el funcionamiento del sistema, las consecuencias de dicha manipulación fueran inocuas. Una tercera, que aunque en su forma original el sistema fuera manipulable, pudiesen arbitrarse sistemas adicionales cuya implantación eliminara o hiciera inofensivas las posibles distorsiones derivadas del comportamiento estratégico de los agentes.

En esta sección nos ocuparemos del primer supuesto, es decir, de la posibilidad de diseñar procedimientos bajo los cuales los agentes contribuirían llanamente a la formación de las decisiones colectivas, porque no dejan espacio a ninguna manipulación. En la siguiente sección veremos hasta qué punto y dentro de qué marcos pueden tratarse la segunda y tercera posibilidad aquí apuntadas (17), (18).

(16) Esta es una hipótesis técnica. $\geq_i(x)$ es el conjunto de los estados al menos tan preferidos como x de acuerdo con \geq_i y $\overline{\geq_i(x)}$ el de los más preferidos que x . $\overline{\geq_i(x)}$ es el cierre del conjunto $\geq_i(x)$.

(17) Al final de la sección 4 se destaca y aclara en qué sentido el estudio de mecanismos no manipulables representa un enfoque limitado de los problemas de incentivos.

(18) Todo lo que sigue es una discusión de las propiedades teóricas de mecanismos, bajo la

Si en las secciones anteriores hemos hablado de entornos económicos más o menos complejos, y de conceptos de equilibrio que les son aplicables, el tratamiento que adoptaremos ahora será algo más abstracto y nos llevará a hablar de funciones (o correspondencias) de decisión social. Empecemos, pues, relacionando uno y otro lenguaje. Un concepto de equilibrio, a medida que se aplica sobre entornos económicos distintos, va generando conjuntos de «asignaciones de equilibrio» para cada uno de ellos. Es, pues, un posible procedimiento, aunque no el único posible, de establecer una correspondencia entre economías (entendidas como especificaciones de sus características definitorias), y asignaciones (entendidas como posibles estados finales de dichas economías). Un concepto de equilibrio define, además, al menos implícitamente, un conjunto de operaciones más o menos complejas a través de las cuales la interacción de los agentes llevaría a seleccionar, para cada economía, aquellas asignaciones que fueran de equilibrio. El conjunto de los instrumentos y reglas de comunicación e interacción entre agentes capaces de identificar asignaciones de equilibrio constituirían un mecanismo. Aun cuando en la descripción de un mecanismo juega una parte importante el proceso de intercambio de información, lo cierto es que muchos temas relevantes —y en particular el de incentivos que nos ocupa— pueden tratarse en buena medida como si sólo las situaciones de equilibrio contaran. Si, obviando otros temas interesantes como los relativos a la mecánica del intercambio informativo, nos concentramos en las consecuencias del mecanismo, podemos limitarnos a analizar la correspondencia que éste establece entre entornos y asignaciones. Además, y para simplificar la exposición, supondremos que se trata de una función, y la llamaremos función de decisión social. Se tratará, pues, de una función que a cada entorno le hace corresponder una asignación.

Salvando, pues, la cuestión de la unicidad del resultado (que marca la diferencia entre funciones y correspondencias), un concepto de equilibrio nos vendría a concretar un mecanismo, y las asignaciones de equilibrio correspondientes a cada entorno irían definiendo una función de decisión social. La relación no es, sin embargo, recíproca. Dada una función de decisión social, no tiene por qué responder a un mecanismo específico derivado de la aplicación de un concepto dado de equilibrio: puede, por ejemplo, corresponder a la aplicación de criterios éticos sin ninguna contrapartida en términos de comportamiento individual bajo restricciones, que es lo que subyace a cada noción de equilibrio. Así, pues, y para terminar con estas aclaraciones, cuando hablamos de funciones de decisión social creemos referirnos a objetos que, en una interpretación posible, están íntimamente ligados a las nociones habituales de equilibrio económico, pero que admiten también otras interpretaciones más normativas. Por ejemplo, la estructura de una función de decisión social se acerca mucho a la de un procedimiento de votación. En ambos, conocidas las características de los participantes, queda determinada la decisión a

hipótesis de que la actuación de los agentes se rige por el propio interés, definido en sentido amplio. Una corriente moderna de estudio, que lentamente se va extendiendo, confía a experimentos de laboratorio la respuesta a cuestiones aún poco esclarecidas como: ¿Hasta qué punto sacan partido los agentes de todas las ventajas que les ofrece el sistema económico? ¿Qué papel juega el altruismo o el control social sobre el comportamiento estratégico de los agentes? Algunos títulos en esta prometedora dirección de la «economía experimental» son Isaac et al (1982), Marwell y Ames (1980) y Smith (1980).

tomar: la diferencia, o mejor, lo que tiene de específico un mecanismo de votación, es que las características relevantes que definen al entorno son las preferencias de los individuos con respecto a las posibles alternativas entre las que decidir, mientras que en una función de decisión tal como la hemos definido, cabría que la descripción del entorno fuese más compleja. Otra característica común es que la información pertinente para reconocer el entorno en que nos movemos puede ser exclusiva de los agentes, y costosa o imposible de recabar sin que éstos cooperen. Esto es claro en el caso de las preferencias individuales, pero se aplica también a muchos otros parámetros: la capacidad productiva actual y futura de las empresas, la capacidad de esfuerzo de obreros y empresarios, etc. Diremos en este caso que los agentes disponen de información privada, y que para su funcionamiento es preciso que éstos revelen sus características. De hecho, y aun cuando nuestro argumento general quiere abarcar otros ámbitos, las formulaciones que siguen se refieren al caso en que las características relevantes son preferencias (19).

Una vez establecida —por acuerdo o imposición— una función de decisión concreta, los agentes sujetos a ella deben adoptar una actitud determinada respecto a su comportamiento. Pueden decidir revelar sus verdaderas características en cualquier circunstancia, en cuyo caso diremos que actúan llanamente. O, por el contrario, pueden optar por actuar estratégicamente; es decir, revelar unas u otras características según su valoración de la situación, y teniendo en cuenta cuánta información dispongan acerca de las intenciones y posibilidades de otros, el funcionamiento del sistema, etc. De hecho, y bajo una hipótesis general de racionalidad cada agente adoptará aquella forma de comportamiento que mejor convenga a sus intereses, y sólo actuará llanamente si las previsibles ganancias de su comportamiento estratégico se ven contrapesadas por su coste. Podría parecer que, al limitar la atención a la comunicación de información privada se estén olvidando otros aspectos relevantes para el estudio de los incentivos. Ciertamente, no todos los temas concebibles pueden contemplarse desde esta perspectiva, pero una interpretación adecuada del lenguaje formal permite abarcar mucho. Por ejemplo, si se tienen en cuenta los costes de inspección se puede considerar como privada aquella información que, aunque posible de obtener coactivamente, exigiese para ello la distracción de recursos excesivos. Asimismo, los costes y probabilidades de ser detectado como infractor por parte de cada agente pueden considerarse implícitos en la descripción de las estrategias y resultados posibles (20).

Conviene, pues, distinguir entre las verdaderas características de un agente —y en particular, los criterios con que valora los resultados de su acción— y su comportamiento efectivo, las características que revela. Supongamos una función

(19) Las precisiones anteriores parecen necesarias porque, aun cuando claramente tienen mucho en común, la microeconomía convencional, el diseño de mecanismos, la teoría de la planificación y la teoría de la elección social siguen esperando que se produzca una plena unificación entre ellas. A falta de dicha unificación, los resultados de cada una de estas corrientes de literatura económica se van desarrollando en paralelo más o menos próximo. Sin pretender tal unificación, los comentarios anteriores intentan al menos esbozar la dirección en que hay que buscarla.

(20) Existe una literatura teórica sobre evasión fiscal, en la que se tienen en cuenta consideraciones de este tipo. No conocemos ningún intento de relacionarla directamente con la literatura considerada aquí, pero presumimos que la conexión entre ambas podría resultar fructífera.

de decisión bajo la cual cualesquiera que fuesen las circunstancias, ningún individuo pudiese nunca modificar a su favor el resultado social derivado de su comportamiento llano, por mucho que intentase actuar estratégicamente. Diríamos, entonces, que la revelación correcta de las preferencias es una estrategia dominante (21) bajo dicha función de decisión, y que ésta no es manipulable. Las funciones no manipulables son interesantes porque en su seno lo racional es actuar llanamente, y cualquier actividad costosa, dirigida a fundamentar comportamientos estratégicos, resulta irracional. Sin embargo, es bien sabido que, en muchas circunstancias los agentes económicos adoptan comportamientos estratégicos. ¿Son estos comportamientos debidos a las características de los procedimientos concretos utilizados en nuestras sociedades, o se trata de un rasgo generalizado de las funciones de decisión? Acaso cualquier función de decisión abre posibilidades estratégicas... En términos generales, así es efectivamente, y diversos autores han venido a demostrarlo rigurosamente para distintos contextos (22). El formato general de los resultados consiste en establecer que, excepto con fuertes restricciones sobre las estrategias posibles (23) o si se aplican procedimientos triviales, existirán para cada mecanismo dado entornos en los cuales algún individuo encontraría ventajoso no revelar sus verdaderas características. Es decir, que, en general, el comportamiento llano no será siempre óptimo para los agentes bajo ningún mecanismo de decisión. Este hecho no implica que dichos agentes vayan a estar siempre falseando sus verdaderas características, pero sí que, en general, sus actos van a ser función de los de los demás, lo cual tiene consecuencias de carácter general sobre el tipo de enfoque a adoptar —una vez reconocido este hecho— en el estudio normativo de las funciones de decisión y de sus propiedades.

a) Dominios generales y mecanismos sin transferencias.

El primer resultado concluyente, y aún hoy fundamental para iniciar una exposición sistemática de los problemas de incentivos en el diseño de mecanismos, es el llamado teorema de Gibbard-Satterthwaite (24).

El marco en que se formula dicho resultado es el habitual en teoría de la elección social, sobre cuya conexión con otros más habituales en economía tendremos algunos comentarios más adelante.

(21) Este lenguaje anuncia el enfoque adoptado en la siguiente sección. Véase al final de aquella la razón de dicha terminología.

(22) Véanse Barberá (1977), Kelly (1978, cap. 6) y Pattanaik (1979) para contextos de elección social, y Satterthwaite y Sonnenschein (1981) para preferencias más próximas a los modelos económicos habituales.

(23) Estas restricciones pueden ser de varios tipos. Unas, como las consideradas en Kalai y Muller (1977) son de difícil interpretación, pero muy generales. Otras, como las propuestas en d'Aspremont y Gérard-Varet (1979), Moulin (1980) y Roberts (1979), reflejan posibles especificaciones de los entornos, de las preferencias o de las condiciones de información bajo las que se van a tomar las decisiones colectivas.

(24) Inicialmente demostrado, independientemente, por uno y otro autor en 1973 y 1975. Demostraciones alternativas de dicho resultado se contienen en Barberá (1983), Batteau et al (1981), Roberts (1979).

Sea $I = \{1, 2, \dots, n\}$ un conjunto de *individuos*. Sea A un conjunto finito de *alternativas* entre las que se debe escoger. Las alternativas se denotan por x, y, z, \dots

Sea P el conjunto de las relaciones binarias completas, reflexivas, transitivas y antisimétricas sobre A . Cada uno de sus elementos representa posibles preferencias estrictas (sin indiferencia) de los individuos respecto a las alternativas.

Dados $P \in P, Y \subseteq A$, el mejor elemento de P en Y se denota por

$$C(P, Y) = \{x \in Y | x P y, \quad \forall y \in Y\}.$$

Los elementos de P^n , es decir, las n -plas de preferencias (P_1, P_2, \dots, P_n) se llaman perfiles.

Una *función de decisión social* le atribuye una decisión (una alternativa) a cada posible perfil de preferencias. Es, pues, una función f de la forma $f: P^n \rightarrow A$.

Dada una función f , el rango de f , r_f , es el conjunto de los valores que toma f para al menos un perfil. Una función de decisión es *manipulable* si y sólo si existen $(P_1 \dots P_i \dots P_n) \in P^n, P'_i \in P$ tales que

$$x = f(P_1 \dots P'_i \dots P_n) \quad P_i f(P_1 \dots P_i \dots P_n) = y$$

Informalmente, vemos que f es manipulable si, dado un perfil de preferencias $(P_1 \dots P_i \dots P_n)$ ante el que f determinaría que el resultado fuese y , existe para un individuo, en este caso el i -ésimo, la posibilidad de declarar otras preferencias, las P'_i , y con ello: 1) modificar el resultado social, y 2) obtener así un resultado x mejor, de acuerdo con sus preferencias, al que resultaría de un comportamiento no estratégico.

Una función de decisión social es *dictatorial* si existe un individuo tal que los resultados de la función coinciden siempre con la alternativa que dicho individuo considera mejor, entre todas aquellas que alguna vez son escogidas. Formalmente, f es dictatorial si existe un $i \in I$ tal que, para todo perfil $(P_1 \dots P_i \dots P_n)$,

$$f(P_1 \dots P_i \dots P_n) = C(P_i, r_f).$$

El teorema de Gibbard-Satterthwaite nos dice que, si pretendemos diseñar un mecanismo capaz de indicar una decisión única para cualquier perfil de preferencias concebible, dicho mecanismo será siempre manipulable, con dos excepciones de poco interés (a menos que: 1) sea dictatorial, o 2) siempre dé lugar, como elección, a una de dos alternativas). Formalmente,

Teorema 3.—Sea f una función de decisión social, cuyo rango contenga más de dos alternativas. Dicha función es o dictatorial o manipulable.

El teorema de Gibbard-Satterthwaite nos indica la dificultad de encontrar procedimientos capaces de incentivar adecuadamente a los individuos para que declaren sus preferencias de modo veraz. En realidad, la formulación de Gibbard es más amplia, y cubre casos en los que las características individuales a revelar

podrían no ser preferencias, o, al menos, no de modo exclusivo. Son muchos ya los trabajos que en una u otra dirección demuestran que el resultado es robusto —se sigue cumpliendo en lo esencial aunque se altere en alguna medida el marco en que se formule— con tal de que las funciones consideradas deban operar sobre dominios de definición suficientemente amplios.

b) Dominios restringidos y mecanismos con transferencias.

En contraste, y si estamos dispuestos a restringirnos al análisis de aquellos casos en los que adquieren sentido las nociones de transferencias de dinero y pagos compensatorios, existen mecanismos capaces de garantizar que los individuos desestimen la posibilidad de actuar estratégicamente.

Presentaremos una formulación bastante amplia de las condiciones bajo las que se pueden inducir comportamientos veraces, y una caracterización de aquellos mecanismos que lo hacen posible. Antes, sin embargo, parece adecuado motivar lo que sigue mediante un ejemplo sencillo (25).

Supongamos una economía formada por tres agentes (1, 2 y 3), que deben decidir entre tres proyectos alternativos (A, B, C). Supongamos que tales proyectos no tienen coste, y que la realización de cada uno de ellos tendría distinto impacto sobre cada agente, impacto que cada uno de ellos es capaz de medir mediante su disposición a pagar para que se realice el proyecto. Sea d_i^x la disposición a pagar del individuo i por el proyecto x . En este contexto, toda regla de asignación eficiente exige que se escoja aquel proyecto que maximice la suma de las disposiciones a pagar —puesto que con dicho proyecto, más eventuales pagos compensatorios, podrían siempre conseguirse asignaciones Pareto-superiores a cualquier otra en que el proyecto escogido diese lugar a menores disposiciones a pagar. Más concretamente, las alternativas globales entre las que escogen los individuos en un mundo donde tiene sentido la noción de disposición a pagar no pueden ser solamente proyectos, sino pares proyecto-dinero, como en el caso descrito ya en la sección 2, teorema 1. La observación anterior nos determina necesariamente una regla para escoger la primera componente de dichos pares: el proyecto que genera mayor disposición global a pagar. Lo que queda por describir es el modo en que se determinan las cantidades de dinero disponibles para cada individuo. Queremos movernos en un contexto en el que las disposiciones a pagar por cada proyecto sean independientes de las tenencias iniciales de dinero por parte de cada agente; en tal caso, basta para describir un mecanismo con fijar reglas de transferencia que, dada cada declaración de preferencias sobre proyectos, determinen las transferencias netas a recibir (o pagar) por parte de cada agente.

Consideremos el siguiente ejemplo, para el que se especifican las verdaderas disposiciones a pagar de los tres agentes por los distintos proyectos.

(25) Este ejemplo es una adaptación del presentado por Tideman en su introducción al número de la revista *Public Choice* (1977), dedicado al mecanismo de revelación de demanda. Una discusión también introductoria, pero más sofisticada de los principios básicos de dichos mecanismos se encuentra en los primeros capítulos de Green y Laffont (1979).

Proyecto \ Agente	A	B	C
1	10	30	50
2	35	15	25
3	26	43	12
Total disposición a pagar	71	88	87

Ya hemos dicho que el proyecto que genera mayor disposición a pagar debería ser el escogido. En este caso es el B. Vemos, sin embargo, que los agentes podrían declarar preferencias distintas y, con ello, obtener cualquier resultado que desearan. Así, por ejemplo, 3 podría lograr que C fuese elegido, ya sea simplemente aumentando su valoración de C hasta 15 o disminuyendo la de B a 40. Claro está que dichas alteraciones en principio no le interesan, porque prefiere B sobre A y C. Pero también 2, para quien B es poco deseable, podrían obtener C declarando una disposición a pagar por C igual a 30, por ejemplo, o incluso A, si declara estar dispuesto a pagar 55 por dicho proyecto. Estas observaciones serán posteriormente generalizadas; diremos que el procedimiento no impone alternativas, ya que, en principio, cualquier alternativa puede obtenerse, para declaraciones adecuadas de las valoraciones.

En aquellos casos, como los mencionados para 2, en que un individuo fuese capaz de conseguir que se adoptasen proyectos mejores que el correspondiente a sus verdaderas preferencias a base de falsearlas, estará o no en su interés hacerlo según cuáles vayan a ser las consecuencias de sus acciones sobre las transferencias a pagar o recibir.

Supongamos que un individuo tuviese que pagar exactamente lo que declara como su disposición, una vez determinado el proyecto. En tal caso, su declaración de preferencias jugaría el papel clásico que desempeña —implícitamente al menos— en un sistema de precios: contribuir a determinar qué acción es eficiente y, a la vez, establecer la magnitud de su contribución a que dicha acción se lleve a cabo. Insistimos en que algo análogo sucede en un mercado: cuando una asignación es eficiente, los precios que la sostienen como equilibrio se igualan a las correspondientes relaciones marginales de sustitución, y, así, estas últimas, expresión de las preferencias de cada individuo, juegan a la vez el papel de determinante de las condiciones de eficiencia y el de determinante de los pagos a realizar. En nuestro caso, si cada agente debe pagar lo que declara como su disposición, el mecanismo que esto define, junto con la regla de escoger el proyecto con mayor disposición global, será manipulable. Veamos un ejemplo: declarando sus verdaderas preferencias, el individuo 1 ve elegido el proyecto B a un coste para él de 30 unidades monetarias. Si, en lugar de declarar sus verdaderas preferencias, hubiese declarado

que su disposición a pagar por A y por B es nula en ambos casos, y que valora C en 25, las disposiciones totales a pagar serían 61 para A, 58 para B y 62 para C. El individuo 1 lograría ver escogido C, que prefiere sobre B, y a un coste de 25, menor aún que lo que le hubiese correspondido declarando sus verdaderas preferencias.

Es posible definir un sistema de transferencias basado en las disposiciones a pagar declaradas, pero que no coinciden con éstas, y con la propiedad de que, cuando el proyecto se decide del modo eficiente ya descrito, lo mejor para cada uno de los agentes es declarar sus verdaderas preferencias. En nuestro caso, un sistema de transferencias con esta propiedad nos viene dado por los llamados impuestos de Clarke, que se calculan del siguiente modo. Para cada individuo, determínese lo que todos los demás, excepto él mismo, estarían dispuestos a pagar por cada proyecto y, sobre esta base, cuál sería el proyecto elegido en su ausencia. Si dicho proyecto es el mismo que se escoge teniendo en cuenta a todos, el individuo no es pivote y no debe pagar nada. Si el proyecto es otro, calcúlese la diferencia entre las disposiciones a pagar de otros por el proyecto que prefieren sin un individuo, y su disposición a pagar por aquel proyecto que, teniendo en cuenta a la totalidad, se acaba por elegir. Esta diferencia, que mide el coste que la influencia de dicho individuo sobre la decisión colectiva impone sobre los demás miembros de la sociedad, es el impuesto de Clarke que le corresponde pagar. La siguiente tabla resume los sencillos cálculos involucrados (recordemos que la decisión óptima, en nuestro ejemplo, es B).

Proyectos		A	B	C		
Disposiciones totales a pagar	sin 1	61	58	37	A	sin 1
	sin 2	36	73	62	B	sin 2
	sin 3	45	45	75	C	sin 3

decisión

$$\text{Impuesto de Clarke} \begin{cases} \text{Agente 1: } 61 - 58 = 3 \\ \text{Agente 2: } 0 \\ \text{Agente 3: } 75 - 45 = 30 \end{cases}$$

El lector podrá comprobar que, con este sistema de transferencias, ningún individuo encontrará factible falsear sus preferencias. Esto es así porque: 1.º) lo que se paga es independiente de la cantidad exacta declarada, con tal de que no varíe la decisión colectiva, y 2.º) el coste de variar la decisión colectiva resulta sistemáticamente superior a las ganancias que se derivan de hacerlo. Veamos, por ejemplo, el caso utilizado para mostrar la manipulabilidad del sistema en que todos pagan según su disposición. Teníamos que, en aquel caso, 1 hubiese derivado ventajas declarando que sus disposiciones a pagar por A, B y C eran, respectiva-

mente, 0, 0 y 25. Calculemos el impuesto de Clarke correspondiente a dichas preferencias. La decisión global pasa a ser C. La decisión, en ausencia de 1, no cambia, y sigue siendo A. El nuevo impuesto de Clarke es, pues, la diferencia entre las disposiciones a pagar conjuntamente de 2 y 3 por A y C, igual a 24. Las ganancias de 1 son, pues, que obtiene C en lugar de B —ganancias iguales a 20—. Los costes para obtenerlas son la diferencia entre el impuesto de Clarke que corresponde a su falseamiento, 24, y el que pagaría declarando sus verdaderas preferencias, 3. Estos costes, 21, son superiores a las ganancias del falseamiento.

Lo aquí descrito no es una feliz casualidad, sino un ejemplo de cómo, en contextos donde quepa hablar con propiedad de disposiciones a pagar, y donde pueda descomponerse la descripción de las decisiones relevantes en términos de una decisión pública y un conjunto de transferencias privadas, es posible diseñar mecanismos que induzcan a la revelación correcta de preferencias. En qué contexto exacto, y con qué mecanismos puede lograrse, es lo que expresa con bastante generalidad un teorema, debido a K. Roberts, que pasamos a exponer (26).

Las economías que considera K. Roberts como datos, y para los que estudia la posibilidad de diseñar los mecanismos no manipulables son, esencialmente, las ya descritas en la sección 2, y para las cuales hemos expuesto el teorema de Mas Colell sobre la conexión entre equilibrios y óptimos.

Se parte de un número finito de agentes $I = \{1 \dots n\}$, y un número también finito (≥ 3) de «proyectos» o decisiones públicas $K = \{\dots x, y, z, \dots\}$. Los estados relevantes para el bienestar de los individuos pueden describirse como pares, que especifican en cada caso un proyecto y una cantidad de «dinero». Los estados sociales quedarán determinados por un proyecto o decisión —de aplicación común a todos los agentes— y por n cantidades $t_1 \dots t_n$, cada una de ellas representativas de la cantidad de dinero de que dispone el correspondiente individuo en dicho estado.

Supondremos que las preferencias de los individuos sobre pares de este tipo pueden representarse mediante funciones de utilidad aditivamente separables, de la forma

$$U_i(x, t_i) = u_i(x) + t_i$$

En estas condiciones, conocer la función $u_i(\cdot)$ equivale a conocer totalmente las preferencias del individuo i , y esta función, en presencia de p proyectos, viene dada por un vector p -dimensional. Un perfil de preferencias es, aquí, un vector np -dimensional. A los perfiles los denotamos por $u = (u_1, u_2, \dots, u_n)$. La notación u_{-i} se refiere al vector obtenido eliminando la i -ésima componente de u . Así, $u = (u_i, u_{-i})$.

(26) Véase Roberts (1979). El artículo de Roberts forma parte de un volumen editado por J. J. Laffont, en el que se recogen muchas contribuciones relevantes a la literatura sobre incentivos. En realidad, la presentación que hace Roberts de su resultado es más general que la aquí expuesta, ya que se incluiría en el marco de la sección 4 de este trabajo. Pero nuestra reformulación, para adoptarla al presente marco expositivo, respeta fielmente el contenido esencial.

Una función de decisión social, es una función $f: U \rightarrow X$, siendo U el conjunto de los perfiles posibles en R^n , cada uno de la forma $u(\cdot)$. Se trata, pues, de una regla para determinar qué proyecto debe seleccionarse, dadas las preferencias de los agentes.

Sin embargo, una decisión sobre proyectos no configura la totalidad de las consecuencias posibles en un proceso de asignación, si es que, simultáneamente, pueden producirse transferencias. Por ello se introduce la noción de función de transferencia,

$$t: U \times I \rightarrow R;$$

para cada perfil de preferencias, u indica la transferencia neta $t(u, i) \equiv t'_i(u)$ que le corresponde al agente i .

Un mecanismo vendría dado por un par (f, t) , donde f es una función de decisión social y t una función de transferencia. Obsérvese que para el teorema de Gibbard-Satterthwaite nos hemos movido en un marco bajo el que no existían transferencias, y sólo resultaba, pues, relevante f .

En el contexto actual, podemos decir que un mecanismo (f, t) no es manipulable, si para todo $i \in I$, $u \in U$, u'_i ,

$$u_i(x) + t(u_i, u_{-i}, i) \geq u_i(y) + t(u'_i, u_{-i}, i)$$

siendo $x = f(u_i, u_{-i})$ e $y = f(u'_i, u_{-i})$.

Diremos que un mecanismo no impone resultados si, para cualquier proyecto $x \in T$, existen preferencias $u \in U$ para las que $f(u) = x$. Con estas definiciones, podemos ahora formular una versión del Teorema de K. Roberts:

Teorema 5.—Una función de decisión social f que no imponga resultados es no manipulable si y sólo si existen números $k \in R^n$, $k_i > 0$, y una función $F: T \rightarrow R$, tales que, para todo $u \in U$,

$$f(u) \in \{x \mid \sum k_i u_i(x) + F(x) \geq \sum k_i u_i(y) + F(y), \quad \forall y \in X\}.$$

La función de transferencias t que, junto con f , constituye el mecanismo capaz de inducir a la revelación de las verdaderas preferencias, debe ser de la forma

$$t(u_1, \dots, u_n, i) = \frac{1}{k_i} \left[\sum_{j \neq i} k_j u_j(x) + F(x) \right] + h_i(u_{-i})$$

siendo $x = f(u)$, y h cualquier función con valores reales definida sobre las posibles disposiciones a pagar manifestadas por los agentes distintos de i .

Interpretemos este resultado. Los valores de $F(\cdot)$ dependen de cada proyecto, pero no de las preferencias de los agentes. Cada valor de k_i puede interpretarse como la ponderación que la «disposición a pagar» de cada individuo por cada proyecto recibe al calcular la disposición total a pagar ponderada, $\sum k_i u_i(x)$ que

corresponde al proyecto. La primera parte de la proposición nos indica, por tanto, que el proyecto elegido debe «maximizar la disposición a pagar de los individuos», en un sentido amplio. Una versión estricta de dicha maximización se limitaría a considerar que todos los k_i son iguales a uno y todos los valores de F iguales a cero. La extensión desde la regla sencilla a la aquí considerada permite englobar aquellos casos en que algunos individuos tienen mayor influencia que otros (mayores valores de k_i), o algunas alternativas mayores posibilidades de ser elegidas [mayores valores de $F(\cdot)$].

Por otra parte, veamos qué forma toma la transferencia recibida por un individuo i cuando la alternativa escogida es la alternativa x .

En tal caso, la transferencia recibida por i toma la forma

$$t(u_1, \dots, u_i, \dots, u_n, i) = \frac{1}{k_i} \left[\sum_{j \neq i} k_j u_j(x) + F(x) \right] + b_i(u_{-i})$$

Supongamos que cambiando sus preferencias a u'_i el individuo i lograra hacer cambiar el valor de $\sum_{j \neq i} k_j u_j(\cdot) + F(\cdot)$ y, con ello, que el proyecto escogido pasara a ser y . La transferencia recibida por i pasaría a ser

$$t(u_1, \dots, u'_i, \dots, u_n, i) = \frac{1}{k_i} \left[\sum_{j \neq i} k_j u_j(y) + F(y) \right] + b_i(u_{-i})$$

El cambio de uno a otro valor de la transferencia puede entenderse como el coste monetario para i de alterar el resultado mediante su cambio de preferencias, y se puede calcular por diferencia.

$$\begin{aligned} t(u_1, \dots, u_i, \dots, u_n, i) - t(u_1, \dots, u'_i, \dots, u_n, i) = \\ = \frac{1}{k_i} \left[\sum_{j \neq i} k_j u_j(x) - \sum_{j \neq i} k_j u_j(y) + F(x) - F(y) \right] \end{aligned}$$

Este valor es una versión generalizada de lo que sería simplemente, para $k_i=1 (\forall i)$, $F(x) \equiv 0$, la diferencia entre las disposiciones globales a pagar por parte de los demás individuos de la sociedad por x y por y . Por tanto, las transferencias recibidas por los individuos, o, mejor dicho, la diferencia entre lo que reciben dada la decisión social que contribuyen a formar y lo que recibirían si, cambiando sus preferencias, hicieran que otro proyecto fuera el escogido, deben ser una medida —una suma ponderada y acaso alterada por un factor de escala $F(\cdot)$ — de los costes que cada individuo, a través de su influencia sobre el proyecto escogido, impone a los demás.

De este resultado de caracterización se derivan múltiples consecuencias. Una, que no desarrollaremos aquí, pero que se sigue fácilmente, la de que no es posible

en general conseguir mecanismos que induzcan a revelar correctamente las preferencias individuales sin complementar la regla de elección de proyectos mediante un sistema de transferencias adecuado. Otra, según acabamos de ver, que dichas transferencias deben modificarse cuando un agente sea pivote, de tal modo que éstas reflejen el coste que la presencia del individuo en la sociedad impone sobre los demás agentes, reflejo que encuentra expresión en la forma funcional indicada por la segunda parte del resultado. Precisamente, uno de los problemas que permite destacar esta formulación con mayor claridad es el derivado de la relativa estrechez de los márgenes que dicha forma funcional deja abiertos para el diseño de mecanismos no manipulables. Ciertamente, a través de la elección de k y de $F(\cdot)$ es posible dar tratamientos diferenciados a distintos individuos y alternativas, pero no demasiado, ya que ni unas ni otras ponderaciones pueden depender de las preferencias. Asimismo, es cierto que las funciones $b_i(\cdot)$ permiten modificar los valores absolutos de las transferencias individuales, pero deben hacerlo de modo que las diferencias entre dichas transferencias, al no depender $b_i(\cdot)$ de las preferencias de i , sean independientes de $b_i(\cdot)$. Precisamente esta falta de correspondencia entre las $b_i(\cdot)$ y las preferencias de i , que impide que el valor de las transferencias sea flexiblemente dependiente de las decisiones públicas adoptadas, hace imposible en general garantizar que las transferencias netas entre individuos asociados con un mecanismo no manipulable puedan ser nulas: en general, dichos mecanismos generan inevitablemente déficit o superávit, y, por tanto, no pueden alcanzar asignaciones (x, u) eficientes en sentido pleno, como las consideradas en el teorema 1, sino simplemente asignaciones que, en lo relativo al proyecto elegido —pero no a las transferencias— satisfacen una versión limitada de aquella condición (27).

Podemos establecer un paralelo claro, y señalar las diferencias esenciales, entre lo que nos dicen el teorema de Gibbard-Satterthwaite por un lado, y el de K. Roberts, por otro. En ambos aparece como elemento clave para el diseño de mecanismos no manipulables el papel de aquellos individuos que, siendo capaces de alterar el resultado social, el proyecto elegido o la alternativa global resultante de dicha elección más las transferencias, según los casos, deben decidir si declarar sus verdaderas características o, por el contrario, falsearlas en su propio beneficio. Estos individuos, *pivotes*, encontrarán siempre oportunidades de manipular, con escasas excepciones poco interesantes (Teorema de Gibbard-Satterthwaite), a menos que el contexto económico sea lo suficientemente específico para que: *a*) tenga sentido hablar de transferencias compensatorias de dinero, y suponer que la disposición a pagar de los individuos es independiente del volumen de transferencias, y *b*) éstas toman formas específicas bajo las que existe cierto coste de eficiencia para lograr incentivos adecuados (Teorema de Roberts).

A la vista de estos resultados procederemos, en la última parte de este trabajo, a describir y comentar un marco más general en cuyo seno, creemos, podrían irse

(27) Intentar acotar el alcance de estos inconvenientes, para su eventual propuesta como procedimientos prácticos, es el espíritu de varias propuestas en el número ya citado de *Public Choice*, y de gran parte del libro de Green y Laffont. No parece que se haya logrado avanzar lo suficiente en este terreno para hacer de los mecanismos de revelación de demanda la base de propuestas institucionales inmediatamente practicables.

reformulando en el futuro todas las cuestiones de la economía normativa para cuya solución resulte relevante tener en cuenta los incentivos de los agentes. Aun cuando dentro del marco propuesto se han dado avances interesantes, pensamos que es más lo por venir que lo ya conocido, y nuestra exposición, salvando algunas referencias, se centrará a partir de aquí más en planteamientos que en resultados.

5. PUESTA EN PRACTICA DE FUNCIONES DE ELECCION: UN MARCO GENERAL

En realidad, el estudio de mecanismos no manipulables constituye sólo un modo muy restringido de abordar las cuestiones generales de diseño de sistemas suscitadas por una consideración explícita del posible comportamiento estratégico de los agentes.

Un planteamiento general debe reconocer distintos niveles de análisis, que pasamos a describir (28). En primer lugar, dado un conjunto de entornos económicos, podemos distinguir para cada uno de ellos entre las asignaciones que son «aceptables» y aquellas que no lo son, de acuerdo con el o los criterios que quieran adoptarse como punto de partida del análisis. En general, llamaremos correspondencia de decisión social a aquella que le atribuye el conjunto de asignaciones «aceptables» a cada entorno dentro de nuestro campo de análisis. Como ya hemos señalado en la sección anterior, una correspondencia de decisión social puede ser simplemente resultado de la aplicación de un criterio normativo (por ejemplo, la correspondencia de Pareto, que a cada entorno le asocia aquellas asignaciones factibles que son eficientes en el sentido de Pareto); o pueden ser la composición de un criterio normativo y de algún procedimiento adicional para seleccionar entre diversas asignaciones que lo satisfacen (por ejemplo, la correspondencia de Walras para entornos donde existen y son eficientes los equilibrios competitivos).

También aquí, y a efectos de simplificación, nos limitaremos a hablar de aquellos casos en los que —por conjunción de criterios normativos y procedimientos adicionales de selección— para cada entorno tengamos un solo estado deseable: es decir, cuando la correspondencia de decisión social sea una *función*.

Supongamos dada una función de decisión. Esta establece una relación sistemática entre características de un entorno y resultados. Por otra parte, su campo de variación establece los entornos admisibles y, por tanto, las características que cada agente puede declarar sin que, salvo inspecciones no contempladas aquí, nadie pueda distinguir si son o no sus verdaderas características: define, pues, las posibles estrategias de un agente que se plantee si revelar o no sus verdaderas características, y cuáles revelar en caso contrario. Cada función de decisión deter-

(28) Aunque reciente, la literatura sobre puesta en práctica (*implementation*) es extensa. El lector interesado puede encontrar artículos y referencias en los dos números monográficos dedicados al tema por la *Review of Economic Studies* (1979) y *Journal of Mathematical Economics* (1982).

mina así un conjunto de estrategias para cada agente y una función de resultados, que juntos constituyen una estructura de juego (29), así llamada porque aporta dos de las tres piezas definitorias de un juego en forma normal, a falta sólo de especificar la valoración que cada agente le atribuya a cada resultado.

Es más, reconocida la posibilidad de que los agentes adopten actitudes estratégicas en un intento de sacar ventajas del modo en que un mecanismo tiene en cuenta sus características, es también concebible que el diseñador social incorpore tales comportamientos a su análisis, e intente forzar la racionalidad de los agentes desde un plano de sofisticación más elevado que si se tratase de diseñar autómatas. Un modo de hacerlo, en general, consiste en diseñar juegos, acaso aparentemente inconexos con la función de decisión social en que estamos interesados, pero que sean capaces de conducir a comportamiento que consigan en la práctica dicha función.

A fin de ser más específicos, vamos a empezar con algunas definiciones (30). Como nuestro propósito es esencialmente conceptual, nos moveremos dentro del marco más sencillo, aquel para el que ya hemos discutido el teorema de Gibbard Satterthwaite.

Sea $A = \{x, y, z, \dots\}$ un conjunto finito de m alternativas.

Sea $I = \{1, 2, \dots, n\}$ el conjunto de los individuos.

Sean P las posibles ordenaciones estrictas de preferencias sobre A , y P^n el conjunto de perfiles de preferencias.

Una función de decisión social para (I, A) es una función

$$f : P^n \rightarrow A.$$

Si tomamos los conjuntos de alternativas y de individuos como datos, una estructura de juego vendrá dada por especificación de las estrategias disponibles para cada agente, y por una función de resultados, que le asocie una alternativa a cada n -pla de estrategias admisibles. Formalmente,

Una estructura (o forma) de juego para (I, A) es una

$$n+1\text{-pla } F = [\varphi_1 \dots \varphi_n; g] \equiv [\varphi; g],$$

donde φ_i es un conjunto finito de estrategias, $\varphi_i = \{s_1 \dots s_{n_i}\}$ y $g : \varphi \rightarrow A$ es una función de resultados.

(29) Traducimos así el término *game form*, introducido por Gibbard (1973) y hoy plenamente aceptado. A veces se utiliza también el término mecanismo.

(30) Las definiciones que siguen permiten enmarcar y clasificar las distintas aportaciones dentro del área, cuyo contenido es muy variado. Algunos trabajos en los que se va decantando el enfoque y la terminología aquí descritos son los siguientes: Barberá (1982), Dasgupta et al (1979), Hurwicz (1979), Hurwicz y Schmeidler (1978), Maskin (1979), Moulin (1979), Sen(gupta) (1983), y, en general, los contenidos en Laffont (1979), *Review of Economic Studies* (1979) y *Journal of Mathematical Economics* (1982).

Sean F el conjunto de todas las formas de juego en (I, A) , y F_f el conjunto de los espacios de estrategias posibles.

Dada una estructura de juego $F=[\varphi; g]$, los elementos de $\varphi=\varphi_1 \times \dots \times \varphi_n$ se denotan como $s=(s_1, s_2, \dots, s_n)=(s_{-i}, s_i)$, donde $s_i \in \varphi_i$ y $s_{-i} \in \varphi_1 \times \dots \times \varphi_{i-1} \times \varphi_{i+1} \times \dots \times \varphi_n \equiv \varphi_{-i}$.

Dos estrategias $s_i, s'_i \in \varphi_i$ son *equivalentes* en $F=[\varphi; g]$ si y sólo si, para todo $s_{-i} \in \varphi_{-i}$, se cumple que $g(s_i, s_{-i}) = g(s'_i, s_{-i})$.

Si, sobre una estructura de juego, se especifican además las preferencias de cada agente sobre los posibles resultados (sus funciones de pago, en sentido ordinal), queda determinado un *juego en forma normal*.

Así, un *juego* (en forma normal) para (I, A) , es una $(2n+1)$ -pla $G=[\varphi_1, \dots, \varphi_n; g; P_1, \dots, P_n] \equiv [F; \underline{P}]$, donde F es una estructura de juego y $\underline{P}=(P_1, \dots, P_n) \in P^n$ es un perfil de preferencias sobre A .

Como nos interesa estudiar mecanismos que tengan en cuenta las preferencias de los agentes (en general, sus características), aquellas estructuras de juego en que las estrategias son, precisamente, declaraciones de preferencias (características), resultan especialmente atractivas.

Diremos que *una estructura de juego F es directa* si, para todos los agentes i , $\varphi_i \equiv P_i$.

Dada una función de decisión social f , induce de modo natural una estructura de juego directa muy específica; aquella para la que, además de ser $\varphi_i \equiv P_i$, la función de resultado sea la propia f .

Diremos entonces que, dado f , *la propia estructura de juego inducida por f es la que viene dada por $\bar{F}=[P^n; f]$.*

Dado un juego en forma normal, admite distintos tipos de solución (o de equilibrio), cada uno de los cuales corresponde a distintas hipótesis acerca del comportamiento estratégico de los agentes. Como nuestro propósito es describir, en términos generales, el tipo de cuestiones que se pueden formular y eventualmente resolver una vez especificado el concepto de solución que en cada caso vaya a ser relevante, daremos una descripción muy amplia de lo que es un concepto de equilibrio o solución.

Un concepto de equilibrio o de *solución* para estructuras de juego sobre (I, A) es una correspondencia.

$$E : F \times P^n \rightarrow F_\varphi,$$

tal que

- i) para todo $F=[\varphi; g]$, $\underline{P} \in P^n$, $E(F, \underline{P}) \subseteq \varphi$, y
- ii) para todo $F=[\varphi; g]$, $\underline{P} \in P^n$, $i \in I$ y $s_{-i} \in \varphi_{-i}$,
 $[s_i \text{ y } s'_i \text{ son equivalentes en } F] \rightarrow [(s_i, s_{-i}) \in E(F, \underline{R}) \leftrightarrow (s'_i, s_{-i}) \in E(F, \underline{R})]$

En definitiva, un concepto de equilibrio o solución determina, para cada juego dentro de la estructura (que viene dado por un perfil de preferencias), un conjunto de estrategias, con el requisito de que, si dos estrategias son en todo punto equivalentes, o ambas o ninguna deben ser de equilibrio.

Estamos ahora en condiciones de introducir los conceptos clave para nuestros propósitos. Diremos que una función de decisión social puede ponerse en práctica (para un concepto determinado de solución), si existe una forma de juego tal que, para cada especificación de las preferencias de los individuos, las soluciones del correspondiente juego existen y, además, dan lugar al mismo resultado que la función de decisión social considerada. Al igual que pasa con otros muchos conceptos teóricos, que una función se pueda poner en práctica no tiene por qué verse como una propiedad excesivamente satisfactoria; más bien se trata de un requisito mínimo, que de no darse cerraría toda posibilidad de que pudiera diseñarse procedimiento alguno, por complejo que fuese, mediante el cual pudiese garantizarse el cumplimiento de la relación deseada entre características y resultados.

Formalmente, una función de decisión social f para (I, A) puede ponerse en práctica, bajo el concepto de solución E , si, y sólo si, existe una estructura de juego $F = \{ \varphi; g \}$ para (I, A) , tal que

$$(\forall P \in P^n) [E(F, P) \neq \emptyset \ \& \ (s \in E(F, P) \rightarrow g(s) = f(P))]$$

Esta noción general de puesta en práctica puede ser objeto de algunas críticas. La posible puesta en práctica mediante juegos es más o menos interesante según el tipo de juego de que se trate. Si éste se basa en estrategias abstractas, procesadas mediante una función de resultados sin ninguna conexión con f , la puesta en práctica de f por tal método supone un elevado grado de maquiavelismo por parte del diseñador social, y cierta pasividad por parte de los agentes. En efecto, la interpretación más natural nos lleva a pensar en un sistema institucional complejo, donde se somete a los individuos a complicadas operaciones que ponen a prueba sus habilidades estratégicas, con el único fin de evitar que éstos consigan las posibles ganancias que serían capaces de obtener si se les indicase el objeto final del ejercicio y se les preguntase simplemente por sus preferencias.

Por ello, serán especialmente interesantes aquellas funciones que puedan ponerse en práctica mediante una estructura de juego directa, es decir, mediante estructuras de juego que tienen como espacios de estrategias los de preferencias, por lo que permiten plantear toda la comunicación entre agentes en términos de la información que, verdaderamente, es relevante. Hablaremos en este caso de funciones que pueden ponerse en práctica *directamente*.

Entre estas últimas resultan atractivas sobre todo aquéllas que pueden ponerse en práctica mediante la propia estructura inducida por f . Diremos entonces que f se pone en práctica *por sí misma*. Siguiendo con la interpretación, si una función cumple este último requisito es porque cuando conocen cómo opera sobre la base del concepto de equilibrio que estemos utilizando en cada caso, los agentes adoptarán acciones de equilibrio conducentes al resultado deseado, ya sea porque: 1) todos ellos deciden declarar las verdaderas características; ya sea porque, 2) las

desviaciones que diversos agentes puedan decidir adoptar son tales que se compensan entre sí, y no alteran el resultado conjunto.

Esta última distinción debe servir para advertir que el papel de la verdad, como un modo de poner en práctica funciones de decisión, no es un asunto trivial. Hasta cierto punto, puede verse que un comportamiento veraz de los agentes es suficiente para que la función pueda ser puesta en práctica por sí misma [observación 1)], pero caben en principio otras posibilidades [observación 2)]. Que ambos conceptos —ser veraz, y poder ponerse en práctica bajo un concepto de equilibrio dado— no son equivalentes, debe quedar claro con la siguiente definición.

Dada una estructura de juego directa, $F=[P^n; g]$ y un concepto de solución E , diremos que con ellos se *induce a la verdad* si, para todo perfil de preferencias $P \in P^n$, $P \in E(F, P)$, es decir, si los agentes consideran sistemáticamente que las preferencias con que ellos valoran el juego constituyen a su vez estrategias de equilibrio.

Introducidos estos conceptos, podemos recuperar nuestra línea principal de argumentación. Al final de la sección anterior decíamos que nuestro propósito en ésta era describir un marco general dentro del que se va a ir reescribiendo la economía normativa. Anteriormente hemos argumentado que dicha reformulación pasaría por distinguir radicalmente entre problemas con o sin contenido estratégico. Ejemplos de resultados sin referencias estratégicas, para el caso de economías con bienes públicos (y, en el segundo de los casos allí tratados, también con efectos externos) son los citados en el apartado segundo de este trabajo: a precios personalizados adecuados, ciertas nociones de equilibrio garantizan, de verse satisfechas, la optimalidad de los estados a los que conducen. Si, ahora, consideramos la correspondencia que, a cada especificación de una economía de los tipos allí considerados, le atribuye el conjunto de asignaciones de equilibrio, podemos formalizar el problema de la plausibilidad de dicha noción de equilibrio identificándolo con el problema de «puesta en práctica» que acabamos de describir, con la salvedad de que, por razones de simplificación, seguiremos hablando desde ahora como si aquella correspondencia fuese, en todo caso, una función; es decir, como si los equilibrios fuesen siempre únicos. Veamos cómo el propio marco analítico nos va obligando a distinguir entre varios tipos de cuestiones —importantes todas, pero distintas— en un análisis de incentivos.

Una primera cuestión se refiere al *concepto de solución* que vaya a aplicarse en el análisis de puesta en práctica. Cada concepto de solución traduce, formalmente, diversas consideraciones acerca de los individuos: si van o no a cooperar, y sobre qué bases; de qué información van a servirse; qué objetivo van a perseguir, etc.

Así, por ejemplo, la solución de Nash supone implícitamente un marco no cooperativo, información sobre las acciones de otros agentes y maximización de niveles de utilidad; las soluciones que suponen que cada individuo adopta estrategias *maxmin* son también no cooperativas, pero suponen ausencia de información acerca de los demás agentes, y que cada uno intenta minimizar pérdidas; el núcleo o core, como concepto de solución, supone cooperación y maximización de ganan-

cias, pero su aplicación nos exigiría especificar lo que cada subconjunto de agentes puede garantizarse por sí mismo, a fin de poder definir los casos en que cada coalición puede objetar contra una asignación, etc. Muchas discusiones aparentemente orientadas por consideraciones «prácticas», acerca de la mayor o menor viabilidad de una institución pueden verse como desacuerdos acerca del grado de cooperación, o de información, o acerca de la capacidad para llevar a cabo y hacer creíbles amenazas, por parte de distintos agentes. Así, el problema del *free-rider* en presencia de bienes públicos suele modelarse señalando que los equilibrios de Lindahl no constituyen una solución de Nash en el juego inducido por aquel mecanismo, pero una opinión menos pesimista sobre el grado de cooperación entre agentes podría expresarse argumentando que el concepto de solución adecuado es el núcleo y que bajo condiciones adecuadas los equilibrios de Lindahl pueden coincidir con dicha solución, minimizando así el problema de asignación y la necesidad de intervencionismo (31). Así, pues, vemos cómo algunas ideas muchas veces confusas y/o contrapuestas podrían incluirse analíticamente como componentes de un modelo formal, a través de la fijación del concepto de solución a adoptar.

Una segunda cuestión se refiere al propio *juego* que se vaya a analizar. Una aproximación inmediata consiste en preguntarse por la posibilidad de que la función en que estemos interesados se ponga en práctica a sí misma. Cuando las imágenes de dicha función fuesen —concretamente— equilibrios derivados del comportamiento de varios agentes, que ésta se pusiera en práctica a sí misma sería la expresión formal de que los individuos encuentran incentivos para actuar según las reglas implícitas en la definición de equilibrio (puesta en práctica veraz) o, por lo menos, para actuar en formas que acaban conduciendo, a pesar de sus cálculos estratégicos, a los mismos resultados que si no los hubieran hecho. Si, por ejemplo, un equilibrio con impuestos pigouvianos pudiese ponerse en práctica a sí mismo, dado un determinado concepto de solución, esto indicaría que el comportamiento postulado por la noción de equilibrio —según el cual los agentes informan de los beneficios o daños recibidos en forma de efectos externos— resulta plausible. Cuando, por el contrario, no pudiese ponerse en práctica a sí mismo, entenderíamos que —bajo las condiciones estratégicas especificadas— les cabe a estos agentes incumplir con las prescripciones del equilibrio, dar falsa información sobre los daños y beneficios recibidos, y manipular con ello la función de decisión social que tratamos de poner en práctica: es un caso en que las instituciones, tal como aparecen descritas en el equilibrio (en nuestro ejemplo, los impuestos pigouvianos), no resultan plausibles.

Pero el juego relevante no tiene por qué ser el inducido por la propia función a poner en práctica. Si cabe hacerlo, puede ser interesante someter a los agentes a juegos más complejos, capaces de poner en práctica una función aun cuando ésta no pueda hacerlo por sí misma. La interpretación de esta cuestión abstracta para el caso en que la función en cuestión tuviese como imagen un determinado equilibrio podría verse a través de un ejemplo: si, en general, los equilibrios de Lindahl con bienes públicos dan lugar al problema del *free rider*, es porque existen situaciones en las que revelar las verdaderas relaciones marginales de sustitución entre bien

(31) Véase Mas Colell (1980).

público y privado no constituye un equilibrio de Nash para los agentes, en el juego directamente inducido por la propia función que a cada economía con bienes públicos le asocia uno de sus equilibrios de Lindahl. Si, a pesar de ello, existiera una estructura de juego capaz de poner en práctica aquella función, sus estrategias deberían verse como modos de responder a cuestiones distintas, en general más complejas, que las relativas a relaciones marginales de sustitución, cuestiones tales que, después de todo, conduzcan a respuestas-solución que dan lugar al resultado deseado.

De ahí que, al decidir si nos interesamos por un tipo u otro de juego con el que poner en práctica una función —el propio inducido, otro directo, o uno general cualquiera— estemos decidiendo sobre el grado de complejidad y de sofisticación con el que consideramos posible influir sobre los incentivos de los agentes para participar en las instituciones (32).

Una tercera cuestión tiene que ver con la validez de determinadas afirmaciones, no necesariamente de alcance universal, bajo *condiciones particulares*. Ya hemos visto que, en determinados contextos, existen funciones de decisión no manipulables, mientras que en otros más generales, el teorema de Gibbard-Satterthwaite nos indica que esencialmente no las hay. Y, como dentro de poco vamos a comentar, de la manipulabilidad o no de una función depende la posibilidad de ponerla en práctica en una forma particular. Del mismo modo, para cualquier forma de puesta en práctica podrá haber conjuntos de circunstancias tales que de darse, hacen plausibles los comportamientos predichos implícitamente por la función que nos interese. Por ejemplo, cuando las preferencias son suficientemente homogéneas y pueden representarse unimodalmente (*single-peaked*), es posible decidir sobre la cantidad de un bien público por mayoría simple, y este método es muy poco manipulable (33). Otro ejemplo de caso particular especialmente relevante al tratar de incentivos es el que viene asociado al *tamaño de la sociedad*. Toda discusión del comportamiento competitivo y de su plausibilidad se refiere en algún punto al papel de los grandes números: en sociedades con muchos agentes, se aduce, cada uno de ellos tiene poco poder y no le compensa ejercer el escaso o nulo poder estratégico del que pudiere disfrutar. Afirmaciones tan generales son difíciles de rebatir o demostrar, pero diversos autores han señalado que desde varios ángulos, la conexión entre comportamiento competitivo y número de agentes es mucho más sutil: puede haber comportamientos competitivos con pequeños números, y no haberlos en sociedades con muchos agentes (34). Y la noción de comportamiento competitivo, entendido aquí como renuncia a manipular ciertas restricciones de un sistema (precios, por ejemplo), y tomarlas como parámetros, no es aquí sino una forma de aproximar cuestiones de manipulación, formas particulares a su vez de la cuestión más amplia de puesta en práctica.

Para terminar este apartado, pongamos en relación la noción de puesta en

(32) Para ejemplos de puesta en práctica mediante juegos más complejos que el directamente inducido por la función deseada, véase Hurwicz (1979), Hurwicz y Schmeidler (1978), Moulin (1979, 1982).

(33) Véanse Moulin (1980) y Pattanaik (1979).

(34) Por ejemplo, Ostroy (1980).

práctica aquí descrita con la de no manipulación considerada en el apartado anterior. Para ello conviene describir la noción de estrategia dominante. Una estrategia es dominante para un individuo, en un juego determinado, si dicho individuo obtiene siempre mejor resultado (o al menos tan bueno) utilizándola que haciendo uso de cualquier otra, y esto con independencia de cuales sean las acciones de todos los demás agentes. Una solución en estrategias dominantes, de existir, es una n -pla de estrategias cada una de las cuales es estrategia dominante para el agente correspondiente, del 1 al n .

Con esta terminología, podemos ver que una función de decisión no es manipulable si la verdad es siempre una estrategia dominante de los juegos inducidos por ella, para cada perfil de preferencias. Asimismo, si una función de decisión no es manipulable, se pondrá en práctica a sí misma, con estrategias dominantes, medianamente la verdad. Esta forma de reformular, respectivamente, la definición y la principal consecuencia de que una función no sea manipulable, nos permite insistir aquí en el interés y las limitaciones del análisis llevado a cabo en la sección anterior.

Quando existen estrategias dominantes para todos los individuos en un juego, dicho juego es trivial: no hay espacio para ninguna consideración estratégica, ya que la mejor acción de cada uno está definida y es independiente de las de los demás. Si, además, tales estrategias coinciden con la verdad, no sólo resulta inambiguo cuál vaya a ser el comportamiento de cada agente, sino que éste pone en práctica la función. Además, la puesta en práctica es descentralizada, en un sentido pleno: cada cual puede decidir su acción sin necesidad de coordinar con nadie. Todo ello hace que, de existir, una función no manipulable disfrute de características estratégicas óptimas, ya que puede ponerse en práctica sencillamente, y bajo un concepto de solución que engloba toda hipótesis razonable sobre el comportamiento individual.

Sin embargo, en muchos contextos no es fácil que existan funciones no manipulables, e incluso allí donde las haya no tienen por qué coincidir con aquéllas que deseemos poner en práctica. Por ello resulta útil, en principio, disponer del marco que acabamos de describir. Pero también quisiéramos observar que, allí donde es posible, la puesta en práctica directa y con estrategias dominantes es la forma más natural de garantizar la plausibilidad de un concepto de equilibrio, a través de la función de decisión social a la que da lugar.

* * *

Queda por ver hasta qué punto los resultados abstractos sobre puesta en práctica irán permeabilizando la economía del bienestar más tradicional, y con qué velocidad. Existen obstáculos evidentes, como los derivados de la complejidad de la teoría de juegos, que viene a sumarse a la ya propia de la economía normativa. No hay una teoría de juegos, sino muchos modelos y formas de pensar sobre estrategia, y a cada uno parece corresponderle una posible ampliación de cada tema relevante en economía normativa. A esta multiplicación de posibles modos de modelar los incentivos en sistemas económicos se le añade un mayor grado de

complejidad técnica, e incluso notacional. Por ello es de prever que pasen bastantes años sin que tengamos de nuevo una formulación tan nítida como, por ejemplo, la de los resultados clásicos de Arrow, para una economía del bienestar que incorpore plenamente a los incentivos como parte integral de la discusión teórica y más aún para que puedan recibir una exposición didáctica y sencilla como la de Bator. Pero, cualquiera que sea el ritmo de avance, creemos que es en esta dirección en la que podrá incorporarse un nuevo nivel de discusión al ámbito teórico y así —esperemos— sustraer un tema económico más de manos de propagandistas de una u otra clase.

BIBLIOGRAFIA

- Arrow, K. J. (1951): «An Extension of the basic theorems of classical welfare economics», en Neyman (ed.), *Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability*, U. California Press.
- y Hurwicz, L. (1960): «Decentralization and Computation in resource allocation», en R. Pfouts (ed.), *Essays in Economics and Econometrics*, U. North Carolina Press.
- D'Aspremont, C., y Gérard-Varet, L. A. (1979): «Incentives and Incomplete Information», *Journal of Public Economics*.
- Barberá, S. (1977): «The manipulability of Social Choice mechanisms that do not leave "too much" to chance», *Econometrica*.
- (1979): «El papel del Estado en la asignación de recursos: algunas recomendaciones desde la Teoría Económica», en *El sector público en una economía de mercado*, Espasa-Calpe, Madrid.
- (1982): «Truthfulness and Decentralized Implementation», a aparecer en *Social Choice and Welfare*.
- (1983): «Strategy-proofness and pivotal voters: a direct proof of the Gibbard-Satterthwaite theorem», *International Economic Review*.
- Bator, F. (1957): «The simple analytics of Welfare Maximization», *American Economic Review*.
- (1958): «The anatomy of market failure», *The Quarterly Journal of Economics*.
- Batteau, P.; Blin, J. M., y Montjardet, B. (1981): «Stability of aggregation procedures, ultrafilters and simple games», *Econometrica*.
- Clarke, E. H. (1971): «Multipart pricing of public goods», *Public Choice*.
- Coase, R. (1960): «The problem of social cost», *Journal of Law and Economics*.
- Dasgupta, P. S., Hammond, P. J., y Maskin, E. (1979): «The implementation of social choice rules: some general results on incentive compatibility», *Review of Economic Studies*.
- Debreu, G. (1959): *Theory of Value*, Yale University Press.
- Dutta, B., y Pattanaik, P. K. (1979): «On nicely consistent voting schemes», *Econometrica*.
- Gibbard, A. (1973): «Manipulation of voting schemes: a general result», *Econometrica*.
- Green, J., y Laffont, J. J. (1979): *Incentives in public decision making*, North Holland.
- Groves, T. (1973): «Incentives in Teams», *Econometrica*.
- y Loeb, M. (1975): «Incentives and Public Inputs», *Journal of Public Economics*.
- y Ledyard, J. (1977): «Optimal allocation of public goods: a solution to the "free-rider" problem», *Econometrica*.
- Hurwicz, L. (1972): «On informationally decentralized systems», en Randner y Mc. Guire (eds.), *Decision and Organization*, North Holland.
- (1973): «The design of mechanisms for resource allocation», *American Economic Review*.
- (1979): «Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points», the *Review of Economic Studies*.
- y Schmeidler, D. (1978): «Construction of outcome functions guaranteeing existence and Pareto optimality of Nash equilibria», *Econometrica*.

- Isaac, R. M., McCue, K., y Plott, C. R. (1982): «Public goods provision in an experimental environment», *Social Science Working Paper* 428, California Institute of Technology.
- Journal of Economic Theory* (1980): Número monográfico sobre «Non-cooperative approaches to the theory of perfect competition».
- Journal of Mathematical Economics* (1982): Special Issue on Implementation.
- Kalai, E., y Muller, E. (1977): «Characterizations of Domains Admitting Nondictatorial Social Welfare Functions and Nonmanipulable Voting Procedures», *Journal of Economic Theory*.
- Kelly, J. (1978): *Arrow Impossibility Theorems*, Academic Press.
- Laffont, J. J. (ed.) (1979): *Aggregation and Revelation of Preferences*, North Holland.
- McKenzie, L. W. (1981): «The classical theorem on existence of competitive equilibrium», *Econometrica*.
- Marwell, G., y Ames, R. (1980): «Economists free-rider: does anybody else?», presentado a las reuniones de la Public Choice Society, San Francisco.
- MasColell, A. (1976): «En torno a una propiedad poco atractiva del equilibrio competitivo», *Moneda y Crédito*.
- (1980): «Efficiency and decentralization in the pure theory of public goods», *The Quarterly Journal of Economics*.
- Maskin, E.: *Implementation and strong Nash equilibrium*, en Laffont (1979).
- Milleron, J. C. (1972): «Theory of Value with public goods: a survey article», *Journal of Economic Theory*.
- Moulin, H. (1979): «Implementing efficient, anonymous and neutral social choice functions», *Journal of Mathematical Economics*.
- (1980): «Strategy proofness and single-peakedness», *Public Choice*.
- (1981): *Théorie des jeux pour l'économie et la politique*, Hermann.
- (1982): «Voting with proportional veto power», *Econometrica*.
- Ostroy, J. (1980): «The no-surplus condition as a characterization of perfectly competitive equilibrium», *Journal of Economic Theory*.
- Pattanaik, P. K. (1979): *Strategy and Group Choice*, North Holland.
- Peleg, B. (1978): «Consistent voting schemes», *Econometrica*.
- Postlewaite, A. (1979): «Manipulation via endowments», *Review of Economic Studies*.
- Public Choice* (1977): Special Supplement, Spring.
- Review of Economic Studies* (1979): Symposium on Incentive Compatibility.
- Roberts, J., y Postlewaite, A. (1976): «The incentives for pricetaking behaviour in large exchange economies», *Econometrica*.
- Roberts, K.: «The characterization of implementable choice rules», en Laffont (1979).
- Samuelson, P. A. (1954): «The pure theory of public expenditure», *Review of Economics and Statistics*.
- Satterthwaite, M. A. (1975): «Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions», *Journal of Economic Theory*.
- y Sonnenschein, H. (1981): «Strategy-proof mechanisms at differentiable points», *Review of Economic Studies*.
- Sen(gupta), M. (1983): «Implementable social choice rules: characterization and correspondence theorems under strong Nash equilibrium», *Journal of Mathematical Economics*.
- Schmeidler, D., y Sonnenschein, H. (1978): «Two proofs of the Gibbard-Satterthwaite theorem», en Leinfellner y Gottinger (eds.), *Decision Theory and Social Ethics*.
- Smith, V. (1980): «Experiments with a decentralized mechanism for public good decisions», *American Economic Review*.
- Tulkens, H. (1978): «Dynamic processes for public goods: an institution oriented survey», *Journal of Public Economics*.