

# DISEÑO DE MECANISMOS: CÓMO IMPLEMENTAR OBJETIVOS SOCIALES\*

**Eric S. Maskin\*\***

*Institute for Advanced Study, Princeton*

La teoría del diseño de mecanismos puede ser considerada como la parte "ingenieril" de la teoría económica. Por supuesto, gran parte del trabajo teórico se centra en las instituciones económicas *existentes*. El teórico quiere explicar o predecir los resultados económicos o sociales que generan estas instituciones. Ahora bien, en la teoría del diseño de mecanismos la dirección de la investigación se invierte. Comenzamos identificando nuestro resultado deseado u *objetivo social*. A continuación nos preguntamos si se podría diseñar o no una institución apropiada (mecanismo) para lograr ese objetivo. En el caso de que la respuesta sea afirmativa, tratamos de conocer la forma que podría tener tal mecanismo. En este trabajo, ofrezco una breve introducción a la parte del diseño de mecanismos denominada *teoría de la implementación*, que, dado un objetivo social, describe cuándo podemos diseñar un mecanismo cuyo resultados *predichos* (es decir, el conjunto de resultados de equilibrio) coincide con los resultados *deseables*, de acuerdo con ese objetivo.

*Palabras clave:* discurso Nobel, Eric S. Maskin, diseño de mecanismos, objetivos sociales, reglas de elección social, teoría de la implementación.

---

(\*) © Fundación Nobel 2007 (<http://www.nobelprize.org>). Este artículo es una versión revisada del discurso pronunciado por el profesor Eric S. Maskin en Estocolmo, el 8 de diciembre de 2007, cuando recibió el Premio en Ciencias Económicas del Banco de Suecia instituido en memoria de Alfred Nobel. El discurso se publica en *Revista Asturiana de Economía* con el consentimiento del autor y la autorización de la Fundación Nobel. La traducción ha sido realizada por Mario Piñera.

(\*\*) School of Social Science, Institute for Advanced Study. Profesor Albert O. Hirschman de Ciencias Sociales, Institute for Advanced Study, y profesor visitante, Princeton University. Se reconoce con gratitud la ayuda a la investigación recibida de la National Science Foundation vía la subvención SES-0318103.

## 1. INTRODUCCIÓN

La teoría del diseño de mecanismos puede ser considerada como la parte “ingenieril” de la teoría económica. Por supuesto, gran parte del trabajo teórico se centra en las instituciones económicas *existentes*. El teórico quiere explicar o predecir los resultados económicos o sociales que generan estas instituciones. Ahora bien, en la teoría del diseño de mecanismos la dirección de la investigación se invierte. Comenzamos identificando nuestro resultado deseado u *objetivo social*. A continuación nos preguntamos si se podría diseñar o no una institución apropiada (mecanismo) para lograr ese objetivo. En el caso de que la respuesta sea afirmativa, tratamos de conocer la forma que podría tener tal mecanismo.

En este trabajo, ofrezco una breve introducción a la parte del diseño de mecanismos denominada *teoría de la implementación*, que, dado un objetivo social, describe cuándo podemos diseñar un mecanismo cuyo resultados *predichos* (es decir, el conjunto de resultados de equilibrio) coincide con los resultados *deseables*, de acuerdo con ese objetivo. Trataré de mantener las cuestiones técnicas al mínimo y, en general, las confinaré a las notas<sup>1</sup>.

## 2. RESULTADOS, OBJETIVOS Y MECANISMOS

Lo que denominamos “resultado” dependerá lógicamente del contexto. Así, para un gobierno al que se le encomienda proporcionar bienes públicos, un resultado consistirá en las cantidades proporcionadas de bienes tales como autopistas interurbanas, defensa y seguridad nacional, protección medioambiental y enseñanza pública, junto con las disposiciones mediante las cuales son financiadas. Para un electorado que trata de cubrir un puesto político, un resultado es sencillamente la elección de un candidato para dicho puesto. Para un subastador que trata de vender una colección de activos, un resultado es la asignación de dichos activos entre los potenciales compradores, junto con los pagos que realizan dichos compradores. Finalmente, en el caso de un comprador de una casa y un constructor que están considerando la construcción de una nueva casa, un resultado es una especificación de las características de la casa y de la remuneración del constructor.

---

(1) Hay muchas revisiones de la literatura y manuales excelentes de la teoría de la implementación que entran en un detalle mucho mayor –tanto técnico como conceptual– del que presento aquí; véanse en especial: Postlewaite (1985), Groves y Ledyard (1987), Moore (1992), Palfrey (1992), capítulo 10 de Osborne y Rubinstein (1994), Allen (1997), Corchón (1996), Jackson (2001), Palfrey (2001), Serrano (2004), capítulos 2 y 3 de Austen-Smith y Banks (2005), capítulo 6 de Bergin (2005), capítulos 14-16 de Feldman y Serrano (2006), capítulo 10 de Rasmusen (2006), Baliga y Sjöström (2007) y Corchón (2008). Véanse también Dasgupta, Hammond y Maskin (1979), Maskin y Sjöström (2002), Baliga y Maskin (2003), y mi vieja revisión Maskin (1985).

De modo similar, el nivel con el que juzgamos la “deseabilidad” u “optimalidad” de un resultado dependerá también del entorno. Al evaluar las elecciones de bienes públicos, a menudo se invoca el criterio de maximización del “beneficio social neto”: ¿la decisión respecto a los bienes públicos maximiza el beneficio social bruto menos el coste de proveer los bienes? En lo que respecta a la elección de políticos, a menudo se considera como un desideratum normal la propiedad de que un candidato derrote a cada competidor en una competición cara a cara (esto es, surgiría un *ganador Condorcet*) (véase Dasgupta y Maskin, 2008). En la subasta de activos, hay dos criterios diferentes mediante los que se valora habitualmente un resultado: (i) si los activos terminan en las manos de los postores que los valoran más (es decir, si la asignación es *eficiente*); y alternativamente (ii) si el vendedor obtiene el mayor ingreso posible de las ventas (esto es, si se logra la *maximización del ingreso*). Finalmente, para el comprador y el constructor de la casa, un resultado será considerado generalmente “óptimo” si agota las ganancias potenciales del intercambio entre las partes, es decir, si la especificación de la casa y la remuneración son en su conjunto Pareto eficientes y racionales desde un punto de vista individual.

Un *mecanismo* es una institución, procedimiento, o juego para determinar resultados. Como es lógico, quién elige el mecanismo –esto es, quién lo diseña– es algo que dependerá, una vez más, del entorno. En el caso de los bienes públicos, normalmente pensamos que el gobierno que provee los bienes elige también el método mediante el cual se determinan los niveles de provisión y financiación. De forma similar, cuando se trata de ventas de activos –en los que la *subasta* es el mecanismo habitual– a menudo el vendedor del activo tiene la última palabra respecto a las reglas, es decir, es el único que elige el formato de la subasta.

En comparación, en el caso de las elecciones políticas nacionales un mecanismo es un *procedimiento electoral*, por ejemplo la regla de la mayoría relativa, dos vueltas, u otros similares. Además, normalmente el procedimiento se establece mucho antes, de hecho algunas veces en la constitución del país. Así, en este caso deberíamos pensar en los artífices de la constitución como los diseñadores del mecanismo.

Finalmente, en el ejemplo de la construcción de la casa, un mecanismo es un contrato entre el comprador de la casa y el constructor y recoge los derechos y responsabilidades de cada uno. En la medida en que es de suponer que dichas partes son las únicas que negocian este contrato, ellas mismas son las diseñadoras del mecanismo en este último caso.

Ahora, en el marco público, si el gobierno conoce desde el principio la elección del bien público que es óptima, entonces hay un mecanismo sencillo –de hecho, trivial– para lograr el óptimo: el gobierno solo tiene que aprobar una ley que ordene dicho resultado. De forma similar, si el subastador tiene conocimiento previo de los postores que valoran más los activos, puede sencillamente adjudicárselos directamente a dichos postores (con o sin pago).

La dificultad fundamental –que le da al tema del diseño de mecanismos su interés teórico– es que normalmente el gobierno o el subastador *no* cuentan con dicha información. Después de todo, la elección que maximiza el beneficio neto de los bienes públicos depende de las *preferencias* de los ciudadanos respecto a dichos bienes, y no hay ninguna razón particular por la que el gobierno debería conocer estas preferencias. De la misma manera, normalmente no esperaríamos que un subastador conociera en cuanto valoran los activos en venta los diferentes compradores potenciales de los mismos.

Debido a que, en general, los diseñadores de mecanismos no conocen de antemano cuáles son los resultados óptimos, tienen que avanzar de una forma más indirecta que la que correspondería simplemente a prescribir los resultados por decreto; en particular, los mecanismos diseñados deben generar la información necesaria conforme se va ejecutando. El problema se agrava por el hecho de que los individuos que *sí* tienen esta información fundamental – los ciudadanos en el caso del bien público o los compradores en el ejemplo de la venta del activo – tienen sus propios objetivos y consecuentemente es posible que no tengan el incentivo necesario para comportarse de una forma que revele lo que saben. Así, los mecanismos deben ser *compatibles con los incentivos*. Gran parte del trabajo correspondiente al diseño de mecanismos, incluyendo el mío, ha tratado de responder a tres preguntas básicas:

(A) ¿Cuándo es posible diseñar mecanismos compatibles con los incentivos para el logro de objetivos sociales?

(B) ¿Qué forma podrían tomar dichos mecanismos cuando existan?

y

(C) ¿Cuándo se descarta teóricamente la posibilidad de encontrar tales mecanismos?

El que sea posible diseñar tales mecanismos puede parecer, en principio, sorprendente. ¿Como, después de todo, puede un diseñador de mecanismos lograr un resultado óptimo sin conocer exactamente a lo que está apuntando? Por lo tanto, puede ser útil considerar un caso concreto sencillo.

### 3. UN EJEMPLO

Considérese una sociedad formada por dos consumidores de energía, Alice y Bob. Una autoridad relacionada con la energía se responsabiliza de elegir el tipo de energía que utilizarán Alice y Bob. La lista de posibilidades –de la que la autoridad debe elegir una– incluye gas, petróleo, energía nuclear y carbón.

Supongamos que hay dos estados posibles del mundo. En el estado 1, los consumidores le dan relativamente poco peso al futuro, es decir, tienen en términos comparativos unas tasas de descuento intertemporal elevadas. Por el contrario, en el estado 2 le conceden gran importancia al futuro, lo que significa que sus tasas de descuento son, consecuentemente, reducidas.

Supondremos que en lo que se refiere a la energía lo que le preocupa básicamente a Alice es su comodidad. Esto significa que, en el estado 1, ella clasificará al gas por encima del petróleo, al petróleo por encima del carbón y al carbón por encima de la energía nuclear, debido a que, conforme nos movemos hacia abajo por su clasificación, la fuente energética se convierte bien en más sucia o bien en más incómoda a la hora de utilizarla. Por el contrario, en el estado 2 su clasificación es

energía nuclear  
gas  
carbón  
petróleo

debido a que prevé que los avances técnicos llevarán finalmente a que el gas, el carbón y especialmente la energía nuclear sean más fáciles de utilizar –y en este estado, ella le da mayor importancia a los beneficios *futuros*.

Bob está interesado especialmente en la *seguridad*. Esto supone que en el estado 1, cuando le da mayor importancia al presente, prefiere la energía nuclear respecto al petróleo, el petróleo respecto al carbón, y el carbón respecto al gas. Pero si impera el estado 2 –y por tanto el futuro es en términos comparativos importante– su clasificación es:

petróleo  
gas  
carbón  
energía nuclear

lo que refleja el hecho de que, a largo plazo, cabe esperar que el problema de deshacerse de los residuos nucleares ocupe un lugar preponderante, mientras que la seguridad del petróleo y del gas es probable que mejore un tanto.

En resumen, las clasificaciones de los consumidores en los dos estados se muestran en el cuadro 1.

Supongamos que la autoridad relacionada con la energía está interesada en seleccionar una fuente energética con la que ambos consumidores estén razonablemente contentos. Si interpretamos “razonablemente contentos” en el sentido de lograr la primera o la segunda opción de cada uno, el petróleo es la elección óptima en el estado 1, mientras que el gas es el mejor resultado en el estado 2. En términos de la teoría de la implementación, decimos que la regla de elección social de la autoridad prescribe petróleo en el estado 1 y gas en el estado 2. Por ello, si  $f$  es la regla de elección social, ésta viene dada por el cuadro 2<sup>2</sup>.

---

(2) En un marco más general, en el que  $\Theta$  es el conjunto de estados posibles del mundo y  $A$  es el conjunto de resultados posibles, una regla de elección social  $f$  es una correspondencia (una función multivalorada)  $f: \Theta \rightarrow A$ , en la que, para cada  $\theta$ ,  $f(\theta)$  se interpreta como el conjunto de resultados óptimos en el estado  $\theta$  (estamos aceptando la posibilidad de que se pudiera considerar óptimo más de un resultado en un determinado estado).

**Cuadro 1**

Estado 1		Estado 2	
Alice	Bob	Alice	Bob
gas	energía nuclear	energía nuclear	petróleo
petróleo	petróleo	gas	gas
carbón	carbón	carbón	carbón
energía nuclear	gas	petróleo	energía nuclear

**Cuadro 2**

$f$ (estado 1) = petróleo	$f$ (estado 2) = gas
---------------------------	----------------------

Supongamos, sin embargo, que la autoridad no conoce el estado (aunque Alice y Bob sí lo conocen). Esto significa que no conoce la alternativa que prescribe la regla de elección social, es decir, si el óptimo es el petróleo o el gas.

Probablemente, el mecanismo más sencillo sería que la autoridad le pidiera a cada consumidor que anunciara el estado; y con dicha información eligiera petróleo si ambos consumidores dijese "estado 1", eligiera gas si ambos dijese "estado 2", y lo echara a cara o cruz entre ambos si obtuviera respuestas diferentes. Pero obsérvese que en este mecanismo Alice tiene el incentivo para decir "estado 2" con independencia del estado real y de lo que diga Bob, debido a que ella prefiere el gas al petróleo en ambos estados. En efecto, diciendo "estado 2" en vez de "estado 1", eleva la probabilidad de su resultado preferido desde 0 hasta 0,5, si Bob dice "estado 1", y desde 0,5 hasta 1, si Bob dice "estado 2". Por lo tanto, esperaríamos que Alice dijera "estado 2" en ambos estados. De forma similar, Bob diría siempre "estado 1", porque prefiere el petróleo al gas en ambos estados. Considerados en su conjunto, los comportamientos de Alice y Bob llevan a que, en cada estado, el resultado sea 50-50 en clave aleatoria entre el petróleo y el gas. Esto es, sólo hay una probabilidad del 50% de que el resultado sea óptimo, y por lo tanto este mecanismo es manifiestamente demasiado simple.

Supongamos, por lo tanto, que la autoridad consigue que los consumidores participen en el mecanismo dado por el cuadro 3:

**Cuadro 3**

		Bob	
		Izquierda	Derecha
Alice	Arriba	petróleo	carbón
	Abajo	energía nuclear	gas

Esto es, Alice elige su estrategia entre “Arriba” y “Abajo” y, simultáneamente, Bob la elige entre “Izquierda” y “Derecha”; y el resultado de dichas elecciones viene dado en la celda correspondiente de la matriz<sup>3</sup>.

Obsérvese que en el estado 1 lo mejor para Bob es elegir Izquierda, con independencia de lo que haga Alice: si ella opta por Arriba, Izquierda lleva al resultado petróleo (el que prefiere Bob), mientras que Derecha lleva a carbón. Si ella opta por Abajo, la consecuencia de optar por Izquierda es la energía nuclear (el resultado preferido por Bob), mientras que Derecha lleva a gas. Es decir, Izquierda es la “estrategia dominante” para Bob en el estado 1. Por otra parte, dado que Bob opta por Izquierda, lo mejor para Alice es elegir Arriba en vez de Abajo, porque prefiere el petróleo antes que la energía nuclear. Así, en el estado 1, la predicción evidente es que Alice optará por Arriba y Bob optará por Izquierda, es decir, (Arriba, izquierda) es el único equilibrio de Nash<sup>4</sup>. Además –y este es el punto fundamental– el resultado final, petróleo, es óptimo en el estado 1.

En lo que respecta al estado 2, vemos que Abajo es la estrategia dominante para Alice en dicho estado. Si Bob opta por Izquierda, entonces ella está mejor con Abajo que con Arriba porque prefiere la energía nuclear antes que el petróleo. Y si Bob opta por Derecha, entonces Abajo lleva a gas, lo que ella prefiere en comparación con el resultado de Arriba, carbón. Si Alice elige Abajo, Bob está mejor si opta por Derecha, debido a que prefiere el gas antes que la energía nuclear. Por lo tanto, en el estado 2, el (único) equilibrio de Nash es (Abajo, Derecha): Alice opta por Abajo y Bob opta por Derecha. Además, esto lleva al resultado óptimo, gas.

Hemos visto que, en cualquier estado, el mecanismo del cuadro 3 logra el resultado óptimo a pesar de que (i) el diseñador del mecanismo (la autoridad relacionada con la energía) ni siquiera conoce el estado real, y (ii) Alice y Bob sólo están interesados en sus propias preferencias, no en las de la autoridad. Más precisamente, debido a que los resultados del equilibrio de Nash del mecanismo del cuadro 3 coinciden con los resultados óptimos en cada estado, decimos que el meca-

(3) De forma más general, para una sociedad que tiene  $n$  individuos, un mecanismo es una correspondencia  $g: S_1 \times \dots \times S_n \rightarrow A$  en el que, para todo  $i$ ,  $S_i$  es el espacio de estrategias del individuo  $i$  y  $g(s_1, \dots, s_n)$  es el resultado recomendado por el mecanismo si los individuos realizan las estrategias  $(s_1, \dots, s_n)$ .

(4) En general, un equilibrio de Nash es una especificación de estrategias –una para cada individuo– respecto a las cuales ningún individuo tiene el incentivo de desviarse unilateralmente. Por ello, si  $u_i(a, \theta)$  es el pago del individuo  $i$  procedente del resultado  $a$  en el estado  $\theta$ , las estrategias  $(s_1, \dots, s_n)$  constituyen un equilibrio de Nash del mecanismo  $g$  en el estado  $\theta$  si  $u_i(g(s_1, \dots, s_i, \dots, s_n), \theta) \geq u_i(g(s_1, \dots, s'_i, \dots, s_n), \theta)$ , para todo  $i$  y todo  $s'_i \in S_i$ .

nismo *implementa* la regla de elección social de la autoridad en el equilibrio de Nash<sup>5, 6</sup>.

#### 4. UNA BREVE HISTORIA DEL DISEÑO DE MECANISMOS

La historia intelectual de la teoría del diseño de mecanismos se remonta al menos hasta los socialistas utópicos del siglo XIX tales como Robert Owen y Charles Fourier. Rechazando lo que ellos consideraban como los males del pujante sistema capitalista, estos pensadores señalaron que el socialismo ofrecía una alternativa más humana y algunas veces terminaron involucrados en el establecimiento de comunidades experimentales tales como *New Harmony*, en Indiana.

El Debate sobre la Planificación, que alcanzó su mayor intensidad en los años 30 del pasado siglo, tuvo una influencia más directa en la teoría moderna. Los principales contrincantes fueron, por una parte, Oskar Lange y Abba Lerner, que alegaron con energía que, realizada adecuadamente, la planificación central podría replicar el funcionamiento de los mercados libres (Lange, 1936; Lerner, 1944). En efecto, sugirieron que la planificación podría corregir graves “fallos de mercado” –en particular aquellos que se pusieron de manifiesto en la Gran Depresión– y de ese modo superar potencialmente al mercado. En la otra parte, Friedrich von Hayek y Ludwig von Mises negaron de manera incondicional la posibilidad de que un sistema planificado pudiera aproximarse alguna vez al éxito del libre mercado (von Hayek, 1944; von Mises, 1920).

El debate fue importante y fascinante, pero, para ciertos espectadores, tales como Leonid Hurwicz, fue también bastante frustrante. Ello fue así porque le faltó precisión conceptual: no se definieron términos fundamentales como “descentralización”. Además, las razones aducidas por ambas partes eran a menudo muy incompletas. En parte, esto fue así porque sencillamente carecían del aparato técnico –en particular, la teoría de juegos y la programación matemática– necesario para generar conclusiones realmente convincentes.

Este es el momento en el que Leo Hurwicz entra en escena. Inspirado por el debate, trató de proporcionar definiciones inequívocas de los conceptos centrales y dicho esfuerzo culminó en sus dos grandes trabajos,

---

(5) En un marco más general, el mecanismo  $g$  implementa la regla de elección social  $f$  en un equilibrio de Nash si  $f(\theta) = NE_g(\theta)$ , para todo  $\theta$ , donde  $NE_g(\theta)$  es el conjunto de resultados que son un equilibrio de Nash de  $g$  en el estado  $\theta$ .

(6) El equilibrio de Nash es una predicción respecto a cómo se comportaran los individuos en un mecanismo. Pero en la literatura de la implementación se han considerado otros conceptos predictivos –es decir, conceptos de equilibrio– entre ellos, el equilibrio perfecto en subjuegos (Moore y Repullo, 1988), el equilibrio de Nash no dominado (Palfrey y Srivastava, 1991), el equilibrio bayesiano (Postlewaite y Schmeidler, 1986), la resolución por dominación (Moulin, 1979), el equilibrio perfecto con mano temblorosa (Sjöström, 1993) y el equilibrio fuerte (Dutta y Sen, 1991).

Hurwicz (1960) y (1972), en los que introdujo también el concepto clave de compatibilidad de los incentivos.

El trabajo inspirado por Hurwicz y otros ha llevado a un amplio consenso entre los economistas respecto a que von Hayek y von Mises estaban, de hecho, en lo cierto –el mercado es el “mejor” mecanismo– en entornos en los que (i) hay un gran número de vendedores y compradores, con lo que ningún agente individual tiene un poder de mercado importante; y (ii) no hay externalidades importantes, esto es, el consumo, la producción y la información de un agente no afecta a la producción y al consumo de los demás<sup>7</sup>. No obstante, si se viola algún supuesto, existe en general la posibilidad de contar con mecanismos que sean mejores que el mercado<sup>8</sup>.

La investigación de Hurwicz ocasionó una enorme literatura, que en gran parte se ramificó en dos direcciones diferentes. Por un lado, están los estudios que utilizan entornos especiales, altamente estructurados, para estudiar cuestiones particulares tales como de qué forma se pueden asignar los bienes públicos, cómo diseñar subastas, y cómo estructurar contratos. Por el otro lado, hay estudios que obtienen resultados a un nivel general, abstracto, esto es, en ellos se plantean los menos supuestos posibles en lo que se refiere a preferencias, tecnologías, etcétera. Mi propia investigación se ha situado en ambas categorías en diferentes momentos. No obstante, en este trabajo haré hincapié en los resultados generales.

## 5. IMPLEMENTACIÓN DE REGLAS DE ELECCIÓN SOCIAL

Previamente, señalé tres preguntas centrales (A)–(C) respecto a los mecanismos compatibles con los incentivos. Reformuladas en el lenguaje de la teoría de la implementación, estas preguntas se convierten en:

(A') ¿Bajo qué condiciones se puede implementar una regla de elección social?

(B') ¿Qué forma toma un mecanismo de implementación?

(C') ¿Qué reglas de elección social no se pueden implementar?

A mediados de los setenta del pasado siglo me enfrenté a estas preguntas. Al final descubrí que una propiedad llamada *monotonía* (denomi-

(7) Véanse, por ejemplo, Hammond (1979) –quien demuestra que, cuando el número de consumidores tiende a infinito, el mercado competitivo es el único mecanismo compatible con los incentivos que produce resultados racionales desde un punto de vista individual y óptimos de Pareto– y Jordan (1982) –quien muestra lo mismo cuando “la compatibilidad respecto a los incentivos” se sustituye por la “eficiencia respecto a la información”, bajo los supuestos (i) y (ii).

(8) Véanse, por ejemplo, Groves (1973) y Clarke (1971) para el caso de los bienes públicos y Laffont (1985) para el caso de las externalidades informativas.

nada ahora a veces monotónía de Maskin) es la clave para la implementación en el equilibrio de Nash. Supóngase que un resultado  $a$  es óptimo en el estado  $\theta$  (de acuerdo con la regla de elección social  $f$  en cuestión, esto es,  $f(\theta) = a$ ). En ese caso, si  $a$  no desciende en la clasificación de alguien en comparación con cualquier otra alternativa, al pasar desde el estado  $\theta$  al estado  $\theta'$ , la monotónía exige que  $a$  sea también óptimo en el estado  $\theta'$ :  $f(\theta') = a$ . Sin embargo, cuando  $a$  sí desciende en relación con algún resultado  $b$  en la clasificación de alguien, la monotónía no impone ninguna restricción<sup>9</sup>.

Para ver lo que significa la monotónía de una forma más concreta, consideremos nuestro ejemplo anterior relacionado con la energía (véanse los cuadros 1 y 2). Recuérdese que el petróleo es el resultado óptimo en el estado 1. Obsérvese también que el petróleo *desciende* en la clasificación de Alice, tanto respecto al carbón como a la energía nuclear, al pasar del estado 1 al estado 2 (Alice sitúa al petróleo por encima del carbón y la energía nuclear en el estado 1, pero ocurre justamente lo contrario en el estado 2). Así, el hecho de que el gas –no el petróleo– es óptimo en el estado 2 no viola la monotónía. De forma similar, obsérvese que el gas desciende en la clasificación de Bob, tanto respecto al carbón como a la energía nuclear, al pasar del estado 2 al estado 1. Por ello, a pesar de que el gas es óptimo en el estado 2, el hecho de que no es óptimo en el estado 1 tampoco está en conflicto con la monotónía. En efecto, estas comprobaciones establecen que la regla de elección social de la autoridad cumple la monotónía (y por lo tanto la posibilidad de implementarla, que se mostró más arriba, no contradice el Teorema 1 que se expone más abajo).

Ahora bien, supóngase que modificamos un poco el ejemplo, de modo que las clasificaciones y los resultados óptimos sean los dados por el cuadro 4.

**Cuadro 4**

Estado 1		Estado 2	
Alice	Bob	Alice	Bob
gas	energía nuclear	gas	energía nuclear
petróleo	petróleo	petróleo	petróleo
carbón	carbón	energía nuclear	carbón
energía nuclear	gas	carbón	gas
petróleo óptimo		energía nuclear óptima	

(9) En un marco más general en el que  $f$  puede ser multivalorada, la monotónía exige que, para todos los estados,  $\theta, \theta'$  y todos los resultados  $a$ , si  $a \in f(\theta)$  y  $u_i(a, \theta) \geq u_i(b, \theta)$  implica que  $u_i(a, \theta') \geq u_i(b, \theta')$  para todo  $i$  y  $b$ , entonces  $a \in f(\theta')$ .

Con estos cambios, la regla de elección social ya no es monótona. Concretamente, obsérvese que si bien el petróleo es óptimo en el estado 1, no lo es en el estado 2, a pesar de que entre los estados 1 y 2 no desciende ni en la clasificación de Alice ni en la de Bob (dado que el petróleo no desciende, la monotonía exigiría que siguiera siendo óptimo en el estado 2). Por lo tanto, podemos concluir que *no* hay un mecanismo que implemente la regla de elección social del cuadro 4. De forma más general, tenemos:

**TEOREMA 1 (Maskin, 1977).** *Si una regla de elección social se puede implementar, entonces debe ser monótona.*

Para ver por qué la regla de elección social del cuadro 4 no se puede implementar, supóngase por el contrario que *hubiera* un mecanismo de implementación. Entonces, en particular, el mecanismo contendría necesariamente un par de estrategias  $(s_A, s_B)$  –para Alice y Bob, respectivamente– que llevarían al resultado petróleo y constituirían un equilibrio de Nash en el estado 1.

Nótese que  $(s_A, s_B)$  deben constituir también un equilibrio de Nash en el estado 2. Para entender esta afirmación, obsérvese en primer lugar que Bob no tiene incentivo para desviarse unilateralmente de  $s_B$  en el estado 2, dado que (i) no tiene tal incentivo en el estado 1 (por la definición de equilibrio de Nash) y (ii) el orden de sus preferencias es el mismo en ambos estados. Además, Alice no tiene incentivo para desviarse de  $s_A$  en el estado 2. Para ver esto, obsérvese que si, en contra de la afirmación, Alice ganara al desviarse unilateralmente de  $s_A$  en el estado 2, debería estar provocando de ese modo el resultado gas (debido a que es el único resultado que prefiere respecto al petróleo en el estado 2). Pero Alice prefiere también el gas respecto al petróleo en el estado 1, y consecuentemente se beneficiaría de la misma desviación en ese estado, contradiciendo el supuesto de que  $(s_A, s_B)$  constituye un equilibrio de Nash en el estado 1.

Por lo tanto,  $(s_A, s_B)$  es en efecto un equilibrio de Nash en el estado 2. Pero el resultado que genera –petróleo– no es óptimo en dicho estado, demostrando que al final la regla de elección social no es implementable.

Como hemos visto, los cuadros 1 y 2 nos proporcionan un ejemplo de una regla de elección social que es monótona y además implementable. No obstante, no es cierto que *todas* las reglas de elección social monótonas sean implementables; véase Maskin (1977) para un contraejemplo. Sin embargo, tales contraejemplos son bastante artificiosos y, si se impone una condición adicional, a menudo inocua, la monotonía *sí* garantiza la implementación, si hay al menos tres individuos en la sociedad<sup>10</sup>.

(10) Esto no significa que la implementación sea imposible con sólo dos individuos –de hecho, nuestro ejemplo relacionado con la energía de los cuadros 1 y 2 sólo tenía dos individuos. Sin embargo, como veremos más abajo, la implementación se facilita cuando hay tres o más individuos.

La condición adicional se denomina *ausencia de poder de veto*. Supóngase que todos los individuos, excepto quizás uno, están de acuerdo en que un resultado concreto *a* es *el mejor*, en el sentido de que todos ellos ponen dicho resultado *a* en lo más *alto* de su orden de preferencias. En ese caso, si la regla de elección social cumple la condición de ausencia de poder de veto, *a* debe ser óptimo. En otras palabras, el individuo restante no puede “vetarlo”.

El que no haya poder de veto es especialmente inocuo –en efecto, no impone ninguna restricción en absoluto– cuando las preferencias son estrictamente crecientes en los bienes consumidos. En ese caso, cada individuo preferirá una porción mayor de dichos bienes para sí mismo. Así, dos de ellos no pueden estar de acuerdo en que un determinado resultado *a* es el mejor: ambos no pueden tener la porción más grande. Esto significa que, si hay tres o más individuos, la hipótesis postulada por la condición de que no hay poder de veto *no se puede cumplir*, y por lo tanto lógicamente la condición se cumple *automáticamente*.

Un resultado general sobre la posibilidad de implementar reglas de elección social es el siguiente:

**TEOREMA 2 (Maskin, 1977).** *Supóngase que hay al menos tres individuos. Si la regla de elección social cumple la condición de monotonía y no hay poder de veto, entonces se puede implementar.*

Las demostraciones del Teorema 2 se escapan del ámbito de este trabajo (véase Repullo, 1987, para un razonamiento especialmente elegante), pero debería mencionar que por lo general son *constructivas*. Esto es, dada la regla de elección social que se debe implementar, una demostración plantea una fórmula explícita para la construcción de un mecanismo que realice el truco.

Merece la pena señalar por qué el Teorema 2 postula al menos tres individuos. A menudo, en la economía, el pasar de dos a tres personas hace que las cosas sean más difíciles<sup>11</sup>. Sin embargo, en el caso de la teoría de la implementación, en realidad tres individuos hacen que las cosas sean más fáciles. Para comprenderlo, recuérdese que la idea subyacente tras un mecanismo es darle al individuo el incentivo para comportarse de una manera que garantice un resultado óptimo. Esto supone “sancionar” a un individuo por desviarse de la estrategia que se le ha recomendado (es decir, de equilibrio). Ahora bien, si sólo hay dos individuos, Alice y Bob, y uno de ellos se ha desviado, puede ser difícil determinar si fue Alice quien se desvió y si Bob cumplió o viceversa. Este problema de identificación se resuelve una vez que hay tres personas: el que se desvía sobresale más claramente cuando dos o más de los individuos restantes están cumpliendo el equilibrio.

---

(11) Los juegos de suma cero proporcionan un ejemplo clásico de este fenómeno. El teorema del minimax –que simplifica enormemente el análisis del comportamiento en los juegos– se aplica a juegos de suma cero de *dos* personas, pero no, en general, al caso de *tres* o más jugadores.

## 6. COMENTARIOS FINALES

Esta presentación solo ha sido una muy breve introducción a la teoría de la implementación (que en sí misma constituye sólo una parte del campo del diseño de mecanismos). Me he centrado en la investigación que se hizo hace treinta años, lo que quizás le dé un sabor engañosamente "antiguo" al trabajo. En realidad, un aspecto especialmente gratificante de la teoría es que, casi cincuenta años después del artículo de Hurwicz (1960), el tema sigue siendo vibrante e importante desde un punto de vista intelectual: aparecen continuamente nuevos trabajos sobre la implementación. Será interesante ver hacia dónde va el campo en los próximos cincuenta años.

## REFERENCIAS BIBLIOGRÁFICAS

- Allen, Beth (1997): "Implementation Theory with Incomplete Information", en Hart, S. y Mas-Colell, A. (eds.), *Cooperation: Game Theoretic Approaches*, Springer, Berlin.
- Austen-Smith, David y Banks, Jeffrey (2005): *Positive Political Theory II: Strategy and Structure*, University of Michigan Press, Ann Arbor.
- Baliga, Sandeep y Maskin, Eric S. (2003): "Mechanism Design for the Environment", en Måler, K. G. y Vincent, J. R. (eds.), *Handbook of Environmental Economics*, vol. 1, North-Holland, Amsterdam, pp. 305-324.
- Baliga, Sandeep y Sjöström, Tomas (2007): "Mechanism Design: Recent Developments", en Blume, L. y Durlauf, S. (eds.), *The New Palgrave Dictionary of Economics*, 2ª edición, Palgrave MacMillan, Londres.
- Bergin, James (2005): *Microeconomic Theory: A Concise Course*, Oxford University Press, Oxford.
- Clarke, Edward (1971): "Multipart Pricing of Public Goods", *Public Choice*, vol. 11, nº 1, pp. 19-33.
- Corchón, Luis (1996): *The Theory of Implementation of Socially Optimal Decisions in Economics*, Macmillan, Londres.
- Corchón, Luis (2009): "The Theory of Implementation", en Meyers, Robert A. (ed.), *The Encyclopedia of Complexity and System Science*, Springer, Berlín.
- Dasgupta, Partha; Hammond, Peter y Maskin, Eric S. (1979): "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility", *Review of Economic Studies*, vol. 46, nº 2, pp. 185-216.
- Dasgupta, Partha y Maskin, Eric S. (2008): "On the Robustness of Majority Rule", *Journal of the European Economic Association*, vol. 6, nº 5, pp. 949-973.

- Dutta, Bhaskar y Sen, Arunava (1991): "Implementation under Strong Equilibrium: A Complete Characterization", *Journal of Mathematical Economics*, vol. 20, n° 1, pp. 46-67.
- Feldman, Allan y Serrano, Roberto (2006): *Welfare Economics and Social Choice Theory*, Springer, Berlín.
- Groves, Theodore (1973): "Incentives in Teams", *Econometrica*, vol. 41, n° 4, pp. 617-631.
- Groves, Theodore y Ledyard, John (1987): "Incentive Compatibility since 1972", en Groves, T.; Radner, R. y Reiter, S. (eds.), *Information, Incentives and Economic Mechanisms: Essays in Honour of Leonid Hurwicz*, University of Minnesota Press, Minneapolis, pp. 48-111.
- Hammond, Peter (1979): "Straightforward Individual Incentive Compatibility in Large Economies", *Review of Economic Studies*, vol. 46, n° 2, pp. 263-282.
- Hayek, Friedrich A. von (1944): *The Road to Serfdom*, Routledge, Londres.
- Hurwicz, Leonid (1960): "Optimality and Informational Efficiency in Resource Allocation Processes", en Arrow, Kenneth J.; Karlin, S. y Suppes, P. (eds.), *Mathematical Methods in Social Sciences*, Stanford University Press, Stanford, pp. 27-46.
- Hurwicz, Leonid (1972): "On Informationally Decentralized Systems", en McGuire, C. B. y Radner, Roy (eds.), *Decision and Organization: a Volume in Honor of Jacob Marshak*, North-Holland, Amsterdam, pp. 297-336.
- Jackson, Matthew (2001): "A Crash Course in Implementation Theory", *Social Choice and Welfare*, vol. 18, n° 4, pp. 655-708.
- Jordan, James (1982): "The Competitive Allocation Process is Informationally Efficient Uniquely", *Journal of Economic Theory*, vol. 28, n° 1, pp. 1-18.
- Laffont, Jean-Jacques (1985): "On the Welfare Analysis of Rational Expectations Equilibria with Asymmetric Information", *Econometrica*, vol. 53, n° 1, pp. 1-29.
- Lange, Oskar (1936): "On the Economic Theory of Socialism", *Review of Economic Studies*, vol. 4, pp. 53-71.
- Lerner, Abba (1944): *The Economics of Control*, McMillan, Nueva York.
- Maskin, Eric S. (1977): "Nash equilibrium and welfare optimality", *Review of Economic Studies*, vol. 69, n° 1, 1999, pp. 23-38.
- Maskin, Eric S. (1985): "The Theory of Implementation in Nash Equilibrium: A Survey", en Hurwicz, Leonid; Schmeidler, David y Sonnenschein, Hugo (eds.), *Social Goals and Social Organization, Essays in Memory of Elisha Pazner*, Cambridge University Press, Cambridge, pp. 173-204.

- Maskin, Eric S. y Sjöström, Tomas (2002): "Implementation Theory", en Arrow, K. J. ; Sen, A. y Suzumura, K. (eds.), *Handbook of Social Choice and Welfare*, vol. 1, Elsevier, Amsterdam, pp. 237-288.
- Mises, Ludwig von (1920): "Die Wirtschaftsrechnung im Sozialistischen Gemeinwesen", *Archiv für Sozialwissenschaften*, vol. 47 (traducción al inglés en Hayek, Friedrich A. von (ed.), *Collectivist Economic Planning*, Routledge, Londres, 1935).
- Moore, John (1992): "Implementation, Contracts, and Renegotiation in Environments with Complete Information", en Laffont, J. J. (ed.), *Advances in Economic Theory*, vol. 1, Cambridge University Press, Cambridge, pp. 182-282.
- Moore, John y Repullo, Rafael (1988): "Subgame Perfect Implementation", *Econometrica*, vol. 56, n° 5, pp. 1191-1220.
- Moulin, Hervé (1979): "Dominance Solvable Voting Schemes", *Econometrica*, vol. 47, n° 6, pp. 1337-1351.
- Osborne, Martin y Rubinstein, Ariel (1994): *A Course in Game Theory*, MIT Press, Cambridge.
- Palfrey, Thomas R. (1992): "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design", en Laffont, J. J. (ed.), *Advances in Economic Theory*, vol. 1, Cambridge University Press, Cambridge, pp. 283-323.
- Palfrey, Thomas R. (2001): "Implementation Theory", en Aumann R. y Hart, S. (eds.), *Handbook of Game Theory*, vol. 3, North-Holland, Amsterdam, pp. 2271-2326.
- Palfrey, Thomas R. y Srivastava, Sanjay (1991): "Nash Implementation using Undominated Strategies", *Econometrica*, vol. 59, n° 2, pp. 479-501.
- Postlewaite, Andrew (1985): "Implementation via Nash Equilibria in Economic Environments", en Hurwicz, Leonid; Schmeidler, David y Sonnenschein, Hugo (eds.), *Social Goals and Social Organization, Essays in Memory of Elisha Pazner*, Cambridge University Press, Cambridge, pp. 205-228.
- Postlewaite, A. y Schmeidler, David (1986): "Implementation in Differential Information Economies", *Journal of Economic Theory*, vol. 39, n° 1, pp. 14-33.
- Rasmusen, Eric (2006): *Games and Information: An Introduction to Game Theory*, Blackwell Publishing, Oxford.
- Repullo, Rafael (1987): "A Simple Proof of Maskin's Theorem on Nash Implementation", *Social Choice and Welfare*, vol. 4, n° 1, pp. 39-41.
- Serrano, Roberto (2004): "The Theory of Implementation of Social Choice Rules", *SIAM Review*, vol. 46, pp. 377-414.
- Sjöström, Tomas (1993): "Implementation in Perfect Equilibria", *Social Choice and Welfare*, vol. 10, n° 1, pp. 97-106.

### ABSTRACT

The theory of mechanism design can be thought of as the “engineering” side of economic theory. Much theoretical work, of course, focuses on *existing* economic institutions. The theorist wants to explain or forecast the economic or social outcomes that these institutions generate. But in mechanism design theory the direction of inquiry is reversed. We begin by identifying our desired outcome or *social goal*. We then ask whether or not an appropriate institution (mechanism) could be designed to attain that goal. If the answer is yes, then we want to know what form that mechanism might take. In this paper, I offer a brief introduction to the part of mechanism design called *implementation theory*, which, given a social goal, characterizes when we can design a mechanism whose *predicted* outcomes (i.e., the set of equilibrium outcomes) coincide with the *desirable* outcomes, according to that goal.

*Key words:* Nobel Lecture, Eric S. Maskin, Mechanism Design, Social Goals, Social Choice Rules, Implementation Theory.