

Diseño de mecanismos: cómo implementar objetivos sociales*•

*Eric S. Maskin***

Escuela de Ciencia Social,

Instituto para el Estudio Avanzado, Princeton.

La teoría de diseño de mecanismos puede ser pensada como el lado “ingenieril” de la teoría económica. Mucho trabajo teórico, por supuesto, se focaliza sobre las instituciones económicas existentes. El teórico quiere explicar o prever los resultados sociales o económicos que estas instituciones generan. Pero en la teoría de diseño de mecanismos la dirección de la investigación se revierte. Comenzamos por identificar nuestro resultado u objetivo social deseado. Luego nos preguntamos si una institución (mecanismo) apropiada puede ser o no diseñada para alcanzar ese resultado. Si la respuesta es sí, entonces queremos saber qué forma podría tomar ese mecanismo.

En este trabajo, ofrezco una breve introducción a la parte de diseño de mecanismos llamada *teoría de implementación*, la cual, dado un objetivo social, caracteriza cuando podemos diseñar un mecanismo cuyos resultados predichos (esto es, el conjunto de resultados de equilibrio) coinciden con los resultados deseables, de acuerdo con ese objetivo. Trato de mantener los tecnicismos al mínimo y, en general, los confino a las notas al pie¹.

* Este artículo es una versión revisada de la Conferencia Nóbel de Eric Maskin pronunciada el 8 de diciembre de 2007 en Estocolmo.

• “Mechanism Design: How to Implement Social Goals” © The Nobel Foundation 2007. Traducción al español de Bernardo Díaz de Astarloa.

** Profesor Albert O. Hirschman de Ciencias Sociales, Instituto para el Estudio Avanzado y Profesor Visitante, Universidad de Princeton. Se agradece el apoyo de la beca de investigación #SES-0318103 de la NSF.

1. Hay muchas revisiones y tratamientos de libro de texto de la teoría de la implementación que entran considerablemente en más detalle –tanto técnico como conceptual– de lo que yo hago aquí; véase en particular: Andrew Postlewaite (1985), Theodore Groves y John Ledyard (1987), John Moore (1992), Thomas Palfrey (1992), capítulo 10 de Martin Osborne y Ariel Rubinstein (1994), Beth Allen (1997), Luis Corchon (1996), Matthew Jackson (2001), Palfrey (2002), Roberto Serrano (2004), capítulos 2 y 3 de David Austen-Smith y Jeffrey Banks (2005), capítulo 6 de James Bergin (2005), capítulos 14-16 de Allan Feldman y R. Serrano (2006), capítulo 10 de Eric Rasmusen (2006), Sandeep Baliga y Tomas Sjöström (2007) y Corchon (2008). Véase también Partha Dasgupta, Peter Hammond, y Eric Maskin (1979), Maskin y Sjöström (2002), Baliga y Maskin (2003), y mi vieja revisión Maskin (1987).

I. Resultados, objetivos y mecanismos

Lo que queremos decir por un “resultado” dependerá naturalmente del contexto. Así, para un gobierno encargado de proveer bienes públicos, un resultado consistirá en las cantidades provistas de tales bienes, como autopistas interurbanas, defensa y seguridad nacional, protección ambiental y educación pública, junto con las disposiciones mediante las cuales son financiadas. Para un electorado que busca ocupar un cargo político, un resultado es simplemente la elección de un candidato para ese cargo. Para un subastador que vende una colección de activos, un resultado corresponde a una asignación de estos activos entre potenciales compradores, junto con los pagos que estos compradores realizan. Finalmente, en el caso de un comprador de una casa y un constructor que contemplan la construcción de una nueva casa, un resultado es una especificación de las características de la casa y la remuneración del constructor.

De manera similar, los estándares por los cuales juzgamos la “deseabilidad” u “optimalidad” de un resultado también dependerán del escenario. Al evaluar las elecciones de bienes públicos, se invoca frecuentemente el criterio de maximización del “superávit social neto”: ¿la decisión de bienes públicos maximiza el beneficio social bruto menos el costo de proveer los bienes? En cuanto a elegir políticos, la propiedad de que un candidato derrotaría a cada competidor en una competencia cabeza a cabeza (es decir, que resultaría un *ganador Condorcet*) es vista a veces como un desiderátum natural (véase Partha Dasgupta y Eric Maskin, próximamente). En la subasta de activos, hay dos criterios diferentes mediante los cuales un resultado es típicamente juzgado: (i) si los activos son puestos en las manos de los oferentes que los valoran más (es decir, si la asignación es *eficiente*) y alternativamente (ii) si el vendedor recauda el mayor ingreso posible de las ventas (es decir, si se alcanza la *maximización del ingreso*). Finalmente, para el comprador y el constructor de la casa, un resultado será normalmente considerado óptimo si agota las potenciales ganancias de intercambio entre las partes, es decir, la especificación de la casa y la remuneración son ambas Pareto óptimas e individualmente racionales.

Un *mecanismo* es una institución, un procedimiento o juego para determinar los resultados. No es sorprendente, entonces, que quien podrá elegir el mecanismo –esto es, quién será el diseñador del mecanismo– dependerá, nuevamente, del entorno. En el caso de los bienes públicos, normalmente pensamos en el gobierno proveyendo los bienes públicos y al mismo tiempo eligiendo el método mediante el cual se determinan los niveles de provisión y financiamiento. De manera similar, cuando hablamos de ventas de activos –donde una *subasta* es el mecanismo típico– el vendedor del activo es usualmente quien tiene el poder sobre las reglas, es decir, quien elige el formato de la subasta.

En el caso de las elecciones políticas nacionales, por contraste, un mecanismo es un *procedimiento electoral*, por ejemplo, mayoría relativa, segunda vuelta, o similares. Además, el procedimiento está normalmente prescripto mucho tiempo por adelantado; de hecho, algunas veces en la constitución del país. Así, aquí deberíamos pensar en los que formularon 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 dispone los derechos y responsabilidades de cada uno. Ya que estos grupos son presumiblemente los que negocian este contrato, son

ellos mismos los diseñadores del mecanismo en este último escenario.

Ahora, en el marco público, si el gobierno sabe al comienzo qué elección de bienes públicos es óptima, entonces hay un mecanismo simple –de hecho, trivial– para alcanzar el óptimo: el gobierno sólo tiene que aprobar una ley que implemente este resultado. De manera similar, si el subastador tiene un conocimiento previo de qué oferentes más valoran los activos, él puede simplemente concederlos directamente a estos oferentes (con un pago o sin este).

La dificultad básica –que da al diseño de mecanismos su interés teórico– es que el gobierno o el subastador típicamente no tendrán esta información. Después de todo, la elección de bienes públicos que maximiza el superávit neto depende de las *preferencias* de los ciudadanos sobre tales bienes, y no hay ninguna razón particular por la cual el gobierno deba saber estas preferencias. De la misma manera, normalmente no esperaríamos que un subastador supiera cuanto valoran los activos los diferentes compradores potenciales.

Porque los diseñadores de mecanismos por lo general no conocen de antemano qué resultados son óptimos, tienen que proceder más indirectamente que solo prescribiendo resultados por mandato; en particular, los mecanismos diseñados deben generar la información necesaria a medida que se ejecutan. El problema es exacerbado por el hecho de que los individuos que sí tienen esa información crítica –los ciudadanos en el caso de los bienes públicos o los compradores en el ejemplo de la venta de activos– tienen sus propios objetivos y, entonces, pueden no tener el incentivo por comportarse de modo que revele lo que saben. Entonces, los mecanismos deben ser incentivo-compatibles. Mucho del trabajo en diseño de mecanismos, incluyendo el mío, ha estado dirigido a responder tres preguntas básicas:

- (A) ¿Cuándo es posible diseñar mecanismos incentivo-compatibles para alcanzar objetivos sociales?
- (B) ¿Qué forma podrían tomar estos mecanismos cuando existan?
- (C) ¿Cuándo encontrar tales mecanismos es descartado teóricamente?

Que de hecho sea posible diseñar tales mecanismos puede, en principio, parecer sorprendente. ¿Cómo, después de todo, puede un diseñador de mecanismos alcanzar un resultado óptimo sin saber exactamente qué es a lo que está apuntando? Así, puede ser de ayuda considerar un ejemplo simple y concreto.

II. Un ejemplo

Consideremos una sociedad que consiste de dos consumidores de energía, Alice y Bob. Una autoridad energética está encargada de elegir el tipo de energía que será usada por Alice y Bob. La lista de opciones –de la cual la autoridad debe hacer una única selección– son gas, petróleo, energía nuclear y carbón.

Supongamos que hay dos estados posibles del mundo. En el estado 1, los consumidores colocan relativamente menos peso sobre el futuro; es decir, tienen tasas de descuento temporales comparativamente más altas. En el estado 2, por el contrario, le dan bastante importancia al futuro, lo que significa que sus tasas de descuento son en correspondencia más bajas.

Alice, imaginaremos, está preocupada sobre todo por la conveniencia cuando se trata de energía. Esto significa que, en el estado 1, preferirá el gas al petróleo, el petróleo al carbón y el carbón a la energía nuclear, porque a medida que nos movemos hacia abajo en su *ranking*, la fuente de energía se vuelve o más sucia o más engorrosa de usar. En el estado 2, por contraste, su *ranking* es

- energía nuclear
- gas
- carbón
- petróleo

porque ella anticipa que las mejoras técnicas eventualmente harán que el gas, el carbón y especialmente la energía nuclear sean más fáciles de utilizar –y, en este estado, le da una importancia particular a los beneficios *futuros*.

Bob está interesado particularmente en la *seguridad*. Esto implica que en el estado 1, cuando pone más peso sobre el presente, prefiere la energía nuclear al petróleo, el petróleo al carbón y el carbón al gas. Pero si se obtiene el estado 2 –de modo que el futuro es comparativamente más importante– su *ranking* es:

- petróleo
- gas
- carbón
- energía nuclear

lo que refleja el hecho de que, en el largo plazo, puede esperarse que el problema de des- echar el residuo nuclear sea dominante, pero es probable que la seguridad en el petróleo y el gas mejore un tanto.

Para resumir, los *rankings* de los consumidores en los dos estados se presentan en la Tabla 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 |

Asumamos que la autoridad está interesada en seleccionar una fuente de energía con la que ambos consumidores están razonablemente felices. Si interpretamos “razonablemente felices” como alcanzar la primera o la segunda opción de cada uno, entonces petróleo es la elección óptima en el estado 1, mientras que gas es el mejor resultado en el estado 2. En el lenguaje de la teoría de implementación, decimos que la *regla de elección social* de la autoridad prescribe petróleo en el estado 1 y gas en el estado 2. Así, si f es la regla de elección social, esta viene dada por la Tabla 2².

2. En un planteo más general, donde Θ es el conjunto de posibles estados del mundo y A es el conjunto de resultados posibles, una regla de elección social f es una correspondencia (una función de conjuntos) $f: \Theta \rightarrow A$, donde, para cualquier θ , $f(\theta)$ se interpreta como el conjunto de resultados óptimos en el estado θ (permitimos la posibilidad de que más de un resultado pueda considerarse óptimo en un estado dado).

$$f(\text{estado 1}) = \text{petróleo} \quad f(\text{estado 2}) = \text{gas}$$

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

Probablemente el mecanismo más sencillo sería que la autoridad le pidiera a cada consumidor que anunciase el estado, a partir del cual elegiría petróleo si ambos consumidores dijeren “estado 1”, gas si ambos dijeren “estado 2” y arrojaría una moneda para decidir si obtuviese una respuesta mixta. Pero nótese que en este mecanismo Alice tiene el incentivo a decir “estado 2” más allá del estado real y más allá de lo que Bob diga, porque prefiere el gas al petróleo en los dos estados. En efecto, diciendo “estado 2” en lugar de “estado 1” aumenta la probabilidad de su resultado preferido, de 0 a 0,5 si Bob dice “estado 1” y de 0,5 a 1 si Bob dice “estado 2”. En consecuencia, esperaríamos que Alice reportara “estado 2” en ambos estados. De manera similar, Bob siempre reportaría “estado 1”, porque prefiere el petróleo al gas en cualquier estado. Tomándolos juntos, los comportamientos de Alice y Bob implican que, en cada estado, el resultado es una randomización 50-50 entre petróleo y gas. Es decir, hay sólo una probabilidad de 50% de que el resultado sea óptimo, y entonces este mecanismo es manifiestamente demasiado simple.

Supongamos, entonces, que la autoridad hace a los consumidores participar del mecanismo dado por la Tabla 3:

| | | | |
|-------|--------|-----------------|---------|
| | | Bob | |
| | | Izquierda | Derecha |
| Alice | Arriba | Petróleo | Carbón |
| | Abajo | Energía nuclear | Gas |

Esto es, Alice elige “Arriba” o “Abajo” como su estrategia; simultáneamente, Bob elige “Izquierda” o “Derecha” como su estrategia, y el resultado de esas elecciones viene dado en la celda correspondiente de la matriz³.

Obsérvese que, en el estado 2, Bob está mejor eligiendo Izquierda más allá de lo que haga Alice: si juega Arriba, entonces Izquierda lleva a petróleo como resultado (que Bob prefiere), mientras que Derecha da lugar a carbón. Si juega Abajo, entonces energía nuclear (el resultado preferido por Bob) es la consecuencia de elegir Izquierda, mientras que Derecha lleva a gas. Es decir, Izquierda es la “estrategia dominante” para Bob en el estado 1. Además, dado que Bob irá hacia la Izquierda, Alice esta mejor eligiendo Arriba en lugar de Abajo, porque prefiere el petróleo a la energía nuclear. Entonces, en el estado 1, la predicción clara es que Alice juegue Arriba y Bob juegue Izquierda; es decir, (Arriba, Izquierda) es el único equilibrio de Nash⁴. Además, –y este es el punto crítico– el resultado final, petróleo, es óptimo en el estado 1.

3. Más generalmente, un mecanismo para una sociedad con n individuos es un mapeo $g: S_1 \times \dots \times S_n \rightarrow A$, donde, 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 juegan las estrategias (s_1, \dots, s_n) .

4. En general, un equilibrio de Nash es una especificación de estrategias –una para cada individuo– de la cual ningún individuo tiene incentivos a desviarse unilateralmente. Entonces, si $u(a, \theta)$ es el pago del individuo i del resultado a en el estado θ , las estrategias (s_1, \dots, s_n) constituyen un equilibrio de Nash del mecanismo g en el estado θ 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 $s'_i \in S_i$.

Al pasar al estado 2, vemos que Abajo es la estrategia dominante para Alice en ese estado. Si Bob juega Izquierda, entonces ella está mejor con Abajo que con Arriba porque prefiere la energía nuclear al petróleo. Y si Bob juega Derecha, entonces Abajo lleva a gas, que ella prefiere al resultado de Arriba, carbón. Con Alice eligiendo Abajo, Bob está mejor jugando Derecha, porque gas es mejor para él que la energía nuclear. Entonces, en el estado 2, el (único) equilibrio de Nash es (Abajo, Derecha): Alice juega Abajo y Bob juega Derecha. Además, esto resulta en el óptimo resultado, gas.

Hemos visto que en cualquier estado, el mecanismo de la Tabla 3 alcanza el resultado óptimo incluso cuando (i) el diseñador del mecanismo (la autoridad energética) no conoce siquiera el estado real y (ii) Alice y Bob están interesados sólo en sus propias preferencias, no aquellas de la autoridad. Más precisamente, porque los equilibrios de Nash del mecanismo de la Tabla 3 coinciden con los resultados óptimos en cada estado, podemos decir que el mecanismo *implementa* la regla de elección social de la autoridad en equilibrio de Nash^{5,6}.

III. Una breve historia del diseño de mecanismos

La historia intelectual de la teoría de diseño de mecanismos se remonta al menos hasta los socialistas utópicos del siglo diecinueve como Robert Owen y Charles Fourier. Ante el rechazo por lo que veían como los demonios del capitalismo floreciente, estos pensadores argumentaban que el socialismo ofrecía una alternativa más humana y a veces se veían envueltos en montar comunidades experimentales, como New Harmony, Indiana.

Una influencia más directa sobre la teoría moderna fue la Controversia de la Planificación, que alcanzó su mayor intensidad en la década de los años treinta. Los principales antagonistas eran Oskar Lange y Abba Lerner, para un bando, quienes argumentaban fervientemente que, bien hecha, la planificación central podía replicar el desempeño del libre mercado (Lange, 1936 y Lerner, 1944). De hecho –sugerían– la planificación podría corregir serias “fallas de mercado” –notablemente aquellas que podían observarse en la Gran Depresión– y de esa forma potencialmente sobrepasar a los mercados. Del otro bando, Friedrich von Hayek y Ludwig von Mises negaban incondicionalmente la posibilidad de que un sistema planificado pudiera alcanzar alguna vez el éxito del libre mercado (von Hayek, 1944 y von Mises, 1920).

La controversia fue importante y fascinante, pero para ciertos observadores, como Leonard Hurwicz, fue también algo frustrante. Esto fue porque carecía de precisión conceptual: términos críticos como “descentralización” fueron dejados sin definir. Además, los argumentos aducidos por cada bando eran muchas veces demasiado incompletos. En parte, esto era porque simplemente carecían del aparato técnico –en particular, teoría de juegos y programación matemática– para generar conclusiones realmente persuasivas.

Aquí es donde Leo Hurwicz entró en escena. Inspirado por el debate, intentó proveer

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

6. El equilibrio de Nash es una predicción de cómo se comportarán los individuos en un mecanismo. Pero un número de otros conceptos predictivos –es decir, conceptos de equilibrio– han sido considerados en la literatura de implementación, entre ellos el equilibrio perfecto en subjuegos (Moore y Rafael Repullo, 1988), equilibrio de Nash no dominado (Palfrey y Sanjay Srivastava, 1991), equilibrio bayesiano (Postlewaite y David Schmeidler, 1986), solubilidad de dominación (Hervé Moulin, 1979), equilibrio perfecto trembling-hand (Sjöström, 1993) y equilibrio fuerte (Bhaskar Dutta y Arunava Sen, 1991).

definiciones inequívocas de los conceptos centrales, y este esfuerzo culminó en sus dos grandes trabajos, Hurwicz (1960) y (1972), en donde también introdujo la crítica noción de compatibilidad de incentivos.

El trabajo inspirado por Hurwicz y otros ha producido un amplio consenso entre los economistas acerca de que von Hayek y von Mises estaban, de hecho, en lo correcto –el mercado es el “mejor” mecanismo– en escenarios donde (i) hay un gran número de compradores y vendedores, de modo que ningún agente individual tiene poder de mercado significativo y (ii) no hay externalidades significativas; esto es, el consumo, la producción y la información de un agente no afectan la producción o el consumo de otros⁷. Sin embargo, mecanismos que mejoren el mercado son generalmente posibles si alguno de los supuestos se viola⁸.

El trabajo de Hurwicz dio lugar a una enorme literatura, que se ha ramificado en dos direcciones diferentes. Por un lado, hay trabajos que hacen uso de escenarios especiales, altamente estructurados, para estudiar preguntas particulares, como cómo asignar bienes públicos, cómo diseñar subastas y cómo estructurar contratos. Por otro lado, hay estudios que obtienen resultados a un nivel general, más abstracto. Es decir, hacen la menor cantidad de supuestos posibles sobre las preferencias, tecnologías, etcétera. Mi propio trabajo ha caído en las dos categorías en momentos distintos. Pero, en este trabajo enfatizaré resultados generales.

IV. Implementación de reglas de elección social

Más arriba planteé tres preguntas centrales (A)-(C) acerca de los mecanismos incentivo-compatibles. Expresándolo nuevamente en el lenguaje de la teoría de implementación estas preguntas se convierten en:

- (A') ¿Bajo qué condiciones puede ser implementada una regla de elección social?
- (B') ¿Qué forma toma un mecanismo de implementación?
- (C') ¿Qué reglas sociales no pueden ser implementadas?

A mediados de la década de 1970 luché con estas preguntas. Eventualmente, descubrí que una propiedad llamada *monotonidad* (ahora a veces llamada *monotonidad Maskin*) es la clave para la implementabilidad en equilibrio de Nash. Supongamos que el resultado a es óptimo en el estado θ de acuerdo con la regla de elección social f en cuestión; es decir, $f(\theta) = a$. Entonces, si a no cae en el ranking de nadie relativo a cualquier otra alternativa al pasar del estado θ al estado θ' , la *monotonidad* requiere que a sea también óptimo en el estado θ' : $f(\theta') = a$. Sin embargo, si a sí cae relativo a algún resultado b en el ranking de alguien, la *monotonidad* no impone restricción alguna⁹.

Para ver lo que significa *monotonidad* más concretamente, consideremos nuestro anterior ejemplo de la energía (véanse las Tablas 1 y 2). Recordemos que petróleo es el

7. Véase, por ejemplo, Peter Hammond (1979) –que muestra, aproximadamente, que el mercado competitivo es el único mecanismo incentivo-compatibles que produce resultados individualmente racionales y Pareto óptimos– y James Jordan (1982) –que muestra lo mismo cuando “incentivo-compatibles” se reemplaza por “eficiente en términos de información”, bajo los supuestos (i) y (ii).

8. Véase, por ejemplo, Theodore Groves (1973) y Edward Clarke (1971) para el caso de bienes públicos y Jean-Jaques Laffont (1985) para el caso de externalidades informativas.

9. En un planteo más general en el cual f puede ser una correspondencia, la *monotonidad* requiere que, para todos los estados θ, θ' y todos los resultados a , si $a \in f(\theta)$ y $u_i(a, \theta) \geq u_i(b, \theta)$ implica $u_i(a, \theta') \geq u_i(b, \theta')$ para todo i y b , entonces $a \in f(\theta')$.

resultado óptimo en el estado 1. Nótese también que el petróleo cae en el *ranking* de Alice, relativo tanto al carbón como a la energía nuclear, al pasar del estado 1 al estado 2 (Alice coloca al petróleo por encima del carbón y la energía nuclear en el estado 1, pero justamente lo opuesto es cierto en el estado 2). Entonces, el hecho de que gas –no petróleo– es óptimo en el estado 2 no viola la monotonicidad. De manera similar, obsérvese que el gas cae en el *ranking* de Bob, relativo tanto al carbón como a la energía nuclear, al pasar el estado 2 al 1. Así, incluso aunque gas es óptimo en el estado 2, el hecho de que no es óptimo en el estado 1 tampoco entra en conflicto con la monotonicidad. De hecho, estas verificaciones establecen que la regla de elección social de la autoridad satisface monotonicidad (y entonces la posibilidad de implementarla, que fue mostrada más arriba, no contradice el Teorema 1 más abajo).

Pero supongamos que modificamos un poco el ejemplo, de modo que los *rankings* y los resultados óptimos vienen dados por la Tabla 4. Con estos cambios, la

| 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 óptimo | |

regla de elección social ya no es monótona. Específicamente, observemos que aunque petróleo es óptimo en el estado 1, no es óptimo en el estado 2, aunque no cae ni en el *ranking* de Alice ni en el de Bob entre los estados 1 y 2 (dado que el petróleo no cae, la monotonicidad requeriría que siguiera siendo óptimo en el estado 2). Entonces, podemos concluir que *no* hay un mecanismo que implemente la regla de elección social de la Tabla 4. Más generalmente, tenemos:

Teorema 1 (Maskin, 1977): Si una regla de elección social es implementable, entonces debe ser monótona.

Para ver por qué la regla de elección social en la Tabla 4 no es implementable, supongamos por el contrario que *hubiese* un mecanismo de implementación. Entonces, en particular, el mecanismo necesariamente debería contener un par de estrategias (s_A, s_B) –para Alice y Bob, respectivamente– que resultaran en el resultado petróleo y constituyeran un equilibrio de Nash en el estado 1.

Sostengo que (s_A, s_B) también debe constituir un equilibrio de Nash en el estado 2. Para entender esta afirmación, nótese que Bob no tiene incentivos a desviarse unilateralmente de s_B en el estado 2, ya que (i) no tiene tal incentivo en el estado 1 (por definición de equilibrio de Nash) y (ii) su orden de preferencia es el mismo en ambos estados. Además, Alice no tiene incentivos a desviarse de s_A . Para ver esto, observemos que si, al contrario de la afirmación, Alice ganara al desviarse unilateralmente de s_A en el estado 2, debería entonces estar induciendo el resultado gas (porque este es el único resultado que prefiere a petróleo en el estado 2). Pero Alice también prefiere gas a petróleo en el estado 1, y enton-

ces se beneficiaría del mismo desvío en ese estado, contradiciendo el supuesto de que (s_A, s_B) constituye un equilibrio de Nash en el estado 1.

Entonces (s_A, s_B) es un equilibrio de Nash en el estado 2. Pero el resultado que genera –petróleo– no es óptimo en ese estado, por lo que establece que la regla de elección social no es implementable después de todo.

Como hemos visto, las Tablas 1 y 2 proveen un ejemplo de una regla de elección social que es monótona y también implementable. Sin embargo, no es cierto que *todas* las reglas de elección social monótonas sean implementables; véase Maskin (1977) para un contraejemplo. No obstante, tales contraejemplos son algo artificiosos, y si una condición adicional, generalmente inocua, es impuesta, la monotonicidad sí garantiza la implementabilidad, si hay al menos tres individuos en la sociedad¹⁰.

La condición adicional se llama *no poder de veto*. Supongamos que todos los individuos, excepto posiblemente uno, están de acuerdo en que un resultado particular a es *el mejor*, queriendo decir que todos colocan a a en lo más alto de su orden de preferencias. Entonces, si la regla de elección social satisface no poder de veto, a debe ser óptimo. En otras palabras, el individuo que queda no puede “vetarlo”.

No poder de veto es especialmente inocuo –de hecho, no impone restricción alguna– cuando los resultados suponen una distribución de bienes económicos entre los individuos. En ese caso, cada individuo preferirá una porción mayor de esos bienes para sí mismo. Entonces, no habrá dos de ellos que puedan estar de acuerdo en que un resultado dado a es el mejor: los dos no pueden tener la porción más grande a la vez. Esto significa que, si hay tres o más individuos, la hipótesis propuesta por la condición de no poder de veto *no puede ser satisfecha* y, entonces, por lógica la condición se cumple automáticamente.

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

Teorema 2 (Maskin, 1977): *Supongamos que hay al menos tres individuos. Si la regla de elección social satisface monotonicidad y no poder de veto, entonces es implementable.*

Las demostraciones del Teorema 2 están más allá del alcance de este trabajo (véase Repullo 1987, para un argumento especialmente elegante), pero debería mencionar que son generalmente *constructivas*. Es decir, dada la regla de elección social a ser implementada, una demostración presenta una receta explícita para la construcción de un mecanismo que realiza el truco.

Vale la pena resaltar por qué el Teorema 2 propone al menos tres individuos. Generalmente, en economía, moverse de dos a tres personas hace las cosas más difíciles¹¹. Pero, para la teoría de implementación, tres individuos pueden de hecho hacer las cosas más fáciles. Para entender por qué, recordemos que la idea subyacente a un mecanismo es dar a los individuos el incentivo a comportarse de una manera que asegure un resultado óptimo. Esto supone “castigar” a un individuo por desviarse de su estrategia prescrita (de equilibrio). Pero si hay sólo dos individuos, Alice y Bob, y uno de ellos se ha desviado, puede

10. Esto no quiere decir que la implementación es imposible con sólo dos individuos –de hecho, nuestro ejemplo de la energía de las Tablas 1 y 2 tenía sólo dos individuos. Sin embargo, como veremos más abajo, la implementación se facilita si hay tres o más individuos.

11. Los juegos de suma cero proveen un ejemplo clásico de este fenómeno. El teorema mínimas –que simplifica grandemente el análisis del comportamiento en juegos– se aplica a juegos de suma cero de dos personas, pero no, en general, al caso de tres o más jugadores.

ser difícil determinar si fue Alice la que se desvió y Bob quien cumplió o viceversa. Este problema de identificación se resuelve una vez que hay tres personas: el que se desvía sobresale de manera más obvia cuando otros dos o más individuos cumplen con el equilibrio.

V. Comentarios finales

Esta ha sido sólo una muy breve introducción a la teoría de implementación (que en sí misma constituye sólo parte del campo del diseño de mecanismos). Me he concentrado en trabajos que han sido realizados hace más de treinta años, lo cual quizás da un erróneo sabor “antiguo” al trabajo. En efecto, un aspecto especialmente gratificante de la teoría es que casi cincuenta años después de Hurwicz (1960), el tema se mantiene intelectualmente vibrante e importante: nuevos trabajos sobre implementación aparecen todo el tiempo. Será interesante ver dónde va el campo en los próximos cincuenta años.

Referencias

- Allen, Beth (1997), “Implementation Theory with Incomplete Information” en S. Hart y A. Mas-Colell, (eds.), *Cooperation: Game Theoretic Approaches*, Berlín: Springer.
- Austen-Smith, David y Jeffrey Banks (2005), *Positive Political Theory II*, Ann Arbor: University of Michigan Press.
- Baliga, Sandeep y Eric Maskin (2003), “Mechanism Design for the Environment” en K.G. Måler y J. Vincent, eds., *Handbook of Environmental Economics*, Vol. 1, Amsterdam: North-Holland, pp. 305-324.
- Baliga, S. y Tomas Sjöström (2007), “Mechanism Design: Recent Developments” en L. Blume y S. Durlauf (eds.), *The New Palgrave Dictionary of Economics*, 2a. Edición, Londres: McMillan.
- Bergin, James (2005), *Microeconomic Theory*, Oxford: Oxford University Press.
- Clarke, Edward (1971), “Multipart Pricing of Public Goods”, *Public Choice*, pp. 19-33.
- Corchon, Luis (1996), *The Theory of Implementation of Socially Optimal Decisions in Economics*, Londres: Macmillan.
- Corchon, L. (2008), “The Theory of Implementation”, *The Encyclopedia of Complexity and System Science*, Berlín: Springer.
- Dasgupta, Partha, Peter Hammond, y E. Maskin (1979), “The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility”, *Review of Economic Studies*, 46, pp. 185-216.
- Dasgupta, P. y E. Maskin (forthcoming), “On the Robustness of Majority Rule”, *Journal of the European Economic Association*.
- Dutta, Bhaskar y Arunava Sen (1991), “Implementation under Strong Equilibrium: A Complete Characterization”, *Journal of Mathematical Economics*, 20, pp. 46-67.
- Feldman, Allan y Roberto Serrano (2006), *Welfare Economics and Social Choice Theory*, Berlín: Springer.
- Groves, Theodore (1973), “Incentives in Teams”, *Econometrica*, 41, pp. 617-631.
- Groves, T. y John Ledyard (1987), “Incentive Compatibility since 1972” en T. Groves, R. Radner, y S. Reiter (eds.), *Information, Incentives and Economic Mechanisms*, Minneapolis: University of Minnesota Press, pp. 48-111.
- Hammond, Peter (1979), “Straightforward Individual Incentive Compatibility in Large Economies”, *Review of Economic Studies*, 46, pp. 263-282.
- Hurwicz, Leonid (1960), “Optimality and Informational Efficiency in Resource Allocation Processes” en Kenneth Arrow, S. Karlin y P. Suppes, (eds.), *Mathematical Methods in Social Sciences*, Stanford: Stanford University Press, pp. 27-46.
- Hurwicz, L. (1972), “On Informationally Decentralized Systems” en C. McGuire, y R. Radner, (eds.), *Decision and Organization*, Amsterdam: North-Holland, pp. 297-336.
- Jackson, Matthew (2001), “A Crash Course in Implementation Theory”, *Social Choice and Welfare*, 18, pp. 655-708.
- Jordan, James (1982), “The Competitive Allocation Process is Informationally Efficient Uniquely”, *Journal of Economic Theory*, 28, pp. 1-18.
-

- Laffont, Jean-Jacques (1985), "On the Welfare Analysis of Rational Expectations Equilibria with Asymmetric Information", *Econometrica*, 53, pp. 1-29.
- Lange, Oskar (1936), "On the Economic Theory of Socialism", *Review of Economic Studies*, 4, pp. 53-71.
- Lerner, Abba (1944), *The Economics of Control*, New York: McMillan.
- Maskin, Eric (1977, published 1999), "Nash Equilibrium and Welfare Optimality", *Review of Economic Studies*, pp. 23-38.
- Maskin, E. (1985), "The Theory of Implementation in Nash Equilibrium: A Survey" en L. Hurwicz, D. Schmeidler, y H. Sonnenschein (eds.), *Social Goals and Social Organization*, Cambridge: Cambridge University Press.
- Maskin, E. y T. Sjöström (2002), "Implementation Theory" en K. Arrow, A. Sen, y K. Suzumura, (eds.), *Handbook of Social Choice and Welfare*, Vol. I, Amsterdam: Elsevier, pp. 237-288.
- Moore, John (1992), "Implementation, Contracts, and Renegotiation in Environments with Complete Information" en J. J. Laffont (ed.), *Advances in Economic Theory*, Vol. 1, Cambridge: Cambridge University Press, pp. 182-282.
- Moore, J. y Rafael Repullo (1988), "Subgame Perfect Implementation", *Econometrica*, 56, pp. 1191-1220.
- Moulin, Hervé (1979), "Dominance Solvable Voting Schemes", *Econometrica*, 47, pp. 1337-1351.
- Osborne, Martin y Ariel Rubinstein (1994), *A Course in Game Theory*, Cambridge: MIT Press.
- Palfrey, Thomas (1992), "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design" en J. J. Laffont (ed.), *Advances in Economic Theory*, Vol. 1, Cambridge: Cambridge University Press, pp. 283-323.
- Palfrey T. (2001), "Implementation Theory" en R. Aumann y S. Hart, (eds.), *Handbook of Game Theory*, vol. 3, Amsterdam: North-Holland, pp. 2271-2326.
- Palfrey, T. y Sanjay Srivastava (1991), "Nash Implementation using Undominated Strategies", *Econometrica*, 59, pp. 479-501.
- Postlewaite, Andrew (1985), "Implementation via Nash Equilibria in Economic Environments" en L. Hurwicz, D. Schmeidler, y H. Sonnenschein (eds.), *Social Goals and Social Organization*, Cambridge: Cambridge University Press, pp. 205-228.
- Postlewaite, A. y David Schmeidler (1986), "Implementation in Differential Information Economies", *Journal of Economic Theory*, 39, pp. 14-33.
- Rasmusen, Eric (2006), *Games and Information: An Introduction to Game Theory*, Oxford: Blackwell Publishing.
- Repullo, Rafael (1987), "A Simple Proof of Maskin's Theorem on Nash Implementation", *Social Choice and Welfare*, 4, pp. 39-41.
- Serrano, Roberto (2004), "The Theory of Implementation of Social Choice Rules", *SIAM Review*, 46, pp. 377-414.
- Sjöström, Tomas (1993), "Implementation in Perfect Equilibria", *Social Choice and Welfare*, 10, pp. 97-106.
- von Hayek, Friedrich (1944), *The Road to Serfdom*, Londres: Routledge.
- von Mises, Ludwig (1935), "Die Wirtschaftsrechnung im Sozialistischen Gemeinwesen" en F. von Hayek (ed.), *Collectivist Economic Planning*, Londres: Routledge.