Implementación y diseño de mecanismos

Una de niños cuarteados
En el capítulo 3 del Libro de los Reyes del Antiguo Testamento se relata el conocido como "Juicio del Rey Salomón". Dos prostitutas se presentan ante el rey. Una de ellas explica que ambas comparten una casa, en compañía de nadie más, y que ambas han dado recientemente a luz. Continúa explicando que el hijo de la otra mujer ha fallecido y que, mientras dormía, la otra mujer le ha arrebatado a su hijo. Por ello, se presenta ante el rey para pedirle que le devuelva a su hijo. La otra mujer niega los cargos, asegura que el bebé fallecido es de la acusadora y afirma que el bebé que ha sobrevivido es suyo. Entonces el rey ordena que el bebé sea partido por la mitad con una espada y cada parte entregada a una madre. En ese momento, la madre acusada implora al rey que, antes que muerto, prefiere ver a su bebé en manos de la madre acusadora. Ésta, por el contrario, acepta la partición del bebé. La decisión final del rey es entregar el bebé a la madre acusada, aparentemente basándose en la presunción de que la auténtica madre preferiría que su hijo viviera aunque fuera con otra madre.
http://www.newadvent.org/bible/1ki003.htm http://en.wikipedia.org/wiki/King_Solomon

Esta historia permite ilustrar en qué consiste el diseño de mecanismos: en construir un juego la solución del cual conduzca a un resultado previamente seleccionado y pretendido. Este resultado seleccionado es el resultado a implementar (obtener) mediante el juego. En el juicio, el resultado que pretende obtener el rey es entregar el bebé a la auténtica madre. El problema es que el rey no dispone de la información suficiente para tomar esta decisión: es información privada quién sea la madre auténtica.

Un mecanismo (también llamado "forma de juego" o *game form*) consiste en la especificación de estrategias para cada jugador (los mensajes mediante los que éstos puden transmitir su información privada) junto con una regla que determine qué resultados produce cada posible combinación de estrategias de los jugadores. Un mecanismo no es un juego. Pero los jugadores se suponen dotados de preferencias sobre los resultados, de forma que cuando se combina el mecanismo con esas preferencias lo que se obtiene sí es un juego.

El siguiente paso consiste en escoger qué concepto de solución de un juego se adopta. Tres de los conceptos de referencia son el equilibrio en estrategias dominantes (equilibrio dominante), el equilibrio de Nash y el equilibrio bayesiano. Escogido un concepto de solución, ya puede definirse en qué consiste la implementación del resultado deseado mediante el concepto de solución escogido: el resultado deseado se dice que es implementable mediante el mecanismo diseñado en términos del concepto de solución que se haya adoptado si ese resultado es el único que produce el concepto de solución en el juego obtenido cuando se combina el mecanismo propuesto con las preferencias que tengan los jugadores sobre los resultados.

El diseño de mecanismos va en dirección opuesta al análisis de un juego. Cuando se analiza un juego, se parte de las estrategias de que disponen los jugadores y se trata de investigar qué tipo de resultados pueden razonable o justificadamente obtenerse en el juego. Cuando se diseña un mecanismo ocurre lo contrario: se parte de un resultado y se trata de estructurar un mecanismo asignando estrategias a los jugadores para que sus decisiones produzcan el resultado deseado.

Un mecanismo para el juicio del Rey Salomón

El juicio descrito anteriormente pretende, aparentemente, evidenciar la sabiduría del rey. Sin embargo, el rey consiguió implementar el resultado deseado por fortuna, no por un diseño adecuado del mecanismo: el rey fue, más que sabio, afortunado. La razón de su fortuna es que la madre acusadora (la madre falsa) no jugó una mejor respuesta a la estrategia seguida por la madre acusada (la madre auténtica). ¿Qué habría hecho el rey si la madre acusadora hubiese replicado la estrategia que, a primera vista, escogería una madre auténtica? Esto es, ¿qué habría decidido el rey si la acusadora también hubiese implorado que, antes que arrebatarle la vida, el niño fuese entregado a la otra madre?

La Fig. 1 muestra un juego que sí habría permitido implementar el resultado deseado (que el bebé fuese entregado a su madre) empleando como concepto de solución el equilibrio perfecto en subjuegos. El mecanismo en el que se basa el juego es el siguiente. La naturaleza (jugador 0) determina quién es la madre demandante: con probabilidad $0 < p < 1$, la madre demandante es la madre falsa (jugador 1) y, por tanto, el bebé está inicialmente en manos de la madre verdadera; y con probabilidad $1 - p$, la madre demandante es la madre verdadera (jugador 2) y, por tanto, el bebé está inicialmente en manos de la madre falsa. Una vez que la naturaleza establece quién es la madre demandante, ésta decide si manifestar que ella no es la madre (acción *n*) o afirmar que sí lo es (acción *s*). Si la declaración de la madre demandante es "no soy la madre", el juego acaba. Si la declaración es "soy la madre" entonces la otra madre decide si sostener que es la madre o no.

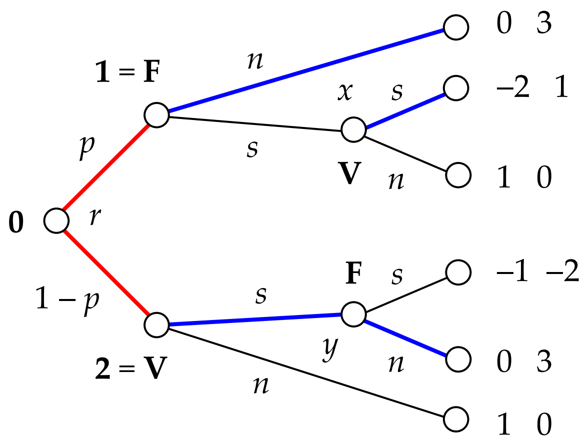

Fig. 1. Un juego de Binmore (2008)

Los resultados del mecanismo son los siguientes. Si el bebé está inicialmente en manos de la madre verdadera (lo que hace que la demandante sea la madre falsa y nos encontremos en la parte alta de la Fig. 1) entonces el bebé queda en manos de la madre verdadera a menos que ésta diga que el bebé no es suyo y la madre demandante diga que sí lo es. Además, si ambas declaran ser las madres, cada una pagará una multa. Por otro lado, si el bebé está inicialmente en manos de la madre falsa (lo que hace que la demandante sea la madre verdadera) entonces el bebé seguirá en manos de la madre falsa a menos que ésta diga que el bebé no es suyo y la madre demandante diga que sí lo es. Como en el caso anterior, si ambas declaran ser las madres, cada una pagará una multa. Cuando se añadan pagos al mecanismo se obtendrá un juego.

En la Fig. 1, los pagos se han establecido suponiendo que el valor de tener el bebé para la madre verdadera (jugador 2) es 3 y que el valor de tenerlo para la madre falsa (jugador 1) es 1. El importe de la multa (que se impone si las dos declaran ser la madres) es de 2 para ambas. Por último, no recibir el bebé ni asumir la multa implica un pago de 0.

Q1. Verifica que los pagos de la Fig. 1 son correctos. En particular, ¿por qué el vector de pagos cuando ambos jugadores declaran ser las madres no es el mismo en la parte superior del juego (vector (–2, 1)) que en la parte inferior (vector (–1, –2))?

Resolviendo el juego por inducción hacia atrás, tomemos el nudo de decisión x. La mejor respuesta en este nudo para el jugador 2 (la madre verdadera) es s. Dada la elección de s en x, la mejor respuesta del jugador 1 en el nudo que precede inmediatamente a x es n. Pasando al nudo de decisión y, la mejor respuesta en este nudo para el jugador 1 (la madre falsa) es n. Dada la elección de n en y, la mejor respuesta del jugador 2 en el nudo justo antes de y es s. Por tanto, <u>decida lo que decida la naturaleza, el resultado siempre es el mismo: la madre falsa niega ser la madre y la madre auténtica afirma serlo</u>. Mediante el equilibrio perfecto en subjuegos, el mecanismo que da lugar al juego de la Fig. 1 ha permitido implementar el resultado deseado: que el bebé siempre vaya a parar a manos de la madre verdadera (vector de pagos (0, 3)).

Q2. Representa el juego de la Fig. 1 como juego simultáneo (en forma matricial) y comprueba que el único equilibrio de Nash con estrategias puras es aquél en el que la madre falsa siempre escoge n y la madre verdadera siempre escoge s. Por tanto, el mecanismo subyacente a ese juego simultáneo también habría permitido implementar (mediante el concepto de solución del equilibrio de Nash con estrategias puras) el resultado de entregar el bebé a la madre auténtica.

Una de profesores examinados

Como segundo ejemplo del problema de la implementación y del diseño de mecanismos, consideremos la siguiente historia, basada en un ejemplo que Eric Maskin presenta en http://nobelprize.org/nobel_prizes/economics/laureates/2007/maskin-slides.pdf (discurso de aceptación del Premio Nobel de Economía de 2007). Un profesor puede ser de 4 tipos: x = exigente, y = despreocupado, z = bromista y v = motivador. Los estudiantes del curso que imparte el profesor tienen preferencias sobre el tipo de profesor y, sobre la base de estas preferencias, evalúan al profesor en una encuesta docente. Para simplificar, supongamos que sólo hay dos estudiantes: él y ella (también podria asumirse que, para algún $k > 1$, hay $2k$ estudiantes que las preferencias de él y $2k$ estudiantes con las preferencias de ella).

Tanto él como ella pueden ser de dos tipos: del tipo interesado t_i por la asignatura o del tipo no interesado t_n. De las cuatro combinaciones posibles de estos dos tipos, supongamos que sólo dos son posibles (o tienen probabilidad positiva): $w = (t_i, t_i)$ y $w' = (t_n, t_n)$. Podemos identificar cada combinación posible de tipos con un estado del mundo: en el estado w, ambos estudiantes están interesados en la asignatura; en el estado w', ninguno está interesado. Las preferencias de los estudiantes en cada estado son las siguientes.

Estado w	Orden de preferencia de él:	$x \to y \to z \to v$	De ella:	$v \to y \to z \to x$
Estado w'	Orden de preferencia de él:	$v \to x \to z \to y$	De ella:	$y \to x \to z \to v$

En la encuesta docente, los estudiantes han de puntuar al profesor, de 1 a 4, sabiendo el tipo del profesor. Cada estudiante da 4 puntos al profesor si su tipo es el más preferido; 3 si su tipo es el segundo más preferido; 2 si es el tercero; y 1 si es el menos preferido (por tanto, si el tipo del profesor ocupa la posición k en el orden de preferencias del estudiante, éste le da $5 - k$ puntos).

La puntuación del profesor en cada estado es la suma de los puntos que recibe de él y de ella en ese estado. El objetivo (el resultado deseado) del profesor es escoger el tipo que, en cada estado, le dé la puntuación más alta posible. El profesor sabe cuáles son las preferencias de él y de ella en cada estado, pero ignora en qué estado está (el tipo de estudiante es información privada). Dado que el profesor sabe qué preferencias tendrían los estudiantes en cada estado, sabe que en el estado w el tipo de profesor que recibe máxima puntuación es y. En w, el tipo y obtendría 6 puntos; x i v obtendrían 5 cada uno; y z obtendría 4. En w', el tipo mejor puntuado sería x.

Por tanto, el profesor desea ser de tipo y (despreocupado) cuando el estado del mundo es w (la asignatura interesa a los estudiantes) y ser de tipo x (exigente) cuando el estado del mundo es w' (la asignatura no interesa a los estudiantes). Su deseo es, pues, implementar y en w i x en w'. El problema del profesor consiste en diseñar un mecanismo mediante el cual las decisiones de los estudiantes generen el resultado pretendido por el profesor: obtener máxima puntuación.

Un mecanismo se llama directo cuando, en el juego que induce el mecanismo, las estrategias de cada jugador consisten en revelar su tipo. En un mecanismo directo, el profesor preguntaría a un estudiante si le interesa la asignatura o no. Si la revelación del estudiante escogido es sincera, cuando el estado del mundo es w, el profesor lo sabrá y escogerá ser del tipo y; y cuando el estado es w', el profesor también lo sabrá y escogerá ser del tipo x. Y problema resuelto.

Por desgracia para el profesor, este mecanismo no incentiva a decir la verdad (técnicamente, no es compatible con los incentivos) cuando los estudiantes saben cuál es el propósito del profesor (ser y en w y ser x en w'). Para comprobar que ambos estudiantes tendrían incentivo a mentir, supongamos que el profesor decide preguntarle a él.

Si el estado es w, él sabe que afirmando "Me interesa la asignatura" el profesor asumirá que el estado es w y sabe que el profesor escogerá ser y. En cambio, él sabe que diciendo "No me interesa la asignatura", el profesor, asumiendo que el estado es w', escogerá ser x. Dado que, en el estado w, él prefiere que el profesor sea del tipo x a que sea del tipo y, él tiene incentivo a mentir en el estado w diciendo que están en el w'.

Si el estado es w', por el mismo razonamiento que en el caso anterior, él sabe que revelando la verdad (diciendo "No me interesa"), el profesor será del tipo x y que mintiendo (diciendo "Me interesa"), el profesor será del tipo y. Puesto que, en el estado w', él prefiere que el profesor sea del tipo x a que sea del tipo y, él no tiene, en el estado w', incentivo a mentir. Por tanto, en w' dirá que están en w'.

Pero entonces el profesor se enfrenta con una dificultad: sea cual sea el estado del mundo, a él le conviene siempre decir que el estado del mundo es w'. Conclusión: lo que diga él no es fiable. En ese caso, el profesor puede dirigirse a ella y preguntarle qué le parece la asignatura.

Q3. Verifica que, en ambos estados, a ella le conviene decir que el estado es w.

En vista de lo anterior, la opinión que exprese ella tampoco es fiable para el profesor. La conclusión final es que el mecanismo directo consistente en preguntar a los estudiantes no permite implementar (cuando los estudiantes escogen mejores respuestas) el resultado deseado por el profesor. La Fig. 2 muestra un mecanismo que sí que lo conseguiría cuando el concepto de solución escogido es el de equilibrio de Nash.

		ella	
		c	d
él	a	y	z
	b	v	x

Fig. 2. Un mecanismo

En este mecanismo, el profesor le dice a él que escoga entre a o b i, a ella, que escoga entre c i d. Lo que representen a, b, c i d es irrelevante: a y c pueden consistir en levantar la mano derecha y b y d en levantar la izquierda; o en escribir las letras a, b, c y d en un papel… lo que sea. Por tanto, $\{a, b\}$ es el conjunto de mensajes de él y $\{c, d\}$ el de ella.

Aparte de indicar el conjunto de mensajes (o estrategias) de cada jugador, el mecanismo debe establecer qué resultado produce cada combinación de mensajes. El conjunto de combinaciones de mensajes viene dado por el producto cartesiano $\{a, b\} \times \{c, d\}$. Los elementos de este conjunto son (a, c), (a, d), (b, c) y (b, d). Cada uno de estos elementos puede asociarse con una casilla de la matriz de la Fig. 2. La función de resultados r completa la descripción del mecanismo asociando un resultado con cada elemento de $\{a, b\} \times \{c, d\}$. En este ejemplo, el conjunto de resultados es el conjunto $\{x, y, z, v\}$ de tipos del profesor. El mecanismo de la Fig. 2 es tal que $r(a, c) = y$, $r(a, d) = z$, $r(b, c) = v$ i $r(b, d) = x$. Por ejemplo, $r(a, c) = y$ significa que si él escoge a y ella escoge c entonces el profesor decide ser del tipo y.

El mecanismo de la Fig. 2 puede interpretarse del modo siguiente. Por un lado, el profesor le da a él el poder de determinar si el profesor escoge ser un tipo del conjunto $\{y, z\}$ o si escoge ser un tipo del conjunto $\{v, x\}$: escogiendo a, él fuerza al profesor a ser y o z; escogiendo b, él hace que el profesor se limite a ser v o x. Por otro lado, el profesor le da a ella el el poder de determinar si el profesor escoge ser un tipo del conjunto $\{y, v\}$ (lo que ella consigue seleccionando c) o si escoge ser un tipo del conjunto $\{z, x\}$ (lo que ella consigue seleccionando d).

El mecanismo de la Fig. 2 se transforma en un juego cuando añadimos las preferencias que tienen los estudiantes sobre los resultados. Para visualizar con más claridad el juego resultante, tomemos las siguientes representaciones numéricas de las preferencias de los estudiantes: la utilidad del resultado más preferido es 3, la del segundo más preferido 2, la del tercero 1 y la del menos preferido, 0 (la función de utilidad sería $u(\alpha) = 4 - k$ si el resultado α ocupa la posición k en el orden de preferencia representado). Por ejemplo, la funció de utilidad de él en el estado w sería $u(x) = 3$, $u(y) = 2$, $u(z) = 1$ y $u(v) = 0$.

La Fig. 3 muestra el juego que juegan los estudiantes en cada estado, en donde el primer número en los vectores de pagos representa la utilidad (o pago) de él y el segundo representa el de ella.

		ella	
		c	*d*
él	*a*	2 2	1 1
	b	0 3	3 0

estado w

		ella	
		c	*d*
él	*a*	0 3	1 1
	b	3 0	2 2

estado w'

Fig. 3

El juego de la Fig. 3 es un juego bayesiano trivial porque cada estudiante sabe cuál es el tipo del otro estudiante. Ello permite resolver todo el juego resolviendo cada matriz por separado. Si el estado es w, ambos estudiantes lo saben y saben que el mecanismo del profesor induce el juego de la matriz izquierda en la Fig. 3. El único equilibrio de Nash de este juego (con estrategias pura o mixtas) és [a, c], puesto que c es una estrategia fuertemente dominante para ella y, dada c, la mejor respuesta de él es a.

Si el estado es w', ambos estudiantes lo saben y saben que el mecanismo del profesor induce el juego de la matriz derecha en la Fig. 3. El único equilibrio de Nash de este juego (con estrategias pura o mixtas) és [b, d], puesto que b es una estrategia fuertemente dominante para él y, dada b, la mejor respuesta de ella es d.

Como consecuencia, si el estado es w, los estudiantes juegan [a, c], que produce el resultado y deseado por el profesor cuando el estado es w. Y si el estado es w', los estudiantes juegan [b, d], que produce el resultado x deseado por el profesor cuando el estado es w'. Conclusión: el mecanismo de la Fig. 2 permite implementar (mediante equilibrios de Nash) el resultado deseado por el profesor: ser y en w y ser x en w'.

Referencias

- Binmore, Ken (2008): L*a teoría de juegos. Una breve introducción*. Alianza Editorial: Madrid.