Universidad Nacional Autónoma de México

Facultad de Ciencias

Incentivos y Diseño de Mecanismos: Los Teoremas de Arrow y de Myerson - Satterthwaite

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Matemático

PRESENTA:

Carlos Edgardo Calvo Hernández

DIRECTOR DE TESIS:

Dr. Salvador Ferrer Ramírez

2015

Ciudad Universitaria, D. F. ,

1. Datos del alumno.
Calvo
Hernández
Carlos Edgardo
55640263
Universidad Nacional Autónoma de México
Facultad de Ciencias
Matemáticas
301561686

2. Datos del tutor.
Ferrer
Ramírez
Salvador

3. Datos del sinodal 1
Dr
Jorge
Ruiz
Moreno

4. Datos del sinodal 2
M en C
María de la Paloma Carmén
Zapata
Y Lillo

5. Datos del sinodal 3
Dr
Salvador
Ferrer
Ramírez

6. Datos del sinodal 4
Dra
Bibiana
Obregón
Quintana

7. Datos del trabajo escrito
Incentivos y Diseño de Mecanismos: Los Teoremas de Arrow y Myerson-Satterthwaite
88 p.
2015

A ti,
quien hizo todo esto posible,
solo puedo decir:
¿Por qué tan lejos?

Agradecimientos

Quiero agradecer a todos y cada uno de aquellos que me han apoyado a lo largo de la vida para que esto fuera posible, familia, amigos y todos aquellos que de alguna manera u otra han tocado mi vida, pues es por todos ustedes que esto ha sido posible. Primero a las dos personas más importantes de mi vida, mis papás, Gustavo y María del Carmen, sin quienes nunca hubiera podido lograr todo lo que he hecho, mucho menos esto. Por apoyarme desde siempre y confiar en mi aunque no me lo ganara, ni lo mereciera. Gracias por su amor, apoyo y comprensión. Los amo.

A Emmanuel Alejandro Albanés y Daniela Hernández sin quienes nunca hubiera terminado la carrera, no tengo palabras. Muchachos, tienen mi más profunda gratitud. A mis amigos y hermanos: Armando Pérez y Paulina González por su constante e incondicional apoyo a lo largo de los años, siempre los querré como a pocos; Paulina Garza por su siempre presente presión y apoyo, no sabes lo que esto significa, se que no siempre me sale, pero le echo muchas ganas; Gabriel Campos y a Joyce Vega por todos esos años que sufrimos juntos en esta escuela; Francisco por tantos años, vidas y demás. A Alberto, Perla, Alfredo, José Antonio, Moises, Azuri, Marco y más por ser mis amigos y por estar ahí, por siempre creer en mi. En general a todos aquellos que hayan tenido poco o mucho que ver en que haya sido posible que terminara aquí. Gracias.

A mi asesor, Salvador, que a pesar del tiempo, los problemas y las necedades hizo que este trabajo fuera posible. Gracias a su constante apoyo y supervisión espero que esto sea útil al menos para alguien. A Jorge Ruiz por todos sus comentarios, las discusiones que tuvimos y el tiempo que dedicó a la revisión de este trabajo. A Bibiana Obregón por su ayuda, tiempo y consejo, hizo que este trabajo llegará a ser la mejor versión que podía ser. En general, a todos los maestros y sinodales que me ayudaron y aconsejaron para terminarlo.

A todos ustedes y a quienes haya olvidado, muchas gracias por contribuir a que este sueño se volviera una realidad.

Índice general

Introducción

Este trabajo de tesis tiene como objetivo final comparar los teoremas de Arrow y de Myerson-Satterthwaite. Comenzaremos dando el marco teórico para ambos teoremas y a partir de ahí, concluiremos con la posibles similitudes y diferencias que existan entre ellos. La intención final es ver qué alcance y limitaciones tienen, tanto por sí solos, como uno frente al otro. Cabe mencionar que no se pretenden contribuciones originales, solamente realizar una comparación que no suele hacerse y que al estar fundamentada en dos teorías diferentes está lejos de lo trivial porque genera un nuevo enfoque hacia el diseño de mecanismos y a la elección social.

Todo parte de la microeconomía, en particular, de la economía del bienestar. Por un lado tenemos a la teoría de elección social en la que está contenido el teorema de Arrow y por otro la teoría del diseño de mecanismos, parte de la teoría de juegos, que da pie al teorema de Myerson-Satterthwaite.

A pesar de que en esencia, y como veremos a lo largo del trabajo, estas dos teorías son diferentes, el objetivo subyacente es denotar cómo la toma de decisiones tiene distintas formas de ser modelada, además de la dificultad de cómo compatibilizar la idea de racionalidad con la de incentivos.

El desarrollo del trabajo es a través de 4 partes; en la primera daremos los antecedentes históricos que precedieron a ambos teoremas, para de esta manera entender el contexto detrás de ellos. En la segunda, daremos una caracterización general de la teoría de elección social. Empezaremos con las definiciones básicas de las relaciones de preferencia y de las funciones de utilidad. Dentro del siguiente capítulo entraremos en el caso más básico de elección social (el caso para dos jugadores), del que partiremos para explicar el caso general de n jugadores, que concluye con el teorema de Arrow, ejemplificando que mediante una votación mayoritaria no es posible llegar a una decisión social, a menos que ciertas condiciones se cumplan.

La tercera parte comienza con las definiciones básicas de la teoría de juegos, esto con el propósito de posteriormente definir los conceptos de la teoría de diseño de mecanismos y después poder mostrar la manera en la que los diferentes resultados o soluciones que se pueden obtener en un juego, modifican la dinámica y el diseño de los mecanismos. Hablaremos de diferentes conceptos de solución cómo equilibrios de Nash, equilibrios en estrategias dominantes y equilibrios bayesianos de Nash. En este punto hablaremos del teorema de Gibbard-Satterthwaite, un resultado del diseño de mecanismos muy relacionado con el teorema de Arrow, con la intención de acercar cada vez más estas dos teorías.

Una vez establecidas estas bases y cuáles son las restricciones que tiene la creación y utilización de mecanismos, daremos la caracterización del escenario de intercambio bilateral, de tal forma dar pie para el enunciado y demostración del teorema de Myerson-Satterthwaite el cual, intuitivamente, nos dice que no hay un método eficiente para que dos agentes intercambien un bien sin que alguno de los dos pierda.

Por último, en la cuarta parte, expondremos las conclusiones de ambos resultados, y si es que existiera, la relación entre ellos así como las implicaciones en un ambiente básico y claramente limitado. Siempre teniendo en mente que la idea parte de las implicaciones que rodean a la toma de decisiones.

Mi interés por el tema nace de observar que muchos de los resultados que podemos encontrar en la literatura microeconómica tratan de resolver problemas similares, siempre utilizando herramientas teóricas diferentes.

Parte I

Antecedentes

Nuestro primer tema de interés será la teoría de elección social, cuyo fundamento es el análisis de colecciones de preferencias, lo cual lo podemos entender como la acumulación de varias clasificaciones de preferencias individuales en una sola clasificación colectiva. Inspirado por el trabajo de Alfred Tarski, en 1950, Kenneth Arrow, a quien se le otorgaría el Premio Nobel de Economía en 1972, introdujo un acercamiento general al estado de colecciones de preferencia para métodos de votación, en particular para las votaciones por mayoría. Fue capaz de demostrar que, sorprendentemente, no existe un método para obtener una ordenación de preferencias sociales, a partir de preferencias individuales. Resultado hoy conocido como el Teorema de Imposibilidad de Arrow.

El teorema de Arrow constituye un tema clásico dentro de la hoy llamada, economía del bienestar. Él consideró una clase de posibles métodos de acumulación, a los que llamó funciones de bienestar social, estas tenían como objetivo encontrar qué alternativas o resultados eran preferidos para la sociedad como conjunto; a raíz de ellas se preguntó cuáles cumplían ciertos supuestos.

Este teorema de "imposibilidad" - o de "posibilidad general", como Arrow lo llamó - responde una pregunta básica en la teoría de decisiones colectivas. ¿Qué procedimientos existen para encontrar una ordenación "social" de preferencias, basada en la información que se tiene disponible acerca de las preferencias de cada individuo? Digamos, existen ciertas alternativas de entre las que se puede elegir, como pueden ser: políticas, proyectos públicos, candidatos en una elección, distribuciones de ingreso y trabajo en una sociedad, etc. En general, el propósito es tomar la decisión con base en las preferencias de los individuos miembros de la sociedad.

La respuesta es sorprendente. El teorema de Arrow dice que si consideramos ciertos supuestos, aparentemente razonables, sobre la autonomía de los agentes y la racionalidad de sus preferencias, entonces no existen procedimientos que respeten estas preferencias y nos de como resultado una ordenación social deseable.

Gran parte de la investigación contemporánea de la teoría de elección ha sido motivada por este teorema, y junto con el escenario técnico propuesto por Arrow, han detonado el estudio de problemas actuales en la economía del bienestar. Dicho escenario da a las ordenaciones sociales un sentido preciso y una respuesta rigurosa. Arrow aseguró que el problema de encontrar un procedimiento de acumulación de preferencias surge de las alternativas entre las que hay que tomar una decisión. Como ya comentamos, la naturaleza de dichas alternativas depende del tipo de problemas de elección que se quiere estudiar. El problema al que se enfrenta Arrow surge una vez que algunas alternativas y algunos agentes han sido fijados; lo que incita a encontrar un procedimiento de acumulación para ellos, pero es importante notar que este problema existe antes de que se pueda conocer la información relevante acerca de las preferencias individuales de los agentes. Con esto en mente, podemos reformular la pregunta hecha anteriormente. ¿Qué procedimientos, si es que existen, logran una ordenación social de ciertas alternativas dadas, basadas en las preferencias individuales acumuladas, sin importar de quién resulten ser estas preferencias?

En la práctica, algunas veces antes de conocer las alternativas y los agentes, debemos elegir el procedimiento que se usará para tomar decisiones sociales. Por ejemplo, en elecciones recurrentes se usa cada vez el mismo mecanismo de votación para determinar al ganador, sin embargo, en cada elección existen conjuntos de candidatos y de votantes diferentes. Este tipo de procedimientos crea un problema para el escenario de Arrow, ya que no son directamente analizables en este. Aún así, el resultado del teorema, sigue siendo relevante; porque nos dice que aunque las alternativas y las personas se fijen, de todos modos, no existe un método completamente adecuado para lograr ordenaciones sociales.

Ahora, consideremos otra forma de tratar de tomar decisiones colectivas, el diseño de mecanismos. Pensemos en dos niños pequeños que pelean por dividir un pastel. Nuestra motivación es encontrar un método para dividirlo equitativamente. Tal vez, quienes ya sean padres, tendrán una idea empírica de cómo hacerlo, que uno de los niños parta y que el otro elija. Esto le da un incentivo al niño que corta para hacerlo lo más parejo posible. Esta decisión tan sencilla es un ejemplo de un mecanismo compatible con incentivos, objetos de estudio primordiales de la teoría del diseño de mecanismos.

En 2007, Leonid Hurwicz, Eric Maskin y Roger Myerson, recibieron el Premio Nobel de Economía "por haber sentado las bases de la teoría del diseño de mecanismos". Una de estas bases es el uso y modelación de instituciones para obtener resultados sociales. Pensemos en estas instituciones o "mecanismos" como procedimientos que a partir de "mensajes" o "señales" de los participantes ofrece un resultado. La idea esencial del diseño de mecanismos es crear instituciones que produzcan resultados deseables, mientras se respeta el hecho que los participantes tienen información privada y son egoístas. Esto nos lleva a buscar diseñar mecanismos que funcionen deseablemente, mientras respetan las restricciones impuestas por la información y el egoísmo de los agentes, lo que resulta sumamente difícil. Si quisiéramos ver la importancia de la información, nos basta pensar en un caso de dos participantes; pensemos la venta de un bien. ¿Si existe información que solo uno de los dos participantes conoce, habrá impedimentos para realizar la venta? Debería ser claro que si el vendedor conoce algo del bien a vender (calidad, tamaño, tiempo, etc.) que el comprador no puede saber, habría un impedimento para la realización de una venta eficiente. En este sentido el teorema de Myerson-Satterthwaite, muestra que al existir información privada sobre los valores se puede prevenir que se obtenga un resultado deseado.

Es claro que la información es importante en este tipo de análisis, pero para entender la transición que ha tenido la economía y su visión, nos podemos remontar a sus inicios, a la antigua Grecia, en particular al libro de Xenofonte “Oeconomicus” (360 A.C), en el que Sócrates entrevista a un ciudadano que tiene dos problemáticas. La primera, motivar a los trabajadores de su granja para que sea productiva y la otra, mantener su estatus político en la ciudad, actividad que le permite conservar la granja. De esto, podemos concluir que la preocupación en este caso es sobre los incentivos y las motivaciones de los individuos, para este caso en particular del ciudadano entrevistado. Aunque podemos rastrear el inicio de la economía a este punto, y hoy en día este tipo de preocupaciones sobre incentivos es parte central de la teoría económica, no siempre ha sido esta la visión de la economía. Hasta hace poco más de doscientos años, la teoría económica tenía otro enfoque, principalmente centrado en la producción, fue durante esta época cuando se empezó a desarrollar como una ciencia social analítica, a través de la creación de metodologías sobre el ingreso y la teoría de precios. A pesar que las preguntas sobre asignación de recursos parecían particularmente dispuestos para el análisis matemático, dado que los flujos de bienes y dineros son medibles y sus supuestos cuantificables, el problema clásico de la economía se mantenía siendo que la habilidad de las personas de satisfacer sus deseos está restringida por recursos limitados. El resultado clásico era que el comercio libre y sin limitaciones puede lograr una *asignación eficiente*.

Fue a partir de las investigaciones de Cournot (1838), se empezó a generar un cambio de enfoque; que de centrarse en la asignación de recursos pasó a un pensamiento sobre el análisis de incentivos. Paulatinamente, los teóricos analizaron las decisiones óptimas de los individuos racionales, esto como una herramienta para entender la oferta y la demanda dentro de la teoría de precios. No fue hasta la primera mitad del siglo XX cuando algunos matemáticos comenzaron a formular modelos para analizar las decisiones competitivas en escenarios más generales, con esto, sentando las bases para la teoría de juegos.

A pesar de estos desarrollos cada vez se hizo más evidente que existía una necesidad sustantiva de modelos analíticos que fueran más allá de los límites de la teoría de precios. Una motivación particular fue, la falta de resultados concluyentes en los debates entre socialismo y capitalismo, los cuales mostraron las limitaciones de la teoría de precios para evaluar instituciones que no eran de precios, como pasaba con la economía socialista. Esta teoría podía mostrar (bajo ciertas condiciones), que el libre mercado puede lograr una asignación eficiente, sin embargo, esto no quiere decir que las economías socialistas no puedan alcanzar resultados similares. Con la necesidad de comparar analíticamente las distintas formas de organización económica, fue que se buscó una nueva y más general infraestructura teórica. A raíz de esto podemos encontrarnos con que el inicio intelectual del diseño de mecanismos, parte de los socialistas utópicos como Robert Owen y Charles Fourier. Sin embargo, la Controversia de la Planeación fue un influencia más directa para el cambio a la teoría moderna. Partiendo de su mayor auge en las década de 1930 y considerando a los principales antagonistas, del lado socialista eran Oskar Lange y Abba Lerner, y del libre mercado Friederich von Hayek y Ludwig von Mises. Los partidarios del socialismo argumentaban que la planeación centralizada podía replicar los resultados del libre mercado, siempre que esto se hiciera correctamente (Maskin 2007). Sugerían que la planeación podía corregir las *fallas del mercado* - en particular las expuestas durante la Gran Depresión - y de esta manera poder sobrepasar al libre mercado. Los oponentes de estas nociones, negaban incondicionalmente que un sistema planeado pudiera aproximar el éxito del libre mercado (von Hayek 1944, von Mises 1920).

A pesar que este debate era, popularmente, visto únicamente como un enfrentamiento entre los defensores del capitalismo contra los del socialismo, como menciona Karen Vaughn (Vaughn 2004); en realidad, una gran parte era entre los miembros del bando socialista, quienes tenían diferentes posiciones acerca de la medida con la cual utilizar los conceptos de mercado y dinero en un sistema socialista, así como si la ley de valor podía o no seguirse utilizando.

No obstante del debate, hubo pocos avances, en el contexto del rigor matemático, durante los primeros cuarenta años. En este sentido, uno de estos avances fue la demostración de Leonid Kantorovich, de cómo una economía, sin considerar dinero, ni valor financiero, podía utilizar procedimientos matemáticos determinados para calcular la combinación de técnicas necesarias para alcanzar cierta producción y objetivos (Cockshott 2008).

Según Eric Maskin (Maskin 2007), la controversia era importante y fascinante, pero para ciertos espectadores, como Leonid Hurwicz, un tanto frustrante debido a la falta de precisión de los conceptos utilizados como son, descentralización, eficiencia, entre otros (Hurwicz 2007). Una razón era que no existía el aparato técnico - en particular, la teoría de juegos y la programación matemática - para poder generar estas definiciones. Inspirado por esto, Hurwicz se abocó a definir sin ambigüedades las ideas centrales, este esfuerzo tuvo como resultado sus artículos sobre el análisis de incentivos (Hurwicz 1960, 1972), en los que introduce una de las nociones clave del diseño de mecanismos, la compatibilidad con incentivos.

Otra de las razones por las que hubo este estancamiento, fue debido a que ninguno de los dos bandos hablaba el mismo idioma que el otro - parcialmente porque el lenguaje adecuado no había sido inventado. En este respecto, Filip Palda comenta (Palda 2013), que Myerson en particular, argumentaba que lo que hacía falta era un mejor entendimiento sobre los problemas de información que prevenían la coordinación entre la gente; enfatizaba que existen problemas particulares que quejan al capitalismo y otros que afectan al socialismo. Uno de estos problemas, común para ambos sistemas, es el monitoreo del desempeño de los administradores de una empresa. Los sistemas capitalistas se encargan de este problema pagándoles con acciones de la compañía, de tal manera que los intereses de ambos estén alineados, removiendo así la necesidad de estar monitoreándolos constantemente. Por el otro lado, el socialismo nunca uso esta estrategia, pero tampoco pudieron descifrar como monitorear adecuadamente el desempeño, esto resultaba en administradores con poder, pero sin responsabilidad. A esta situación, los economistas le llaman "riesgo moral". De hecho, la diseminada elusión de responsabilidades y obligaciones, así como el saqueo de las empresas fue una de las razones del colapso del sistema soviético. Pero, esto no quiere decir que el capitalismo esté libre de fallas, tiene un problema de información muy particular llamado por los economistas "selección adversa", que consiste en que los candidatos para puestos administrativos puede representar falsamente sus habilidades con tal de ser contratados, o los vendedores pueden mentir sobre la calidad de un producto con la finalidad de realizar una venta. El socialismo y el capitalismo tenían diferentes fortalezas y debilidades.

Una vez que la autoridad había implementado las reglas y esquemas para la obtención de un beneficio, el diseño de mecanismos vio estas como interacciones estratégicas que podían ser modeladas como juegos, de tal forma que se pudieran manipular para que las personas involucradas se portaran honesta y obedientemente. Uno de los problemas de esta noción es que, la dificultad del gobierno o la autoridad para diseñar las reglas, además del costo de hacer que todos participen y que se comporten como se espera, puede ser mayor que los beneficios obtenidos. De esta manera, al fusionar la teoría de juegos y la economía de la información, el diseño de mecanismos generó el lenguaje o marco teórico en el cual ambos lados podían comparar los méritos de sus argumentos. Adentrarme más en este debate no es el objetivo de este trabajo, pero sí es importante entender que estas preguntas sobre cómo la gente usaba la información para coordinar sus acciones sigue estando válida y presente en el desarrollo de la teoría de juegos y del diseño de mecanismos.

A pesar de que el trabajo inspirado por Hurwicz y otros ha demostrado que para escenarios en donde (i) exista un gran número de compradores y vendedores, de tal manera que ninguno tenga suficiente poder de mercado y (ii) no existan externalidades significativas, esto es, el consumo, producción e información de un agente no afecte la producción o consumo de otros; el mercado es el "mejor" mecanismo, existen mecanismos que mejoran al mercado, simplemente con violar alguna de estas dos suposiciones.

Partiendo de las investigaciones generales, se dio pie a una gran cantidad de literatura, la cual, en general, se encaminó en dos direcciones. Por un lado están las investigaciones que hacen uso de escenarios especiales, altamente estructurados, con el fin de estudiar preguntas particulares, como son, la manera de asignar bienes públicos, el diseño de subastas y la estructuración de contratos. Por otro lado, ha habido investigaciones sobre resultados generales y abstractos que hacen el menor número de suposiciones posibles para obtener una conclusión, como son: estudios sobre los espacios de soluciones posibles para un mecanismo, las implicaciones generales del principio de revelación, entre otros.

Siguiendo estas líneas de investigación, en particular la primera, surge el teorema de Myerson-

Satterthwaite, que fue demostrado en 1983 en el artículo “Efficient Mechanisms for Bilateral Trading” (Myerson 1983), en donde se ocupan de caracterizar los problemas de negociación bilateral entre un comprador y un vendedor para un único bien. La conclusión que se puede obtener del teorema es que bajo ciertos supuestos no existe una manera, llámese un precio, en que el comprador y el vendedor estén dispuestos a realizar el intercambio, de tal manera que los dos estén “razonablemente de acuerdo” con el precio. Esto es interesante, porque una vez más encontramos un resultado que denota las dificultades que existen para la toma de decisiones en escenarios en donde la información es primordial.

Parte II

El Teorema de Imposibilidad de Arrow

Capítulo 1

Relaciones de Preferencia y Funciones de Utilidad

Antes de poder entrar de lleno a la teoría de elección social, tenemos que definir unos conceptos que ocuparemos a lo largo de este trabajo, las relaciones de preferencia y las funciones de utilidad. Estos parten del estudio de la teoría de las decisiones individuales en un escenario completamente abstracto. El punto de partida de cualquier problema de decisión individual es el *conjunto de posibles alternativas (mutuamente excluyentes)* de donde el individuo debe de escoger. De aquí en adelante denotaremos al conjunto de alternativas como X. Por el momento, este conjunto puede ser cualquier cosa, cualquier conjunto que conste de posibles alternativas para tomar una decisión.

Lo primero que haremos será usar los gustos del tomador de decisión, los cuales están resumidos en su *relación de preferencia*, como las características deseables de los individuos. La teoría se desarrolla imponiendo unos axiomas de racionalidad a las preferencia del tomador de decisión, para después analizar las consecuencias de estas en la decisión realizada. Este manejo basado en preferencia es el que usaremos a lo largo del trabajo.

Ahora, los objetivos del tomador de decisión se encuentran resumidos en una *relación de preferencia*, la cual denotamos por $\succeq$. Técnicamente, $\succeq$ es una relación binaria en el conjunto de alternativas, X, que permite una comparación de parejas de alternativas $x, y \in X$. Se lee $x \succeq y$ como "x es al menos igual de preferida que y". A partir de $\succeq$ podemos obtener dos relaciones en X:

(i) La *relación de preferencia estricta*, $\succ$ definida como:

$$x \succ y \iff x \succeq y \text{ pero no } y \succeq x$$

y se lee "x es preferido a y."

(ii) La *relación de indiferencia*, $\sim$, definido como:

$$x \sim y \iff x \succeq y \text{ y } y \succeq x$$

y se lee "x es indiferente a y."

En gran parte de la teoría microeconómica, se asume que las preferencias individuales son *racionales.* La hipótesis de racionalidad está representada en dos suposiciones básicas de la relación de preferencia $\succeq$: *completez y transitividad.* [1]

Definición 1.1. *La relación de preferencia $\succeq$ es* ***racional*** *sí cumple las siguientes propiedades:*

(i) ***Completez:*** $\forall, x, y \in X$ *tenemos que* $x \succeq y$ *o* $y \succeq x$ *(o las dos).*

(ii) ***Transitividad:*** $\forall, x, y, z \in X$, *sí* $x \succeq y$, $y \succeq z$ *entonces,* $x \succeq z$.

El supuesto que $\succeq$ es completa dice que las preferencias entre dos posibles alternativas están bien definidas. La fuerza del supuesto de completez no debe de ser subestimado. El axioma de completez nos dice que los tomadores de decisión solamente toman decisiones previamente meditadas.

La transitividad también es una suposición fuerte, esta implica que es imposible enfrentar al tomador de decisión a una serie de parejas de alternativas en donde sus preferencias aparenten un ciclo. Por ejemplo, sentir que una manzana es al menos tan buena como un plátano y que un plátano es al menos tan buena como una naranja, pero luego preferir la naranja a la manzana. Comparada con la propiedad de completez, es más importante en el sentido que partes de la teoría económica no sobrevivirían sí no se pudiera suponer que los agentes económicos tengan preferencias transitivas.

[1] Como dice en Mas-Colell (1995), nótese que no hay una terminología unificada en la literatura; *orden débil* y *preorden completo* son alternativas comunes al término *relación de preferencia racional.* También, en algunas presentaciones, se agrega el supuesto que $\succeq$ es *reflexiva* (definida como $x \succeq x, \forall x \in X$) a los de completez y transitividad. Esta propiedad está implicita a partir de la completez, por lo tanto es redundante.

La suposición de que la relación de preferencia $\succeq$ es completa y transitiva tiene implicaciones para las relaciones de preferencia estricta $\succ$ y la de indiferencia $\sim$. Estas las resumimos en la siguiente proposición.

Proposición 1.1. *sí* $\succeq$ *es racional, entonces:*
(i) $\succ$ *es* ***irreflexiva*** *(*$x \succ x$ *nunca sucede),* ***transitiva*** *(sí* $x \succ y$*,* $y \succ z$*, entonces* $x \succ z$*).*
(ii) $\sim$ *es* ***reflexiva*** *(*$x \sim x, \forall x$*),* ***transitiva*** *(sí* $x \sim y$*,* $y \sim z$*, entonces* $x \sim z$*),* ***simétrica*** *(sí* $x \sim y$*, entonces* $y \sim x$*).*
(iii) sí $x \succ y \succeq z$*, entonces* $x \succ z$*.*

La irreflexividad de $\succ$ y, la reflexividad y la simetría de $\sim$ son propiedades sensibles para las relaciones de preferencia estricta y las de indiferencia. Un punto más importante de la proposición anterior es que la racionalidad de $\succeq$ implica que $\succ$ y $\sim$ son transitivas. Además, $\succ$ tiene una propiedad cuasi-transitiva cuando se combina con una relación como $\succeq$.

En economía a menudo se describen relaciones de preferencia a través de una *función de utilidad*. Una función de utilidad $u(x)$ asigna un valor numérico a cada elemento de X, ordenando los elementos de X de acuerdo con las preferencias individuales.

Definición 1.2. *Una función* $u : X \to \mathbb{R}$ *es una función de utilidad que representa una relación de preferencia* $\succeq$ *sí, para todo* $x, y \in X$

$$x \succeq y \iff u(x) \geq u(y)$$

Notemos que una función de utilidad que representa una relación de preferencia $\succeq$ no es única. Para cualquier función estrictamente creciente $f : \mathbb{R} \to \mathbb{R}$, $v(x) = f(u(x))$ es una nueva función de utilidad que representa las mismas preferencias que $u(\cdot)$. Lo más importante es el orden de las alternativas. Las propiedades de las funciones de utilidad que son invariantes bajo cualquier transformación estrictamente creciente se llaman *ordinales*. Las propiedades *cardinales* son aquellas que no son preservadas bajo este tipo de transformaciones. Por lo tanto, la relación de preferencia asociada con una función de utilidad es una propiedad ordinal. Por otro lado, los valores numéricos asociados con las alternativas de X, y por lo tanto, la magnitud de cualquier diferencia en la medida de utilidad entre alternativas son propiedades cardinales.

La habilidad para representar preferencias a través de una función de utilidad está estrechamente ligada al supuesto de racionalidad. En particular, tenemos el siguiente resultado.

Proposición 1.2. *Una relación de preferencia $\succeq$ puede representarse por una función de utilidad solo sí es racional*

Demostración. Para hacer esta demostración vamos a mostrar que sí existe una función de utilidad que representa preferencias $\succeq$, entonces $\succeq$ debe ser completa y transitiva.

Completa. Como $u(\cdot)$ es una función real definida en X, tenemos que para cualquier $x, y \in X$, debe pasar alguna de las siguientes $u(x) \geq u(y)$ o $u(y) \geq u(x)$. Pero como $u(\cdot)$ es una función de utilidad que representa a $\succeq$, entonces $x \succeq y$ o $y \succeq x$ (por la definición anterior). Por lo tanto, $\succeq$ debe de ser completa.

Transitiva. Supongamos que $x \succeq y$ y $y \succeq z$. Como $u(\cdot)$ representa $\succeq$, debemos tener que $u(x) \geq u(y)$ y $u(y) \geq u(z)$. Por lo tanto, $u(x) \geq u(z)$. Como $u(\cdot)$ representa $\succeq$, esto implica que $x \succeq z$. Por lo tanto $\succeq$ es transitiva. $\square$

De aquí en adelante cuando hablemos de relaciones de preferencia y funciones de utilidad, estamos hablando de este tipo de relaciones y funciones. En el siguiente capítulo usaremos las relaciones de preferencia en el contexto de la teoría de elección social.

Capítulo 2

La Teoría de Elección Social: Un caso especial

Empezaremos nuestro análisis de la elección social con el caso más simple: en el que solo hay dos alternativas sobre las cuales decidir. Llamaremos a estas alternativa x y alternativa y. La alternativa x, por ejemplo puede ser el "status quo" y la alternativa y puede ser un proyecto público en particular que se está contemplando implementar.

La información para nuestro problema son las preferencia individuales de los miembros de la sociedad sobre las dos alternativas. Asumimos que existe un número $I < \infty$ de individuos o *agentes*. La familia de las preferencias individuales entre dos alternativas puede describirse con un perfil

$$(\alpha_1, \ldots, \alpha_I) \in \mathbb{R}^I$$

en donde α_i toma los valores 1, 0 o -1 de acuerdo a sí el agente i prefiere la alternativa x sobre la alternativa y, es indiferente entre las dos, o prefiere la alternativa y a la alternativa x respectivamente.

Definición 2.1. *Un* ***funcional de bienestar social o (acumulador de bienestar social)*** *es una regla $F(\alpha_1, \ldots, \alpha_I)$ que asigna una preferencia social, esto es, $F(\alpha_1, \ldots, \alpha_I) \in \{-1, 0, 1\}$ a cada posible perfil de preferencias individuales $(\alpha_1, \ldots, \alpha_I) \in \{-1, 0, 1\}^I$.*

Todos los funcionales de bienestar social que consideraremos respetan las preferencias individuales en el sentido débil de la Definición 2.2

Definición 2.2. *El funcional de bienestar social $F(\alpha_1, \ldots, \alpha_I)$ es **de Pareto** o tiene la **propiedad de Pareto**, sí respeta la unanimidad de las preferencia estrictas de parte de los agentes, esto es, sí $F(1, \ldots, 1) = 1$ y sí $F(-1, \ldots, -1) = -1$*

Ejemplo 2.1. *Funcionales de bienestar social de Pareto abundan. Sea $(\beta_1, \ldots, \beta_I) \in \mathbb{R}^I$ un vector no negativo, no todos cero. Entonces podemos definir*

$$F(\alpha_1, \ldots, \alpha_I) = signo \sum_i \beta_i \alpha_i$$

en donde, para todo $\alpha \in \mathbb{R}$, signo α es igual 1, 0 o -1 dependiendo de sí $\alpha > 0$, $\alpha = 0$ o $\alpha < 0$, respectivamente. Podemos considerar a β_i como una medida de poder del jugador i

*Un caso particular importante es **voto por mayoría**, en donde tomamos $\beta_i = 1$ para toda i. Entonces $F(\alpha_1, \ldots, \alpha_I) = 1$ sí y solo sí el número de agentes que prefieren la alternativa x a la y es mayor que el número de agentes que prefieren y a x. Similarmente, $F(\alpha_1, \ldots, \alpha_I) = -1$ sí y solo sí aquellos que prefieren y a x son más numerosos que aquellos que prefieren x a y. Finalmente, en caso de que sean iguales estos dos números, tenemos $F(\alpha_1, \ldots, \alpha_I) = 0$, esto es, indiferencia social.*

Ejemplo 2.2 (Dictadura)**.** *Decimos que un funcional de bienestar social es **dictatorial** sí existe un agente h llamado **dictador**, tal que, para cualquier perfil $(\alpha_1, \ldots, \alpha_I), \alpha_h = 1$ implica que $F(\alpha_1, \ldots, \alpha_I) = 1$ y similarmente $\alpha_h = -1$ implica que $F(\alpha_1, \ldots, \alpha_I) = -1$. Esto es, la preferencia estricta del dictador prevalece como la preferencia social. Un funcional de bienestar social dictatorial es de Pareto en el sentido de la Definición 2.2. Para el funcional de bienestar social del Ejemplo 2.1, tenemos una dictadura siempre que $\alpha_h > 0$ para algún agente h y $\alpha_i = 0$ para $i \neq h$, entonces $F(\alpha_1, \ldots, \alpha_I) = \alpha_h$.*

El funcional de bienestar social de votación por mayoría juega un rol de punto de referencia en la teoría de elección social. Además de ser de Pareto tiene tres propiedades importantes, las cuales procederemos a establecer formalmente. La primera (simetría entre los agentes) dice que el funcional de bienestar social trata a todos los agentes desde la misma base. La segunda (neutralidad entre alternativas) dice que, similarmente, el funcional de bienestar social no distingue a priori entre dos alternativas. La tercera (sensibilidad positiva) dice que, más fuerte que la propiedad de Pareto de la Definición 2.2, el funcional de bienestar social es sensible a las preferencia individuales.

Definición 2.3. *El funcional de bienestar social $F(\alpha_1, \ldots, \alpha_I)$ es **simétrico entre los agentes (o anónimo)** sí los nombres de los mismos no importan, esto es, sí una permutación de preferencia sobre los agentes no altera las preferencia sociales. Precisando, sea $\pi : \{1, \ldots, I\} \to \{1, \ldots, I\}$ sea una función sobre (i.e. una función que para cualquier i existe una h tal que $\pi(h) = i$). Entonces para cualquier perfil $(\alpha_1, \ldots, \alpha_I)$ tenemos $F(\alpha_1, \ldots, \alpha_I) = F(\alpha_{\pi(1)}, \ldots, \alpha_{\pi(I)})$.*

Definición 2.4. *El funcional de bienestar social $F(\alpha_1, \ldots, \alpha_I)$ es **neutral entre alternativas** sí $F(\alpha_1, \ldots, \alpha_I) = -F(-\alpha_1, \ldots, -\alpha_I)$ para todo perfil $(\alpha_1, \ldots, \alpha_I)$, esto es, las preferencia sociales se invierten cuando invertimos las preferencia de todos los agentes.*

Definición 2.5. *El funcional de bienestar social $F(\alpha_1, \ldots, \alpha_I)$ es **sensible positivamente** sí siempre que $(\alpha_1, \ldots, \alpha_I) \geq (\alpha'_1, \ldots, \alpha'_I)$. $(\alpha_1, \ldots, \alpha_I) \neq (\alpha'_1, \ldots, \alpha'_I)$ y $F(\alpha'_1, \ldots, \alpha'_I) \geq 0$ tenemos $F(\alpha'_1, \ldots, \alpha'_I) = +1$. Esto es, sí x es socialmente preferido o indiferente a y, además de que algunos agentes aumentan su consideración de x, entonces x se vuelve socialmente preferida.*

A continuación demostraremos rápidamente que la votación mayoritaria cumple las tres propiedades anteriores.

Proposición 2.1. *La votación mayoritaria entre dos alternativas es simétrica entre los agentes, neutral entre alternativas y sensiblemente positiva*

Demostración. (i) Simetría: Sea $\pi : \{1, \ldots, I\} \to \{1, \ldots, I\}$ una permutación (sobre), entonces

$$\sum_{i=1}^{I} \alpha_i = \sum_{i=1}^{I} \alpha_{\pi(i)}$$

lo que implica que

$$Signo\left[\sum_{i=1}^{I} \alpha_i\right] = Signo\left[\sum_{i=1}^{I} \alpha_{\pi(i)}\right]$$

lo que implica que

$$F(\alpha_1, \ldots, \alpha_I) = F(\alpha_{\pi(1)}, \ldots, \alpha_{\pi(I)})$$

(ii) Neutralidad:

$$F(\alpha_1, \ldots, \alpha_I) = Signo[\sum_{i=1}^{I} \alpha_i] = Signo[-\sum_{i=1}^{I}(-\alpha_i)] =$$
$$-Signo[\sum_{i=1}^{I}(-\alpha_i)] = -F(-\alpha_1, \ldots, -\alpha_I).$$

(iii) Sensibilidad Positiva: Asumamos que $F(\alpha_1, \ldots, \alpha_I) \geq 0$, entonces

$$Signo\left[\sum_{i=1}^{I} \alpha_i\right] \geq 0$$

lo que implica que

$$\sum_{i=1}^{I} \alpha_i \geq 0$$

Tomemos $(\alpha'_1, \ldots, \alpha'_I) \geq (\alpha_1, \ldots, \alpha_I)$ tal que $(\alpha'_1, \ldots, \alpha'_I) \neq (\alpha_1, \ldots, \alpha_I)$. Entonces, $\sum_{i=1}^{I} \alpha'_i > 0$, lo que implica que $Signo[\sum_{i=1}^{I} \alpha'_i] > 0$ y esto implica que $F(\alpha'_1, \ldots, \alpha'_I) = 1$. $\square$

A continuación veremos la demostración que hace Kenneth May (May 1952), que prueba que cualquier funcional de bienestar social que cumpla las tres propiedades tiene que ser una votacion mayoritaria.

Proposición 2.2 (Teorema de May)**.** *Un funcional de bienestar social $F(\alpha_1, \ldots, \alpha_I)$ es un funcional de bienestar social de votación mayoritaria sí y solo sí es simétrica entre los agentes, neutral entre alternativas y sensiblemente positiva.*

Demostración. Ya probamos que la votación mayoritaria satisface las tres propiedades. Para establecer la suficiencia notemos que la propiedad de simetría entre los agentes significa que la preferencia social depende solamente del número de agentes que prefieren la alternativa x sobre la y, el número total de los que les es indiferente y el total de los que prefieren y sobre x. Dado $(\alpha_1, \ldots, \alpha_I)$ denotamos

$$n^+(\alpha_1, \ldots, \alpha_I) = \#\{i : \alpha_i = 1\}$$

y

$$n^-(\alpha_1, \ldots, \alpha_I) = \#\{i : \alpha_i = -1\}$$

Entonces, la simetría permite que expresemos $F(\alpha_1, \ldots, \alpha_I)$ de la siguiente forma

$$F(\alpha_1, \ldots, \alpha_I) = G(n^+(\alpha_1, \ldots, \alpha_I), n^-(\alpha_1, \ldots, \alpha_I)).$$

Ahora, supongamos que $(\alpha_1, \ldots, \alpha_I)$ es tal que $n^+(\alpha_1, \ldots, \alpha_I) = n^-(\alpha_1, \ldots, \alpha_I)$.
Entonces $n^+(-\alpha_1, \ldots, -\alpha_I) = n^-(\alpha_1, \ldots, \alpha_I) = n^+(\alpha_1, \ldots, \alpha_I) = n^-(-\alpha_1, \ldots, -\alpha_I)$, así que

$$\begin{aligned} F(\alpha_1, \ldots, \alpha_I) &= G(n^+(\alpha_1, \ldots, \alpha_I), n^-(\alpha_1, \ldots, \alpha_I)) \\ &= G(n^+(-\alpha_1, \ldots, -\alpha_I), n^-(-\alpha_1, \ldots, -\alpha_I)) \\ &= F(-\alpha_1, \ldots, -\alpha_I) \\ &= -F(\alpha_1, \ldots, \alpha_I) \end{aligned}$$

La última igualdad se sigue de la neutralidad entre alternativas. Como el único número igual a su negativo es cero, concluimos que sí $n^+(\alpha_1, \ldots, \alpha_I) = n^-(\alpha_1, \ldots, \alpha_I)$, entonces $F(\alpha_1, \ldots, \alpha_I) = 0$.

Ahora supongamos que $n^+(\alpha_1, \ldots, \alpha_I) > n^-(\alpha_1, \ldots, \alpha_I)$. Denotamos $H = n^+(\alpha_1, \ldots, \alpha_I)$, $J = n^-(\alpha_1, \ldots, \alpha_I)$, entonces $J < H$, tal que $G(H, J) = F(\alpha_1, \ldots, \alpha_I)$. Digamos, sin pérdida de generalidad, que $\alpha_i = 1$ para $i \leq H$ y $\alpha_i \leq 0$ para $i > H$. Consideremos un nuevo perfil $(\alpha'_1, \ldots, \alpha'_I)$ definido por $\alpha'_i = \alpha_i = 1$ para $i \leq J < H$, $\alpha'_i = 0$ para $J < i \leq H$ y $\alpha'_i \alpha \leq 0$ para $i > H$. Entonces $F(\alpha'_1, \ldots, \alpha'_I) = 0$. Pero por construcción la alternativa x perdió fuerza en las nuevas preferencia individuales. De hecho, $(\alpha_1, \ldots, \alpha_I) \geq (\alpha'_1, \ldots \alpha'_I)$ y $\alpha_{J+1} = 1 > 0 = \alpha'_{J+1}$. Por lo tanto, por la propiedad de sensibilidad positiva, debemos tener que $F(\alpha_1, \ldots, \alpha_I) = 1$.

Cuando $n^-(\alpha_1, \ldots, \alpha_I) > n^+(\alpha_1, \ldots, \alpha_I)$ entonces $n^+(-\alpha_1, \ldots, -\alpha_I) > n^-(-\alpha_1, \ldots, -\alpha_I)$ y entonces $F(-\alpha_1, \ldots, -\alpha_I) = 1$. Por lo tanto, por la neutralidad entre alternativas:

$$F(\alpha_1, \ldots, \alpha_I) = -F(-\alpha_1, \ldots, -\alpha_I) = -1$$

Concluimos que $F(\alpha_1, \ldots, \alpha_I)$ es de hecho un funcional de bienestar social de votación mayoritaria. $\square$

Ya que hemos establecido las bases del ambiente teórico en el que vamos a trabajar, podemos pasar de lleno a uno de nuestros importantes resultados, el teorema de imposibilidad de Arrow.

Capítulo 3

El Caso General: El Teorema de Imposibilidad de Arrow

Ahora procederemos a estudiar el problema de la agregación de preferencia individuales sobre cualquier número de alternativas. Denotaremos al conjunto de alternativas como X y asumiremos que hay I agentes, indicados por $i = 1, \ldots, I$. Cada agente i tiene una relación de preferencia racional $\succeq_i$ definida sobre X. Tanto la preferencia estricta como la relación de indiferencia derivadas de $\succeq_i$ se denotan por $\succ_i$ y $\sim_i$ respectivamente. Definimos $\succ_i$ dejando que $x \succ_i y$ sí $x \succeq_i y$ se cumpla pero $y \succeq_i x$ no. Esto quiere decir, que x se prefiere a y sí x es al menos tan bueno como y pero y no es tan bueno como x. También, la relación de indiferencia $\sim_i$ se define como $x \sim_i y$ sí $x \succeq_i y$ y $y \succeq_i x$. Además, será conveniente, a menudo, asumir que no hay alternativas distintas que sean indiferentes en una relación de preferencia individuales $\succeq_i$. Por lo tanto, es importante para claridad de la explicación, tener un símbolo para el conjunto de todas las relaciones de preferencia posibles en X, $\mathcal{R}$, y para el conjunto de las posibles relaciones de preferencia en X con la propiedad de que no tienen preferencia indiferentes, $\mathcal{P}$, en donde $\mathcal{P} \subset \mathcal{R}$.

Así como en la sección 2, podemos definir un funcional de elección social como una regla que asigna preferencia sociales a perfiles de preferencia individuales $(\succeq_1, \ldots, \succeq_I) \in \mathcal{R}^I$. La Definición 3.1 generaliza la Definición 2.1 en dos aspectos; admite cualquier número de alternativas, no solamente dos, y permite que el problema de agregación se limite a un dominio dado $\mathcal{A} \subset \mathcal{R}^I$

de perfiles individuales. Aquí nos concentraremos en los dominios más grandes, esto es, cuando $\mathcal{A} = \mathcal{R}^I$ y $\mathcal{A} = \mathcal{P}$.

Definición 3.1. *Un* ***funcional de bienestar social (o acumulador de bienestar social)*** *definido en $\mathcal{A} \subset \mathcal{R}^I$ es una regla $F : \mathcal{A} \to \mathcal{R}^I$ que a cada relación de preferencia racional asigna una relación de preferencia racional $F(\succeq_1, \ldots, \succeq_I) \in \mathcal{R}$. La cual se interpreta como la relación de preferencia social, para cualquier perfil de relaciones de preferencia racional individual $(\succeq_1, \ldots, \succeq_I)$ en el dominio permitido $\mathcal{A} = \mathcal{R}^I$.*

Al igual que en la Sección 2, los individuos están descritos exclusivamente por sus relaciones de preferencia sobre las alternativas.

Para cualquier perfil $(\succeq_1, \ldots, \succeq_I)$, denotamos por $F_p(\succeq_1, \ldots, \succeq_I)$ la relación de preferencia estricta derivada de $F(\succeq_1, \ldots, \succeq_I)$. Esto es decimos que "$x$ es socialmente preferida a y" cuando $xF_p(\succeq_1, \ldots, \succeq_I)y$ sí se cumple $xF(\succeq_1, \ldots, \succeq_I)y$, pero no se cumple $yF(\succeq_1, \ldots, \succeq_I)x$. Leeremos $xF(\succeq_1, \ldots, \succeq_I)y$ como "x es socialmente tan buena como y".

Definición 3.2. *El funcional de bienestar social $F : \mathcal{A} \to \mathcal{R}$ es* ***de Pareto*** *sí, para cualquier par de alternativas $\{x, y\} \subset X$ y cualquier perfil de preferencia $(\succeq_1, \ldots, \succeq_I) \in \mathcal{A}$, tenemos que x es preferida socialmente a y, esto es, $xF_p(\succeq_1, \ldots, \succeq_I)y$ siempre que $x \succ_i y$ para toda i*

Ahora, estableceremos una restricción importante a funcionales de bienestar social que fue sugerida originalmente por Arrow (1963). Esta dice que las preferencia sociales entre cualesquiera dos alternativas depende solamente de la preferencia individual entre estas mismas alternativas. Existen tres líneas posibles de justificación para esta suposición. La primera es estrictamente normativa y tiene un atractivo considerable: argumenta que cuando se establece un orden social entre x y y, la presencia o ausencia de alternativas diferentes a x y y no debería importar, ya que son irrelevantes al problema en cuestión. La segunda es una cuestión práctica. La suposición facilita enormemente la tarea de tomar decisiones sociales ya que ayuda a separar problemas. La determinación de un orden social en un subconjunto de alternativas, no necesita de ninguna información sobre preferencia individuales sobre alternativas fuera del subconjunto. La tercera tiene que ver con incentivos y la trataremos en el siguiente capítulo. La independencia en parejas está íntimamente conectada con el problema de proveer el aliciente adecuado para la revelación honesta de las preferencia individuales.

Definición 3.3. *El funcional de bienestar social* $F : \mathcal{A} \rightarrow \mathcal{R}$ *definida en el dominio* $\mathcal{A}$ *satisface la* ***condición de independencia en parejas (o condición de independencia de alternativas irrelevantes)*** *sí la preferencia social entre cualquiera dos alternativas* $\{x, y\} \subset X$ *dependen solamente del perfil de preferencia individuales sobre las mismas alternativas. Es decir, para cualquier par de alternativas* $\{x, y\} \subset X$ *y para cualquier par de perfiles de preferencia* $(\succeq_1, \ldots, \succeq_I) \in \mathcal{A}$ *y* $(\succeq'_1, \ldots, \succeq'_I) \in \mathcal{A}$ *con la propiedad de que para cada* i*,*

$$x \succeq_i y \iff x \succeq'_i y$$

$$y \succeq_i x \iff y \succeq'_i x.$$

tenemos que

$$xF(\succeq_1, \ldots, \succeq_I)y \iff xF(\succeq'_1, \ldots, \succeq'_I)y$$

y

$$yF(\succeq_1, \ldots, \succeq_I)x \iff yF(\succeq'_1, \ldots, \succeq'_I)x$$

A continuación daremos un camino que garantizará automáticamente que se satisface la independencia de alternativas irrelevantes, el cual consiste en determinar la preferencia social entre cualquiera dos alternativas dadas aplicando una regla de agregación que usa solamente la información sobre el orden de esas dos alternativas en preferencia individuales. Vimos en la Sección 2 que, para cualquier par de alternativas, esto se puede hacer. ¿Podremos proceder de esta manera y aún así terminar con preferencia sociales que sean racionales, esto es, completas y transitivas?

Ejemplo 3.1 (La Paradoja de Condorcet)**.**

Supongamos que queremos probar la votación por mayoría de dos alternativas cualquiera. ¿Esto determina un funcional de bienestar social? Veremos que nos encontramos con el siguiente problema. Tengamos tres alternativas $\{x, y, z\}$ y tres agentes. Las preferencia de los tres agentes son

$$x \succ_1 y \succ_1 z$$

$$z \succ_2 x \succ_2 y$$

$$y \succ_3 z \succ_3 x$$

Entonces la votación por mayoría en parejas nos dice que x debe ser preferida socialmente a y (ya que x tiene una mayoría sobre y y, con mayor razón, y no tiene una mayoría sobre x). Similarmente, y debe de ser preferida socialmente a z (dos votantes prefieren y a z) y z debe ser preferida a x (dos votantes prefieren z sobre x). Pero tenemos que este patrón cíclico viola el requisito de transitividad en preferencias sociales.

De aquí podemos concluir un pequeño resultado que nos dice que las preferencia individuales no necesariamente inducen o llevan a preferencia racionales.

La siguiente proposición es el **Teorema de Imposibilidad de Arrow**, el cual básicamente nos dice que la Paradoja de Condorcet no se debe a ninguna de las propiedades fuertes de la votación mayoritaria (las cuales, recordando de la Proposición 2.2 son simetría entre los agentes, neutralidad entre alternativas y sensibilidad positiva). La paradoja se enfoca en lo central del asunto: con independencia de alternativas irrelevantes no existe un funcional de bienestar social definido en $\mathcal{R}^I$ que satisfaga una mínima forma de simetría entre los agentes, es decir que no haya dictadores, y una forma mínima de sensibilidad positiva, o sea que cumpla la propiedad de Pareto.

Teorema 3.1 (Teorema de Imposibilidad de Arrow)**.** *Supongamos que el número de alternativas es al menos tres y que el dominio de perfiles individuales admisibles, denotado por $\mathcal{A}$, es $\mathcal{A} = \mathcal{R}^I$ o $\mathcal{A} = \mathcal{P}^I$. Entonces cada funcional de bienestar social $F : \mathcal{F} \to \mathcal{R}$ que sea de Pareto y satisfaga la condición de independencia de alternativas irrelevantes es* ***dictatorial*** *en el siguiente sentido: Existe un agente h tal que, para todo $\{x, y\} \subset X$ y cualquier perfil $(\succeq_1, \ldots, \succeq_I) \in \mathcal{A}$, tenemos que x es socialmente preferida a y, esto es, $xF(\succeq_1, \ldots, \succeq_I)y$, siempre que $x \succ_h y$.*

La demostración que haremos será más visual que la que se utiliza usualmente, en particular la que podemos encontrar en Mas-Colell (1995). La idea final detrás de esta demostración se hará aparente en la Sección 4, al ver que la demostración del Teorema 5.1, tiene fundamentos lógicos idénticos (Reny (2000)).

Demostración. De ahora en adelante veremos a I no solamente como el número sino también el conjunto de agentes. Durante toda la demostración nos referiremos a un funcional de bienestar social fijo $F : \mathcal{A} \to \mathcal{R}$ que satisface las condiciones de Pareto y de independencia de alternativas irrelevantes. Para empezar, necesitamos algunas definiciones. De aquí en adelante cuando nos refiramos a parejas de alternativas siempre estaremos hablando de alternativas distintas.

Paso 1: Consideremos cualesquiera dos alternativas $x, y \in X$ y un perfil de preferencia $(\succeq_1, \dots, \succeq_I)$ en donde $x \succeq_i z \succeq_i y, \forall z \in X, \forall i = 1 \dots, I$. Por la propiedad de Pareto de la Definición 3.2 tenemos que x esta estrictamente en el tope de las preferencia sociales.

Consideremos ahora, cambiar las preferencia del agente 1, movamos la alternativa y un lugar hacia arriba. Por la condición de independencia de alternativa irrelevante (IAI) (Definición 3.3), x se mantiene en el tope de las preferencia sociales siempre que y esté por debajo de x en las preferencia del agente 1. Pero cuando y finalmente sobrepasa a x (es decir, $y \succeq_1 x$), entonces IAI implica que $x \succeq_1 y \succ_1 z, \forall z \in X$, lo que quiere decir que en las preferencia sociales x se prefiere a todas las alternativas, menos tal vez y. sí $xF(\succeq_1, \dots, \succeq_I)y$, entonces hacemos los mismo para el agente 2 y luego el 3, etc. Esto hasta que encontremos un agente n, para el cual las preferencia sociales de y cambien con respecto a x, es decir $yF(\succeq_1, \dots, \succeq_I)x$ dado que $y \succeq_n x$. Debe existir este individuo n ya que la alternativa y, al fin y al cabo estará en el tope de todas las preferencia individuales y por la propiedad de Pareto entonces se debe de cumplir que $yF(\succeq_!, \dots, \succeq_I)x$ ya que $x \succeq_i y, \forall i \in I$. sí vemos el Cuadro 3-1 y el Cuadro 3-2 que representan estas situaciones, antes y después del cambio en las preferencia del agente n.

$$
\begin{array}{ccccccc|ccc}
\succeq_1 & \dots & \succeq_{n-1} & \succeq_n & \succeq_{n+1} & \dots & \succeq_I & & & \text{Preferencias Sociales} \\
\text{y} & \dots & \text{y} & \text{x} & \text{x} & \dots & \text{x} & & & \text{x} \\
\text{x} & \dots & \text{x} & \text{y} & \cdot & & \cdot & & & \cdot \\
\cdot & & \cdot & \cdot & \cdot & & \cdot & \rightarrow & & \cdot \\
\cdot & & \cdot & \cdot & \cdot & & \cdot & & & \text{y} \\
\cdot & & \cdot & \cdot & \text{y} & \dots & \text{y} & & & \cdot
\end{array}
$$

Cuadro 3-1:

$$
\begin{array}{ccccccc|ccc}
\succeq_1 & \dots & \succeq_{n-1} & \succeq_n & \succeq_{n+1} & \dots & \succeq_I & & & \text{Preferencias Sociales} \\
\text{y} & \dots & \text{y} & \text{y} & \text{x} & \dots & \text{x} & & & \text{y} \\
\text{x} & \dots & \text{x} & \text{x} & \cdot & & \cdot & & & \text{x} \\
\cdot & & \cdot & \cdot & \cdot & & \cdot & \rightarrow & & \cdot \\
\cdot & & \cdot & \cdot & \cdot & & \cdot & & & \cdot \\
\cdot & & \cdot & \cdot & \text{y} & \dots & \text{y} & & & \cdot
\end{array}
$$

Cuadro 3-2:

Paso 2: Ahora fijémonos en el Cuadro 3-3 y en el Cuadro 3-4. La 3-3 se obtiene del Cuadro

3-1 (y el Cuadro 3-4 del Cuadro 3-2) mediante el cambio de la alternativa x al fondo de las preferencia de los agentes i con $i < n$ y al penúltimo lugar para los agentes $i > n$. Con esto quisiéramos argumentar que estos cambios no afectan las alternativas más elevadas en las preferencia sociales y que estas se preservan para cada uno de los dos casos.

$\succeq_1$	...	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	...	$\succeq_I$		Preferencias Sociales
y	...	y	x	·	...	·		x
·	...	·	y	·		·		y
·		·	·	·		·	$\rightarrow$	·
·		·	·	x		x		·
x	...	x	·	y	...	y		·

Cuadro 3-3:

$\succeq_1$	...	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	...	$\succeq_I$		Preferencias Sociales
y	...	y	x	·	...	·		y
·	...	·	y	·		·		·
·		·	·	·		·	$\rightarrow$	x
·		·	·	x		x		·
x	...	x	·	y	...	y		·

Cuadro 3-4:

Primero, notemos que y, por IAI, será la preferencia social más alta en el Cuadro 3-4 ya que es la más alta en el Cuadro 3-2 y ninguna de las preferencia de los agentes cambia el lugar de y en el movimiento del Cuadro 3-2 al Cuadro 3-4. Después, notemos que los perfiles $(\succeq_1, \ldots, \succeq_I)$ en los Cuadros 3-3 y en 3-4 se encuentra en la preferencia relativa de x y y en $\succeq_n$. Entonces, por IAI, y debe mantenerse socialmente preferido a cualquier otra alternativa, a excepción tal vez de x, para el caso del Cuadro 3-3. Pero sí $yF(\succeq_1, \ldots, \succeq_I)x$ en el Cuadro 3-3 , entonces por IAI, entonces $yF(\succeq_!, \ldots, \succeq_I)x$ en 3-1, lo que es una contradicción. Por lo tanto, $xF(\succeq_!, \ldots, \succeq_I)y$ en primer y segundo lugar, respectivamente para el caso que aparece en el Cuadro 3-3.

Paso 3: Consideremos $z \in X$ tal que $z \neq x \neq y$. De tal manera que a partir de 3-3 tengamos que $z \succeq_i x$ y $z \succeq_i y$ para todo $i \neq n \in I$ y $x \succeq_n z \succeq_n y$. Esto lo podemos ver en el Cuadro 3-5. De esta manera, por IAI, la preferencia social más alta debe de ser x (i.e. $xF(\succeq_1, \ldots, \succeq_I)m, \forall m \in X$).

Paso 4: Ahora intercambiemos el lugar de las alternativas x y y para los agentes $i > n$, de tal manera que tengamos $z \succeq_i y \succeq_i x, \forall i \neq n$ como aparece en el Cuadro 3-6. Como teníamos que

$\succeq_1$	$\dots$	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	$\dots$	$\succeq_I$		Preferencias Sociales
·	...	·	x	·	...	·		x
·	...	·	z	·		·		·
·		·	y	·		·		·
·		·	·	·		·	$\rightarrow$	·
z		z	·	z		z		·
y		y	·	x		x		·
x	...	x	·	y	...	y		·

Cuadro 3-5:

$\succeq_1$	$\dots$	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	$\dots$	$\succeq_I$		Preferencias Sociales
·	...	·	x	·	...	·		x
·	...	·	z	·		·		·
·		·	y	·		·		z
·		·	·	·		·	$\rightarrow$	·
z		z	·	z		z		y
y		y	·	x		x		·
x	...	x	·	y	...	y		·

Cuadro 3-6:

$xF(\succeq_1,\dots,\succeq_I)m, \forall m \in X$ para el Cuadro 3-5, entonces por IAI, $xF(\succeq_1,\dots,\succeq_I)m, \forall m \in X$ para el cambio al Cuadro 3-6, ya que la preferencia relativa entre x y z no cambió, entonces la preferencia social no puede cambiar, pero como $z \succeq_i y, \forall i \in I$ entonces por Pareto, $zF(\succeq_1,\dots,\succeq_I)y, \forall i \in I$. Por lo tanto, $xF(\succeq_1,\dots,\succeq_I)zF(\succeq_1,\dots,\succeq_I)y$ con x la preferencia social más alta.

Paso 5: Consideremos un perfil arbitrario de preferencia individuales $(\succeq_1,\dots,\succeq_I)$ en donde $x \succeq_n y$. sí es necesario, alteremos el perfil haciendo que $x \succeq_n z \succeq_n y$ y $z \succeq_i m, \forall m \in X, \forall i \neq n$. Por IAI, esto no afecta la preferencia social de x y y, es decir, $xF(\succeq_1,\dots,\succeq_I)y$ y como $z \succeq_i x, \forall i \neq n$ y $x \succeq_n z$, entonces IAI implica que $xF(\succeq_1,\dots,\succeq_I)z$, y por Pareto $zF(\succeq_1,\dots,\succeq_I)y$. Por lo que por transitividad podemos concluir que $x(\succeq_1,\dots,\succeq_I)y$ siempre que $x \succ_n y$. sí repetimos este argumento cambiando los roles de y y z, además de recordar que z fue una alternativa arbitraria distinta de x y y, podemos concluir que $xF(\succeq_1,\dots,\succeq_I)m,$ para algún $m \in X$ siempre que $x \succ_n m$. Por lo tanto, podemos decir que el agente n es un *dictador* para x. Como x fue arbitraria hemos probado que para toda $x \in X$, existe un dictador para x. Pero claramente, no puede existir un dictador diferente para cada alternativa. Por lo tanto, hay un único dictador para todas las alternativas.

□

Se puede ver que este resultado se extiende a la agregación de otro tipo de ordenaciones, tan diferentes de preferencia como puede ser: (i) creencias sobre hipótesis, (ii) criterios múltiples que un tomador de decisión puede usar para generar una ordenación amplia de varias opciones de decisión, y (iii) clasificaciones incompatibles a ser conciliadas.

Parte III

El Problema del Diseño de Mecanismos

Capítulo 4

El Diseño de Mecanismos como parte de la Teoría de Juegos

Ahora pensemos un poco en el lenguaje de teoría de juegos, en donde un *juego* se refiere a cualquier situación social que comprenda a dos o más individuos. Los elementos esenciales de un juego son *jugadores*, *movimientos*, *pagos* e *información.* A estas se les conoce como *Reglas del Juego* y el objetivo del modelador es describir una situación en términos de estas reglas para explicar qué pasará en dicha situación. A continuación veremos una breve caracterización de los conceptos de teoría de juegos. Debido al enfoque de este trabajo, nos vamos a centrar en un tipo particular de juegos conocidos como juegos no cooperativos de información incompleta o bayesianos.

Definiciones básicas

En general los juegos se caracterizan a través de los tipos de información que los jugadores tienen a su disposición; se dividen en juegos de información perfecta o imperfecta, con incertidumbre o sin ella, de información simétrica o asimétrica y finalmente de información completa o incompleta. Un juego de información *perfecta* es en donde cada jugador sabe exactamente el punto en el que se encuentra en el juego, es decir, sabe cuales fueron las decisiones que se tomaron antes de las que él va a tomar y cuales son las consecuencias de todas las decisiones, sean propias o de los demás jugadores, de lo contrario el juego es de información *imperfecta.* En un juego de información *incompleta* la naturaleza hace una “decisión” antes de que cualquier jugador haga la suya y esta es información no conocida por al menos uno de los jugadores, esta es la idea intuitiva del concepto de tipo que veremos más adelante. De aquí en adelante siempre consideraremos juegos de información incompleta o bayesianos.

Un juego en donde se asume que los jugadores conocen toda la información relevante acerca de cada uno, incluyendo los pagos que cada uno recibe en las diferentes instancias del juego, se le conoce como *juegos de información completa*. Pero sí pensamos por un momento en las implicaciones de esto, nos deberíamos de poder convencer que esta es una suposición muy fuerte. ¿A caso dos firmas en una industria conocen los costos una de la otra? ¿Una empresa negociando con un sindicato necesariamente conoce la pérdida de utilidad que los miembros del sindicato van a experimentar sí hacen huelga durante un mes? La respuesta, claramente, es no.

Nos va a ser útil introducir la idea de *tipo* de un jugador o agente, el cual determina las preferencias que este tiene sobre las diferentes alternativas de un juego. Esto nos dará un poco de claridad cuando discutamos el diseño de mecanismos más adelante. Sea $\theta_i \in \Theta_i$ denota el tipo del agente i de un conjunto de posibles tipos Θ_i, $\theta_i \in \Theta_i$ es una variable aleatoria escogida por la naturaleza que solamente es observada por el jugador i. Una vez más, las alternativas las denotaremos por $x \in X$ y representaremos las preferencias de los agentes sobre las alternativas a través de la función de utilidad $u_i(x, \theta_i)$.

El concepto fundamental de la elección de los agentes en teoría de juegos está expresada como una *estrategia*. Sin demasiado estructura, una estrategia puede definirse como:

Definición 4.1. *Una estrategia $s_i(\theta_i) \in S_i$ es un plan completo y contingente, o regla de decisión, que define la acción o decisión que cada agente tomará en cada punto distinguible del juego. S_i es el conjunto de estrategias disponibles para el jugador i.*

La presencia de información incompleta abre la posibilidad de que tengamos que considerar las creencias de los jugadores acerca de las preferencias de los otros jugadores, sus creencias acerca de sus propias preferencias y lo que sucederá a partir de esas decisiones. Afortunadamente, existe una estrategia para enfrentar este problema originada por Harsanyi (1967-68) que hace esto innecesario. En ésta, uno imagina que las preferencias de los jugadores están determinadas por la realización de una variable aleatoria. Aunque la realización de la variable aleatoria solo sea observada por un jugador, se asume que la distribución de probabilidad es información pública de todos los jugadores. Mediante esta formulación , la situación de información incompleta la podemos reinterpretar como un juego de información imperfecta: La Naturaleza hace le primer movimiento, escogiendo las realizaciones de las variables aleatorias que determina el tipo de cada jugador y cada jugador observa únicamente la realización de su propia variable aleatoria. Este tipo de juegos se conocen como *juegos bayesianos*.

Debemos notar, que la alternativa x, a la que se hace referencia en la función de utilidad va a ser representada por una combinación de estrategias; para evitar ambigüedades en las siguientes secciones por lo que para cada jugador i se tiene una función de utilidad $u_i(s_i, s_{-i}, \theta_i)$, en donde $s_i = (s_1, \dots, s_I)$ es un perfil de estrategias que consta de una estrategia s_i para cada $i \in I$, y en este mismo sentido $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_I)$ es un perfil de estrategias para todo $j \neq i \in I$, es decir, para los oponentes del agente i. La distribución de probabilidad conjunta de los θ_i está dada por $F(\theta_1, \dots, \theta_I)$, la cual se asume como información pública para los jugadores. Con $\Theta = \Theta_1, \times \cdots \times \Theta_I$, un juego bayesiano se representa de la siguiente manera $\Gamma = [I, \{S_i\}, \{u_i(\cdot)\}, \Theta, F(\cdot)]$.

El pago esperado del jugador i dado un perfil de estrategias para los I jugadores $(s_1(\cdot), \dots, s_I(\cdot))$ está dado por:

$$u_i(s_1(\cdot), \dots, s_I(\cdot)) = E_\theta[u_i(s_1(\theta_1), \dots, s_I(\theta_I), \theta_i)]$$

Definición 4.2. *Una estrategia $s_i \in S_i$ se dice estrictamente dominada para el jugador i en el juego Γ sí existe otra estrategia $s'_i \in S_i$ tal que para toda $s_{-i} \in S_{-i}$,*

$$u_i(s'_i, s_{-i}) > u_i(s_i, s_{-i})$$

En este caso, decimos que s'_i domina estrictamente a s_i.

Con esta definición en mente podemos, de la misma manera, pensar en la *dominancia débil* cuando se da que $u_i(s'_i, s_{-i} \geq u_i(s_i, s_{-i})$.

Ahora podemos definir uno de los conceptos más usados y conocidos de la teoría de juegos, el de *equilibrio de Nash* (Nash1951).

Definición 4.3. *Un equilibrio bayesiano de Nash para el juego Γ es un perfil de estrategias $(s_1(\cdot), \dots, s_I(\cdot))$ que para cada $i = 1, \dots, I$*

$$u_i(s_i(\cdot), s_{-i}(\cdot)) \geq u_i(s'_i(\cdot), s_{-i}(\cdot))$$

para toda $s_i'(\cdot) \in S_I$.

Ahora tenemos las definiciones mínimas de teoría de juegos, podemos proseguir con nuestro análisis del diseño de mecanismos.

Capítulo 5

Mecanismos

Para comenzar, consideremos el escenario con l agentes, ordenados por $i = 1, \ldots, l$. Estos agentes deben hacer una elección colectiva de un conjunto X de posibles alternativas. Anterior a la elección, de manera que, cada agente i observa para sí mismo sus preferencia sobre las alternativas en X. Modelamos esto suponiendo que el agente i privadamente observa un parámetro, o señal θ, que determina sus preferencia. A θ la llamaremos el **tipo** del agente i. El conjunto de posibles tipos del agente i es denotado por Θ. Asumimos que cada agente i es un maximizador de su utilidad esperada, con función de utilidad Bernoulli cuando su tipo es θ es $u_i(x, \theta_i)$. La relación de equivalencia de preferencia de las alternativas en X que está asociada con la función de utilidad $u_i(x, \theta_i)$ está denotada por $\succeq_i (\theta_i)$. El conjunto de posibles relación de preferencia del agente i sobre X es dado por:

$$\mathcal{R}_i = \{\succeq_i : \succeq_i = \succeq_i (\theta_i) \text{ p.a. } \theta_i \in \Theta_i\}$$

Nótese que como θ_i únicamente es observada por el agente i, tenemos un escenario de **información incompleta**. Asumiremos que los tipos de los agentes se obtienen de una distribución común previamente conocida. En particular, denotaremos un perfil de tipos de agentes por $\theta = (\theta_1, \ldots, \theta_l)$, la función de densidad de probabilidad sobre la posible realización de $\theta \in \Theta_1 \times \cdots \times \Theta_l$ es $\phi(\cdot)$. Asumimos que la función de densidad $\phi(\cdot)$ así como los conjuntos $\Theta_1, \ldots, \Theta_l$ y las funciones de utilidad $u_i(\cdot, \theta_i)$ son información pública entre los agentes, pero el valor específico del tipo de cada agente i solo es observado por él.

Debido a que las preferencia de los agentes dependen de $\theta = (\theta_1, \ldots, \theta_l)$, los agentes pueden querer que la decisión colectiva dependa de θ. Para capturar esta dependencia formalmente daremos la definición siguiente:

Definición 5.1. *Una* **función de elección social** *es una función* $f : \Theta_1 \times \cdots \times \Theta_l \to X$ *la cual para cada posible perfil de tipos de los agentes* $(\theta_1, \ldots, \theta_l)$ *asigna una elección colectiva* $f(\theta_1, \ldots, \theta_l) \in X$*.*

Hay que notar que esta definición nos restringe a tratar con funciones de elección social deterministas. Esto es con fines prácticos, para evitar el volver este texto muy engorroso. En secciones posteriores permitiremos que la funciones de elección social asignen **loterías** sobre X.

Una característica deseable de las funciones de elección social es la propiedad de *eficiencia ex post* descrita a continuación:

Definición 5.2. *La función de elección social* $f : \Theta_1 \times \cdots \times \Theta_l \to X$ *es* **eficiente ex post** *(o de Pareto) sí para ningún perfil* $\theta = (\theta_1, \ldots, \theta_l)$ *existe un* $x \in X$ *tal que* $u_i(x, \theta_i) \geq u_i(f(\theta), \theta_i)$, $\forall i$ *y* $u_i(x, \theta_i) > u_i(f(\theta), \theta_i)$ *para algún* i*.*

La definición 5.2 dice que una función de elección social es de Pareto sí esta elige, para cada perfil $\theta = (\theta_1, \ldots, \theta_l)$, una alternativa $f(\theta) \in X$ que es óptima de Pareto dadas las funciones de utilidad de los agentes $u_1(\cdot, \theta_1), \ldots, u_l(\cdot, \theta_l)$.

El problema al que se enfrentan los agentes es que las θ_i no son públicamente observables y entonces para que se elija la elección social $f(\theta_1, \ldots, \theta_l)$ cuando los tipos de los agentes son $(\theta_1, \ldots, \theta_l)$, se debe confiar en que cada agente i revelará su tipo θ_i. Sin embargo, para una función de elección social $f(\cdot)$ dada, puede no ser la mejor opción para algún agente el revelar esta información honestamente.

Con lo anterior en mente, veamos los siguientes ejemplos para desarrollar un poco las ideas anteriores.

Ejemplo 5.1.

Veamos el caso más abstracto, tenemos un conjunto S y para cada agente i un conjunto ordenado $\mathcal{R}_i$ de posibles preferencia racionales en X. Ahora, supongamos que $X = \{x, y, z\}$ e $I = 2$. Supongamos también que el agente 1 tiene un tipo posible, tal que $\Theta_1 = \{\theta_1\}$, y que el agente 2 tiene dos posibles tipos, tal que $\Theta_2 = \{\theta_2', \theta_2''\}$. Los órdenes de preferencia posibles de los agentes $\mathcal{R}_1 = \{\succeq_1 (\theta_1)\}$ y $\mathcal{R}_2 = \{\succeq_2 (\theta_2'), \succeq_2 (\theta_2'')\}$ están caracterizados en el siguiente orden:

$\succeq_1 (\theta_1)$	$\succeq_2 (\theta_2')$	$\succeq_2 (\theta_2'')$
x	z	y
y	y	x
z	x	z

(Una alternativa ubicada más arriba es estrictamente preferida a una más baja, por ejemplo $x \succeq_1 (\theta_1) y \succeq_1 (\theta_1) z$).

Ahora supongamos que los agente desean implementar la función de elección social de Pareto $f(\cdot)$ con

$$f(\theta_1, \theta_2') = y$$

y

$$f(\theta_1, \theta_2'') = x.$$

De esta manera, se tiene que confiar en que el agente 2 revele verbalmente sus preferencia, pero se puede ver que no siempre le va a convenir hacerlo: Cuando $\theta_2 = \theta_2''$, el agente 2 va a querer mentir y reportar que su tipo es θ_2'.

En ambientes abstractos de elección social, surge un caso de interés central cuando $\mathcal{R}_i$ es, para cada agente i, igual a $\mathcal{R}$, el conjunto de todas las posibles relaciones de preferencia racionales en X. En este caso, un agente tiene muchas falsas declaraciones a su disposición e, intuitivamente, puede ser muy difícil para una función de elección social que induzca a los agentes a revelar verbalmente sus preferencia.

Ejemplo 5.2.

Consideremos un escenario en el cual se quiere asignar una unidad de un bien indivisible a uno de I agentes. Se pueden hacer transferencias monetarias. Un resultado puede ser representado por un vector $x = (y_1, \ldots, y_I, t_1, \ldots, t_I)$, en donde $y_i = 1$ sí el agente i obtiene el bien y $y_i = 0$ sí no lo obtiene y t_i es la transferencias monetaria recibida por el agente i. El conjunto de alternativas relevantes es:

$$X = \{(y_1, \ldots, y_I, t_1, \ldots, t_I) : y_i \in \{0,1\} \text{ y } t_i \in \mathbb{R} \forall i, \sum_i y_i = 1 \text{ y } \sum_i t_i \leq 0\}$$

Asumimos que la función de utilidad del tipo θ_i que es Bernoulli y tiene la forma cuasilineal

$$u_i(x, \theta_i) = \theta_i y_i + (m_i + t_i)$$

en donde m_i es, una vez más, la aportación de capital del agente i al capital total del intercambio. Aquí $\theta_i \in \mathbb{R}$ se puede ver como la valuación que da el agente al bien y tomamos el conjunto de posibles valuaciones del agente i como $\Theta_i = [\underline{\theta}_i, \bar{\theta}_i] \subset \mathbb{R}$, en donde $\underline{\theta}_i$ y $\bar{\theta}_i$ es el valor mínimo y máximo, respectivamente, de las posibles valuaciones del jugador i.

En esta situación, una función de elección social

$$f(\theta) = (y_1(\theta), \ldots, y_I(\theta), t_1(\theta), \ldots, t_I(\theta))$$

es eficiente ex post sí siempre asigna el bien al agente que tenga la mayor valuación (o a alguno de ellos sí son varios) y sí no hay desperdicio del principal; esto es, $\forall \theta = (\theta_1, \ldots, \theta_I) \in \Theta_1 \times \cdots \times \Theta_I$,

$$y_i(\theta)(\theta_i - \text{ Max}\{\theta_1, \ldots, \theta_I\}) = 0, \forall i$$

y

$$\sum_i t_i(\theta) = 0$$

Hay dos casos particulares que han tenido mucha atención en la literatura los cuales vale la pena mencionar. El primero es el caso de Comercio Bilateral. En este tenemos $I = 2$; interpretamos al agente 1 cómo el poseedor inicial del bien (el "vendedor") y al agente 2 como el comprador potencial. Cuando $\underline{\theta}_2 > \bar{\theta}_1$ es seguro que habrá ganancias del intercambio sin importar de como sean θ_1 y θ_2; cuando $\underline{\theta}_1 > \bar{\theta}_2$ es seguro que no habrá ganancias del intercambio y finalmente, sí $\underline{\theta}_2 < \bar{\theta}_1$ y $\underline{\theta}_1 < \bar{\theta}_2$ entonces puede que haya o no ganancias del intercambio, todo depende de como sea θ.

El segundo caso particular es el escenario de Subastas. En este, un agente, a quien llamaremos agente 0, es interpretado como el vendedor del bien (el "subastador") y se asume que no obtiene valor de él (en un caso más general, el vendedor puede tener un valor conocido de $\theta_0 = \bar{\theta}_0$ diferente de cero). Los otros agentes, $1, \ldots, I$ son compradores potenciales (los "subastantes").

Para ilustrar el problema de la revelación de la información en este ejemplo, consideremos una subasta con dos compradores ($I = 2$). Ahora suponemos que las valuaciones θ_i (privadas) de ambos compradores son obtenidas independientemente de la distribución uniforme en $[0, 1]$ y este hecho es conocido por todos los agentes. Consideremos la función de elección social $f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$ en donde

$$y_1(\theta) = 1 \text{ sí } \theta_1 \geq \theta_2; \ =0 \text{ sí } \theta_1 < \theta_2 \tag{5.1}$$

$$y_2(\theta) = 1 \text{ sí } \theta_1 < \theta_2; \ =0 \text{ sí } \theta_1 \geq \theta_2 \tag{5.2}$$

$$y_0(\theta) = 0 \quad \forall \theta \tag{5.3}$$

$$t_1(\theta) = -\theta_1 y_1(\theta) \tag{5.4}$$

$$t_2(\theta) = -\theta_2 y_2(\theta) \tag{5.5}$$

$$t_0(\theta) = -(t_1(\theta) + t_2(\theta)) \tag{5.6}$$

En esta función de elección social el vendedor entrega el bien al comprador con la mayor valuación (al comprador 1 sí es que hay un empate) y este comprador hace un pago al vendedor igual a su valuación (el otro comprador no hace ningún pago al vendedor). Notemos que $f(\cdot)$ no solamente es eficiente ex post sino que también es sumamente atractiva para el vendedor: sí $f(\cdot)$ puede implementarse, el vendedor capturará todo el capital que sea generado por el bien.

Supongamos que tratamos de implementar esta función de elección social. Asumimos que los compradores son maximizadores de su utilidad esperada. Ahora nos preguntamos; ¿sí el comprador 2 siempre anuncia su verdadero valor, el vendedor 1 verá que es óptimo hacer lo mismo? Para cada valor de θ_1, el problema del comprador 1 es escoger la valuación que anunciará, digamos que es $\hat{\theta}_1$, así que para resolver

$$\max_{\hat{\theta}_1} \left(\theta_1 - \hat{\theta}_1\right) \text{ Prob} \left(\theta_2 \leq \hat{\theta}_1\right)$$

o

$$\max_{\hat{\theta}_1} \left(\theta_1 - \hat{\theta}_1\right) \hat{\theta}_1$$

La solución a este problema es que el comprador 1 establezca $\hat{\theta}_1 = \theta_1/2$. Entonces vemos que el comprador 2 siempre dirá la verdad, la honestidad *no* es un óptimo para el comprador 1. Un punto similar aplica para el comprador 2. Intuitivamente, para esta función de elección social, un comprador tiene incentivo a dar una valuación menor para así disminuir la transferencia que tenga que hacer en el caso que tenga la valuación mayor y obtenga el bien. El costo de hacer esto es que obtiene el bien con menos regularidad, pero este es un costo, que hasta cierto punto, vale la pena tomar. Por lo que podemos ver, que una vez más hay un problema con implementar ciertas funciones de elección social en escenarios en los que la información es privada.

Aunque los compradores tengan un incentivo para mentir dada la función de elección social descrita en 5.1, esto no es cierto para todas las funciones de elección social en este tipo de subastas. Para observar este punto, supongamos que tratamos de implementar la función de elección social $\hat{f}(\cdot)$ que tiene la misma regla de asignación que en 5.1, pero tiene las funciones de transferencia

$$\begin{aligned} t_1(\theta) &= -\theta_2 y_i(\theta) \\ t_2(\theta) &= -\theta_1 y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)) \end{aligned}$$

En esta función de elección social, en vez de que el comprador i pague al vendedor una cantidad igual a su propia valuación θ_i sí gana el objeto, ahora pagará θ_j, en donde $j \neq i$, esto es, paga un monto igual a la **segunda valuación más alta**. Consideremos los incentivos del comprador 1 para decir la verdad ahora. sí el comprador 2 anuncia su valuación como $\hat{\theta}_2 \leq \theta_1$, el comprador 1 puede recibir la utilidad de $(\theta_1 - \theta_2) \geq 0$ reportando verázmente que su valuación es θ_1. Para cualquier otro anuncio, la utilidad resultante del comprador 1 es la misma (sí anuncia una valuación de al menos $\hat{\theta}_2$) o cero (sí anuncia una valuación menor a $\hat{\theta}_2$). Así que sí $\hat{\theta}_2 \leq \theta_1$, anunciar la verdad es débilmente menor para el comprador 1. En otro caso, sí el comprador 2 anuncia que su valuación es $\hat{\theta}_2 > \theta_1$, entonces la utilidad esperada del comprador 1 es 0 sí revela su verdadera valuación, sin embargo, el comprador 1 solo puede recibir una utilidad negativa mintiendo sobre que él obtiene el bien (decir que su valuación es al menos $\hat{\theta}_2$). Concluimos que decir la verdad es óptimo para el comprador 1 sin importar lo que el comprador 2 reporte. Formalmente, en el lenguaje de teoría de juegos, la honestidad es una estrategia débilmente dominante para el comprador 1. Una conclusión similar sigue para el comprador 2. Entonces, esta función de elección social es implementable aunque las valuaciones de los compradores sean información privada: es suficiente con pedir a cada comprador que reporte su tipo y luego elegir $\hat{f}(\cdot)$.

El ejemplo 5.2 sugiere que cuando los tipos de los agentes son privados, el problema de revelación de la información puede restringir el conjunto de funciones de elección social que pueden ser exitosamente implementadas. Con estos ejemplos como motivación podemos plantear la pregunta central que será el foco principal de este capítulo. *¿Qué funciones de elección social pueden implementarse cuando los tipos de los agentes son información privada?*.

Para responder esta pregunta necesitamos, en principio, pensar en todos los posibles modos en los cuales una función de elección social puede implementarse. En los ejemplos anteriores hemos imaginado implícitamente un escenario muy simple en el cual se le pide al agente i que revele directamente θ_i y luego, dados los anuncios $(\theta_1, \ldots, \theta_I)$, la alternativa $f(\theta_1, \ldots, \theta_I) \in X$ es elegida. Sin embargo, este no es el único modo de implementar una función de elección social. En particular, dada una función de elección social esta, puede ser implementada *indirectamente* haciendo que los agentes interactúen a través de algún tipo de institución en la cual haya reglas que gobiernen las acciones que los agentes puedan tomar y cómo estas acciones se traducen en un resultado social. Para ilustrar este punto, los Ejemplos 5.3 y 5.4 estudian dos instituciones de subasta comúnmente usadas.

Ejemplo 5.3.

Consideremos una vez más el escenario de subasta introducido en el Ejemplo 5.2. En una *subasta cerrada de primer precio* cada comprador potencial i tiene permitido presentar una oferta $b_i \geq 0$. A continuación las ofertas son abiertas y el comprador con la oferta más grande obtiene el bien y paga al vendedor una cantidad igual a su oferta.

Para especificar, consideremos una vez más el caso en el que hay dos posibles compradores $(I = 2)$ y cada θ_i se obtiene independientemente de la distribución uniforme en $[0, 1]$. Buscaremos un equilibrio en el cual las estrategias $b_i(\cdot)$ de cada comprador es de la forma $b_i(\theta_i) = \alpha_i \theta_i$ con $\alpha_i \in [0, 1]$. Supongamos que la estrategia del comprados 2 tiene esta forma y consideremos el problema del comprador 1. Para cada θ_i que quiera resolver

$$\max_{b_i \geq 0} (\theta_1 - b_1) \ \text{Prob} (b_2(\theta_2) \leq b_1)$$

Como la mayor oferta posible del comprador 2 es α_2 (él presenta la oferta α_2 cuando $\theta_2 = 1$), es evidente que el comprador 1 nunca debería de ofertar más de α_2. Más aún, como θ_2 es uniformemente distribuida en $[0, 1]$ y $b_2(\theta_2) \leq b_1$ sí y solo sí $\theta_2 \leq (b_1/\alpha_2)$, podemos escribir el problema del comprador 1 como

$$\max_{b_1 \in [0, \epsilon_2]} (\theta_1 - b_1) \left(\frac{b_1}{\alpha_2} \right)$$

La solución a este problema es

$$b_1(\theta_1) = \begin{cases} \frac{1}{2}\theta_1 & \text{sí } \frac{1}{2}\theta_1 \leq \alpha_2, \\ \alpha_2 & \text{sí } \frac{1}{2}\theta_1 > \alpha_2 \end{cases} \tag{5.7}$$

Y con un razonamiento similar

$$b_2(\theta_2) = \begin{cases} \frac{1}{2}\theta_2 & \text{sí } \frac{1}{2}\theta_2 \leq \alpha_1, \\ \alpha_1 & \text{sí } \frac{1}{2}\theta_2 > \alpha_1 \end{cases} \tag{5.8}$$

Haciendo $\alpha_1 = \alpha_2 = \frac{1}{2}$, podemos ver que las estrategias $b_i(\theta_i) = \frac{1}{2}\theta_i$ para $i = 1, 2$ constituye un equilibrio bayesiano de Nash para esta subasta. Por lo tanto, existe un equilibrio bayesiano de Nash de esta subasta cerrada de primer precio que indirectamente rinde los resultados especificados por la función de elección social $f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$ en donde

$$y_1(\theta) = 1 \text{ sí } \theta_1 \geq \theta_2; \ =0 \text{ sí } \theta_1 < \theta_2 \tag{5.9}$$

$$y_2(\theta) = 1 \text{ sí } \theta_1 < \theta_2; \ =0 \text{ sí } \theta_1 \geq \theta_2 \tag{5.10}$$

$$y_0(\theta) = 0 \ \ \forall\theta \tag{5.11}$$

$$t_1(\theta) = -\frac{1}{2}\theta_1 y_1(\theta) \tag{5.12}$$

$$t_2(\theta) = -\frac{1}{2}\theta_2 y_2(\theta) \tag{5.13}$$

$$t_0(\theta) = -(t_1(\theta) + t_2(\theta)) \tag{5.14}$$

Ejemplo 5.4.

Una vez más, consideremos el escenario de subasta introducido en el Ejemplo 5.2. En una **subasta de segundo precio a sobre cerrado**, cada comprador potencial i tiene permitido presentar una oferta cerrada $b_i \geq 0$. Las ofertas se abren y el comprador con la más alta obtiene el bien, pero ahora paga al vendedor una cantidad igual a la **segunda oferta** más alta.

Utilizando un razonamiento similar al final del Ejemplo 5.3, la estrategia $b_i(\theta_i) = \theta_i$ para todo $\theta_i \in [0, 1]$ es una estrategia débilmente dominante para cada comprador i. Entonces, cuando $I = 2$ la subasta cerrada de segundo precio implementa la función de elección social $f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta),$
$t_1(\theta), t_2(\theta))$ en donde

$$y_1(\theta) = 1 \text{ sí } \theta_1 \geq \theta_2; \ =0 \text{ sí } \theta_1 < \theta_2 \tag{5.15}$$

$$y_2(\theta) = 1 \text{ sí } \theta_1 < \theta_2; \ =0 \text{ sí } \theta_1 \geq \theta_2 \tag{5.16}$$

$$y_0(\theta) = 0 \ \ \forall\theta \tag{5.17}$$

$$t_1(\theta) = -\theta_2 y_1(\theta) \tag{5.18}$$

$$t_2(\theta) = -\theta_1 y_2(\theta) \tag{5.19}$$

$$t_0(\theta) = -(t_1(\theta) + t_2(\theta)) \tag{5.20}$$

Los Ejemplos 5.3 y 5.4 ilustran que, en términos generales, tenemos que considerar no so-

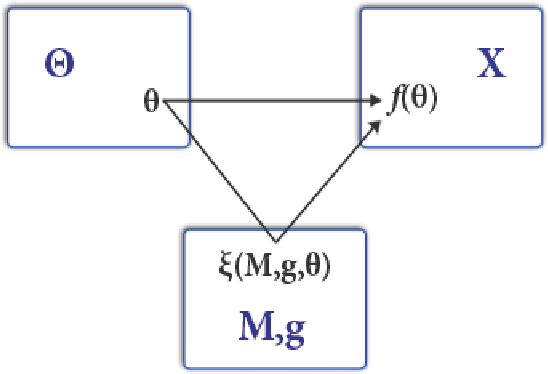

Figura 5.1: Este diagrama ilustra un mecanismo. El espacio Θ denota el espacio de tipos y X el de resultados. La función de elección social $f(\cdot)$ mapa un perfil de tipos θ a un resultado $f(\theta)$. Los agentes mandan mensajes M en un juego g. El equilibrio en el juego $\xi(M, g, \theta)$ puede ser diseñado para implementar alguna función de elección social $f(\theta)$.

lamente la posibilidad de implementar directamente funciones de elección social pidiendo a los agentes que revelen sus tipos, sino que también implementarlas indirectamente a través de instituciones en las cuales los agentes interactúen. La representación formal de dichas instituciones es conocida como *mecanismo*.

Definición 5.3. *Un* ***mecanismo*** $\Gamma = (S_1, \ldots, S_i, g(\cdot))$ *es una colección de* I *conjuntos de estrategias* $(S_1, \ldots, S_i)$ *y una función de resultados* $g : S_1 \times \cdots \times S_I \to X$.

Un mecanismo puede verse como una institución con reglas que gobiernan el procedimiento para hacer elecciones colectivas. Las acciones permitidas para cada agente i se resumen con el conjunto de estrategias S_i, y la regla para como las acciones de los agentes se transforman en una elección social está dada por la función de resultados $g(\cdot)$.

Formalmente, el mecanismo Γ combinado con los posibles tipos $(\Theta_1, \ldots, \Theta)$, la densidad de probabilidad $\phi(\cdot)$ y las funciones de utilidad Bernoulli $(u_1(\cdot),$
$\ldots, u_I(\cdot))$ definen un juego bayesiano de información incompleta. Esto es, haciendo $\hat{u}_i(s_1, \ldots, s_i, \theta_i) = u_i(g(s_1, \ldots, s_I), \theta_i)$ el juego

$$[I, \{S_i\}, \{\hat{u}_i(\cdot)\}, \Theta_1 \times \cdots \times \Theta_I, \phi(\cdot)]$$

es exactamente un tipo de juego bayesiano. Nótese que un mecanismo puede en principio ser un procedimiento dinámico complejo, en el cual los elementos de los conjuntos de estrategias S_i consistirían en planes contingentes de acción.

Para el escenario de subasta, la subasta cerrada de primer precio es el mecanismo en el que $S_i = \mathbb{R}_+$ para todo i y, dadas las ofertas $(b_1, \ldots, b_I) \in \mathbb{R}_+^I$, la función de resultados $g(b_1, \ldots, b_I) = (\{y_i(b_1, \ldots, b_I)\}_{i=1}^I, \{t_i(b_1, \ldots, b_I)\}_{i=1}^I)$ es tal que

$$
\begin{aligned}
y_i(b_1, \ldots, b_I) &= 1 \text{ sí y solo sí } i = \text{mín}\{j : b_j = \text{máx}\{b_1, \ldots, b_I\}\}, \\
t_i(b_1, \ldots, b_I) &= -b_i y_i(b_1, \ldots, b_I)
\end{aligned}
$$

En la subasta cerrada de segundo precio, por otro lado, tenemos los mismos conjuntos de estrategias y funciones $y_i(\cdots)$, pero tenemos $t_i(b_1, \ldots, b_I) = \text{máx}\{b_j : j \neq i\} y_i(b_1, \ldots, b_I)$.

Una estrategia para el agente i en el juego de información incompleta creado por un mecanismo Γ, es una función $s_i : \Theta_i \to S_i$ que da una de las elecciones en S_i del agente i por cada tipo en $\Theta_i$que pueda tener. *Grosso modo*, decimos que un mecanismo **implementa** la función de elección social $f(\cdot)$ para cada posible perfil de tipos $\theta = (\theta_1, \ldots, \theta_I)$. Esto se formaliza en la Definición 5.4.

Definición 5.4. *El mecanismo* $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ ***implementa*** *la función de elección social* $f(\cdot)$ *sí existe un equilibrio* $(s_1^*(\cdot), \ldots, s_I^*(\cdot))$ *del juego inducido por* Γ *tal que* $g(s_1^*(\theta_1), \cdots, s_I^*(\theta_I)) = f(\theta_1, \ldots, \theta_I)$ *para todo* $(\theta_1, \ldots, \theta_I) \in \Theta_1 \times \cdots \times \Theta_I$.

Sin embargo, notemos, que en la Definición 5.4 no especificamos que queremos decir con "equilibrio". Esto es porque no existe un solo concepto de equilibrio que sea universalmente aceptado como el concepto de solución para juegos. Como resultado, la literatura de diseño de mecanismos ha investigado la pregunta de la implementación para una variedad de conceptos de solución. Más adelante nos enfocaremos en dos conceptos de solución importantes: equilibrios en estrategias dominantes y equilibrios bayesianos de Nash.

Veamos también que la noción de implementación que hemos adoptado en la Definición 5.4 es en cierto sentido, débil; en particular, el mecanismo Γ puede tener *más de un equilibrio*, pero la Definición 5.4 solamente requiere que *uno de estos equilibrios* induzca resultados de acuerdo con $f(\cdot)$. Implícitamente, esta definición asume que, sí muchos equilibrios existen, los agentes jugarán el equilibrio que el diseñador del mecanismo quiera. A lo largo de este texto nos apegaremos a esta noción de implementación.

La identificación de todas las funciones de elección social que son implementables puede parecer una tarea de enormes proporciones porque, en principio, puede aparentar que necesitamos considerar todos los posibles mecanismos – un conjunto muy grande. Afortunadamente, un resultado importante conocido como el **principio de revelación**, nos dice que podemos restringir nuestra atención al tipo de mecanismos muy simples que estuvimos considerando implícitamente desde un principio, esto quiere decir, mecanismos en los cuales se le pide al agente que revele su tipo y dados los anuncios $(\hat{\theta}_1, \ldots, \hat{\theta}_I)$, la alternativa elegida es $f(\hat{\theta}_1, \ldots, \hat{\theta}_I) \in X$. Estos se conocen con el nombre de *mecanismos de revelación directa* y formalmente constituyen un caso especial de los mecanismos de la Definición 5.3.

Definición 5.5. *Un* ***mecanismo de revelación directa*** *es un mecanismo en el cual* $S_i = \Theta_i$ *para todo* i*,* $g(\theta) = f(\theta)$ *para toda* $\theta \in \Theta_1 \times \cdots \times \Theta_I$*.*

Mas aún, como veremos, el principio de revelación también nos dice que podemos restringir un poco mas nuestra atención a mecanismos de revelación directa en los que *decir la verdad es una estrategia óptima para cada agente*. Este hecho motiva la noción de *implementación veraz* que introducimos en la Definición 5.6 (una vez más seremos vagos en la definición de equilibrio que usaremos, más adelante especificaremos conceptos de solución).

Definición 5.6. *La función de elección social* $f(\cdot)$ *es* ***verazmente implementable*** *(o* ***compatible con incentivos****) sí el mecanismo de revelación directa* $\Gamma = (\Theta_1, \ldots, \Theta_I, f(\cdot))$ *tiene un equilibrio* $(s_1^*(\cdot), \ldots, s_I^*(\cdot))$ *en el cual* $s_i^*(\theta_i) = \theta_i$ *para toda* $\theta_i \in \Theta_i$ *y toda* $i = 1, \ldots, I$*; esto es, sí la honestidad de cada agente* i *constituye un equilibrio de* $\Gamma = (\Theta_1, \ldots, \Theta_I, f(\cdot))$*.*

Como indicio de porqué podemos limitar nuestro análisis a los mecanismos de revelación directa que induzcan la honestidad, brevemente verificamos que las funciones de elección social que son implementadas indirectamente a través de subastas cerradas de primer y segundo precio de los Ejemplos 5.3 y 5.4, también pueden ser implementadas verazmente usando un mecanismo de revelación directa. De hecho, para la subasta cerrada de segundo precio del Ejemplo 5.4 ya vimos este hecho, porque la función de elección social implementada por la subasta de segundo precio es exactamente la función que vimos en el Ejemplo 5.2, en donde decir la verdad es una estrategia débilmente dominante para ambos compradores. El Ejemplo 5.5 considera la subasta cerrada de primer precio.

Ejemplo 5.5. ***Implementación Veraz de la Función de Elección Social Implementada por la Subasta Cerrada de Primer Precio.***

Cuando enfrentamos un mecanismo de revelación directa $(\Theta_1, \ldots, \Theta_I, f(\cdot))$ con

$$f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$$

que satisface las condiciones de la ecuación 5.15 del Ejemplo 5.4, el anuncio óptimo $\hat{\theta}_1$ del comprador 1 cuando tiene tipo θ_1 resuelve

$$\max_{\hat{\theta}_1} \left(\theta_1 - \frac{1}{2}\hat{\theta}_1\right) \text{ Prob } \left(\theta_2 \leq \hat{\theta}_1\right)$$

o

$$\max_{\hat{\theta}_1} \left(\theta_1 - \frac{1}{2}\hat{\theta}_1\right) \hat{\theta}_1.$$

La condición de primer orden para este problema da $\hat{\theta}_1 = \theta_1$. Así que la honestidad es la estrategia óptima del comprador 1 dado que el comprador 2 siempre diga la verdad. Una conclusión muy similar sigue para el comprador 2. Por lo que la función de elección social implementada por la subasta cerrada de primer precio (en un equilibrio bayesiano de Nash) a través del mecanismo de revelación directa. Esto es, la función de elección social 5.15 es compatible con incentivos.

Debido al principio de revelación, como veremos más adelante, podremos restringir nuestro análisis para identificar aquellas funciones de elección social que pueden implementarse verázmente.

Finalmente, notamos que, en algunas aplicaciones, la participación en el mecanismo puede ser *voluntaria*, así que una función de elección social no solo tiene que inducir la revelación veraz de información pero también debe satisfacer ciertas *restricciones de participación* (o de *racionalidad individual*), sí va a ser implementada exitosamente. A continuación, nos enfocaremos exclusivamente en el problema de revelación de información. Más adelante introduciremos estas restricciones de participación.

En la Definición 5.4 se menciona que la posibilidad de implementar una función de elección depende de que exista un perfil de estrategias en equilibrio para el juego inducido, pero no se menciona qué pasa con equilibrios en particular. Ahora veremos que pasa en equilibrios de estrategias dominantes, de donde llegaremos al Teorema de Gibbard - Satterthwaite que está fuertemente relacionado con el Teorema de Imposibilidad de Arrow.

Recordemos de teoría de juegos, que una estrategia es débilmente dominante para un jugador en un juego, sí le da al menos un pago tan grande como cualquier otra de sus posibles estrategias para cualquier posible estrategia que su rival pueda jugar. Lo que nos lleva a la Definición 5.7.

Definición 5.7. *El perfil de estrategias* $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_I^*(\cdot))$ *es un equilibrio en estrategias dominantes del mecanismo* $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ *sí,* $\forall i$ *y toda* $\theta_i \in \Theta_i$*,*

$$u_i(g(s_i^*(\theta_i), s_{-i}), \theta_i) \geq u_i(g(s_i', s_{-i}), \theta_i)$$

para toda $s_i' \in S_i$ *y toda* $s_{-i} \in S_{-i}$*.*

Ahora adaptaremos la Definición 5.4 a la noción de equilibrio en estrategias dominantes.

Definición 5.8. *El mecanismo* $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ *implementa en estrategias dominates la función de elección social* $f(\cdot)$ *sí existe un equilibrio* $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_I^*(\cdot))$ *de* Γ *en estrategias dominantes, tal que* $g(s^*(\theta)) = f(\theta), \forall \theta \in \Theta$*.*

El concepto de implementación en estrategias dominantes es de interés especial ya que sí podemos encontrar un mecanismo Γ que implemente $f(\cdot)$ en estrategias dominantes, entonces este mecanismo implementa $f(\cdot)$ de una forma muy fuerte y robusta. Esto es cierto en varios sentidos. Primero, podemos estar suficientemente confiados en que un agente racional que tenga una estrategia (débilmente) dominante en verdad la jugará. A diferencia de las estrategias en equilibrios de Nash, un jugador no necesita predecir correctamente la jugada de su oponente para justificar su jugada de una estrategia dominante. Segundo, a pesar que asumimos que los agentes conocen la distribución de probabilidad $\phi(\cdot)$ sobre los tipos $(\theta_1, \ldots, \theta_I)$, por lo tanto puede deducir la distribución de probabilidad condicional correcta sobre θ_{-i}, sí Γ implementa $f(\cdot)$ en estrategias dominantes, esta implementación será robusta aún sí los agentes tienen creencias incorrectas, o tal vez contradictorias, de esta distribución. En particular, las creencias del agente i con respecto a la distribución de θ_{-i} no afecta la dominancia de su estrategia $s_i^*(\cdot)$. Tercero, sigue que sí Γ implementa $f(\cdot)$ en estrategias dominates entonces lo hace sin importar la distribución de probabilidad $\phi(\cdot)$. Por lo que, el mismo mecanismo puede usarse para implementar esa $f(\cdot)$ para cualquier $\phi(\cdot)$. Una ventaja de esto es que sí el diseñador del mecanismo es externo (digamos, el "gobierno"), no necesita conocer $\phi(\cdot)$ para implementar $f(\cdot)$ con éxito.

Como mencionamos en los puntos anteriores, para poder identificar sí es que una función de elección social $f(\cdot)$ es implementable, en principio, necesitamos considerar todos los posibles mecanismos. Afortunadamente, resulta que para implementar en equilibrios en estrategias dominantes es suficiente pedir que sea verázmente implementable, en el sentido de la Definición 5.6, a través del uso del *Principio de Revelación.*

Definición 5.9. *La función de elección social $f(\cdot)$ es verázmente implementable en estrategias dominantes (o compatible con incentivos en estrategias dominantes, o a prueba de estrategias, u honesto) sí $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i$ y $i = 1, \ldots, I$ es un equilibrio en estrategias dominantes del mecanismo de revelación directa $\Gamma = (\Theta_1, \ldots, \Theta_I, f(\cdot))$. Esto es, que para toda i y toda $\theta_i \in \Theta_i$,*

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)$$

para toda $\hat{\theta}_i \in \Theta_i$ y toda $\theta_{-i} \in \Theta_{-i}$.

La habilidad de restringirnos, sin pérdida de generalidad, a sí es que $f(\cdot)$ es verázmente implementable es una consecuencia de lo que conocemos como *principio de revelación en estrategias dominantes.*

Proposición 5.1 (El Principio de Revelación para Estrategias Dominantes)**.** *Supongamos que existe un mecanismo $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ que implementa la función de elección social $f(\cdot)$ en estrategias dominantes. Entonces, $f(\cdot)$ es verázmente implementable (o compatible con incentivos) en estrategias dominantes*

sí definimos los *conjuntos contorno inferiores* de la alternativa x cuando el agente i tiene tipo θ_i como:

$$L_i(x, \theta_i) = \{z \in X : u_i(x, \theta_i) \geq u_i(z, \theta_i)\}$$

Usando este conjunto contorno obtenemos la caracterización del conjunto de funciones de elección social, que son compatibles con incentivos en estrategias dominantes en la Proposición 5.2.

Proposición 5.2. *La función de elección social $f(\cdot)$ es verazmente implementable (o compatible con incentivos) en estrategias dominantes sí y solo sí para toda i, $\theta_{-i} \in \Theta_{-i}$ y toda pareja de tipos del agente i, $\theta_i', \theta_i'' \in \Theta_i$, tenemos*

$$f(\theta_i'', \theta_{-i}) \in L_i(f(\theta_i', \theta_{-i}), \theta_i') \text{ y } f(\theta_i', \theta_{-i}) \in L_i(f(\theta_i'', \theta_{-i}), \theta_i'') \tag{5.21}$$

La idea detrás de esta proposición se ilustra en el Cuadro 5-1, en la cual representamos la función de elección social $f(\cdot)$ para cada posible configuración de los tipos (θ_1, θ_2) en una situación en donde existen dos agentes ($I = 2$), dos posibles valores de θ_1 y tres de θ_2. Consideremos los incentivos del agente 1 para decir la verdad. sí la honestidad es una estrategia débilmente dominante para él, entonces cuando su tipo cambie de θ_1' a θ_1'', debe experimentar un débil cambio en sus preferencia entre los resultados $f(\theta_1', \theta_2)$ y $f(\theta_1'', \theta_2)$ para cada posible valor de θ_2. Un punto similar aplica para el agente 2.

	θ_2'	θ_2''	θ_2'''
θ_1'	$f(\theta_1', \theta_2')$	$f(\theta_1', \theta_2'')$	$f(\theta_1', \theta_2''')$
θ_1''	$f(\theta_1'', \theta_2')$	$f(\theta_1'', \theta_2'')$	$f(\theta_1'', \theta_2''')$

Cuadro 5-1:

Ahora usaremos estos últimos resultados para explorar un poco más a detalle las características de las funciones de elección social compatibles con incentivos impresentables en estrategias dominantes.

El Teorema de Gibbard-Satterthwaite

El Teorema de Gibbard-Satterthwaite fue descubierto independientemente en la década de 1970 por los autores que le dan nombre (Gibbard 1973, Satterthwaite 1975). El interés por este resultado es por ser un teorema de imposibilidad, similar en espíritu, al teorema de Arrow

(Proposición 3.1) y que le ha dado impulso a la investigación de incentivos e implementación. También, veremos que una sola demostración nos da ambos resultados, el de Arrow y el de Gibbard-Satterthwaite. Muestra que, para una clase muy general de problemas no hay esperanza de poder implementar satisfactoriamente funciones de elección social en estrategias dominantes.

Recordemos, que $\mathcal{R}$ y $\mathcal{P}$ son el conjunto de relaciones de preferencia posibles en X y, el conjunto de posibles relaciones de preferencia en X con la propiedad que no existen preferencia indiferentes. Denotamos por $f(\Theta) = \{x \in X : f(\theta) = x$ para algún $\theta \in \Theta$ a la imagen de $f(\cdot)$.

Definición 5.10. *La función de elección social $f(\cdot)$ es **dictatorial** sí existe un agente i tal que, para toda $\theta = (\theta_1, \ldots, \theta_I) \in \Theta$,*

$$f(\cdot) \in \{x \in X : u_i(x, \theta_i) \geq u_i(y, \theta_i), \forall y \in X\}$$

En palabras: Una función de elección social es dictatorial sí hay un agente para el cual $f(\cdot)$ siempre elige una de sus mejores alternativas.

Definición 5.11. *La función de elección social $f(\cdot)$ es **monótona** sí, para cualquier θ, sí θ' es tal que $L_i(f(\theta), \theta) \subset L_i(f(\theta), \theta'), \forall i$ [i.e. sí $L_i(f(\theta), \theta)$ está debilmente incluido en $L_i(f(\theta), \theta'), \forall i$], entonces $f(\theta') = f(\theta)$.*

La monotonía requiere lo siguiente. Supongamos que $f(\cdot) = x$, y que los tipos de los I agentes cambian a $\theta' = (\theta'_1, \ldots, \theta'_I)$ con la propiedad de que ninguna alternativa que era débilmente peor que x en θ, se vuelve estrictamente preferida a x cuando cambia el tipo a θ'. Entonces, x debe ser la elección social.

Definición 5.12. *La función de elección social $f(\cdot)$ es **eficiente de Pareto** (o simplemente* de Pareto*) sí siempre que $u_i(x, \theta) \geq u_i(y, \theta), \forall y \in X, \forall i \in I$ entonces, $f(\theta) = x$.*

Lo que quiere decir que siempre que la alternativa x está en lo más alto de las preferencia de todos los jugadores, entonces $f(\cdot)$ debe elegir x como elección social.

Con estas definiciones ahora enunciaremos y demostraremos el teorema de Gibbard-Satterthwaite.

Teorema 5.1 (El Teorema de Gibbard-Satterthwaite)**.** *Supongase que X es finito y contiene al menos tres elementos, $\mathcal{R}_i = \mathcal{P}, \forall i$ y que $f(\Theta) = X$. Entonces la función de elección social*

$f(\cdot)$ es verázmente implementable (o compatible con incentivos) en estrategias dominantes sí y solo sí es dictatorial.

No usaremos la demostración usual pues, como comentamos en la demostración del Teorema 3.1, el objetivo es mostrar que este teorema y el de Arrow tienen fundamentos lógicos idénticos, para lo cual, usaremos como referencia la demostración de Reny (Reny 2000). Partiremos la prueba en tres partes, con el objetivo de que sea más clara.

Demostración. Paso 1. sí $X \geq 3$ y la función de elección social $f : \mathcal{R}^I \rightarrow X$ es de Pareto y monótona, entonces f es dictatorial

Consideremos cualesquiera dos alternativas $x, y \in X$ y un perfil de preferencia θ, en donde $u_i(x,\theta) > u_i(z,\theta) > u_i(y,\theta), \forall z \in X, \forall i \in I$. Por eficiencia de Pareto, tenemos que $f(\theta) = x$.

Ahora cambiemos al tipo θ' moviendo el orden de las preferencia del agente 1, haciendo que $L_1(y,\theta') \geq L_1(y,\theta) > L_1(z,\theta), \forall z \in X$, subiendo la posición de y un lugar a la vez. Por la monotonía, tenemos que $f(\theta') = x$ siempre que $L_1(x,\theta') \geq L_1(y,\theta')$, pero una vez que $L_1(y,\theta') > L_1(x,\theta')$, la monotonía implica que $f(\theta') = y$ o x. sí $f(\theta') = x$, entonces haremos el mismo procedimiento para $L_i(y,\theta')$ con $i = 2, 3, \ldots, n$, hasta que dado que $L_n(y,\theta') > L_n(x,\theta')$ tengamos $f(\theta') = y$ (Debe existir ese agente n ya que la alternativa y al fin y al cabo va a estar en el tope de las preferencia de cada agente, entonces por Pareto la elección social debe de ser y). La Tabla 5-2 y la5-3 muestran las situaciones antes y después de que $L_n(y,\theta') > L_n(x,\theta')$.

$\succeq_1$	$\ldots$	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	$\ldots$	$\succeq_I$		Elección Social
y	$\ldots$	y	x	x	$\ldots$	x		
x	$\ldots$	x	y	·		·		
·		·	·	·		·	$\rightarrow$	x
·		·	·	·		·		
·		·	·	y	$\ldots$	y		

Cuadro 5-2:

Ahora fijémonos en la Tabla 5-4 y en la 5-5.

Ahora consideremos cambiar el orden de las preferencia del agente 1 subiendo el orden de y un lugar a la vez. Por la monotonía, la elección social se mantiene igual siempre que se cumpla que $x \succeq_1 y$. En el momento en el que se de el caso que $y \succeq_1 x$, la monotonía implica una de dos cosas; que $f(\theta) = y$ o que $f(\theta) = x$ entonces, se repite el procedimiento con el agente 2,

$\succeq_1$	$\dots$	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	$\dots$	$\succeq_I$		Elección Social
y	$\dots$	y	y	x	$\dots$	x		
x	$\dots$	x	x	$\cdot$		$\cdot$		
$\cdot$		$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\rightarrow$	y
$\cdot$		$\cdot$	$\cdot$	$\cdot$		$\cdot$		
$\cdot$		$\cdot$	$\cdot$	y	$\dots$	y		

Cuadro 5-3:

$\succeq_1$	$\dots$	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	$\dots$	$\succeq_I$		Elección Social
y	$\dots$	y	x	$\cdot$	$\dots$	$\cdot$		
$\cdot$	$\dots$	$\cdot$	y	$\cdot$		$\cdot$		
$\cdot$		$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\rightarrow$	x
$\cdot$		$\cdot$	$\cdot$	x		x		
x	$\dots$	x	$\cdot$	y	$\dots$	y		

Cuadro 5-4:

luego el 3, etc. hasta que para algún agente $n \leq I$, se de que $f(\theta) = y$, cuando se cambia y en las preferencia de este. (Debe existir un agente n tal, ya que llegará un punto en el que la alternativa $y \succeq_i z, \forall z \in X, i \in I$ y por la condición de Pareto, entonces $f(\theta) = y$.) Por esto, tenemos dos escenarios, el primero es $f(\theta'_1, \dots, \theta'_{n-1}, \theta_n, \dots, \theta_I) = x$, en donde θ'_i es el cambio de y en las preferencia del agente i al que denotaremos como $f_1(\theta)$ y el segundo en donde $f(\theta'_1, \dots, \theta'_n, \dots, \theta_I) = y$ al que denotaremos $f_2(\theta)$.

sí nos fijamos en $f_i(\theta)$ y movemos la alternativa x al final de las preferencia del agente $i < n$ y al penúltimo lugar para el agente $i > n$, manteniendo $x \succeq_n y$ (llamémosle $f'(\theta)$ a este cambio), a partir de esto quisiéramos argumentar que seguimos manteniendo $f_!(\theta) = x,$ para x, y. Ahora fijémonos en $f_2(\theta)$ y hagamos el mismo cambio, pero ahora mantengamos $y \succeq_n x$.

Primero, notemos que la elección social para $f'_2(\theta)$ debe ser $f'_2(\theta) = y$ por la propiedad de monotonía, ya que $f_2(\theta) = y$ antes de cambiar a $f'_2(\theta)$ y el orden relativo entre x y y no cambia para ningún agente. Ahora, notemos que los perfiles entre $f'_1(\theta)$ y $f'_2(\theta)$ solo van a diferir en el orden de x y y en las preferencia del agente n, como $f'_2(\theta) = y$ entonces por monotonía, $f'_1(\theta) = x$ o $f'_1(\theta) = y$. Pero, sí pasa que $f'_1(\theta) = y$, entonces por monotonía, $f_1(\theta) = y$, pero esto es una contradicción. Por lo tanto, $f'_1(\theta) = x$.

Ahora, consideremos $z \in X$ diferente de x y y, tal que $z \succeq_i y \succeq_i x, \forall i = 1, \dots, n-1$, $z \succeq_j x \succeq_y, \forall j = n+1, \dots, I$ y $x \succeq_n z \succeq_n y$, pero como el orden de x no cambia con respecto a ninguna otra alternativa entonces $f''_1(\theta) = x$.

$\succeq_1$	$\dots$	$\succeq_{n-1}$	$\succeq_n$	$\succeq_{n+1}$	$\dots$	$\succeq_I$		Elección Social
y	$\dots$	y	x	$\cdot$	$\dots$	$\cdot$		
$\cdot$	$\dots$	$\cdot$	y	$\cdot$		$\cdot$		
$\cdot$		$\cdot$	$\cdot$	$\cdot$		$\cdot$	$\rightarrow$	y
$\cdot$		$\cdot$	$\cdot$	x		x		
x	$\dots$	x	$\cdot$	y	$\dots$	y		

Cuadro 5-5:

A partir de lo anterior, intercambiemos el orden de x y y para los agentes $i > n$ de tal manera que tengamos que $z \succeq_i y \succeq_i x, \forall i \neq n$ y como $f_1''(\theta) = x$ y el único cambio que hicimos fue en el orden de x con respecto a y, entonces $f_1'''(\theta) = x$ o $f_1'''(\theta) = y$, pero no puede ser que $f_1'''(\theta) = y$ ya que $z \succeq_i y, \forall i \in I$ y la monotonía implicaría que $f_1'''(\theta) = y$ aunque z fuera movida al primer lugar de las preferencia de cada agente, lo que contradice la propiedad de Pareto. Por lo tanto, $f_1'''(\theta) = x$.

Nótese que un perfil de preferencia θ^* arbitrario para el agente n con x en primer lugar, puede obtenerse a partir del último cambio que hicimos sin perjudicar el orden de las preferencias de x, con respecto a cualquier otra alternativa para cualquier individuo $i \neq n$. Por lo tanto, la monotonía implica que la elección social $f(\theta^*) = x$ siempre que $x \succeq_n y, \forall x \neq y \in X$ para el agente n. Por lo que, podemos decir que este agente es un dictador para la alternativa x. Como x fue arbitraria, hemos probado que para cada alternativa $x \in X$, existe un dictador para x. Claramente no puede haber distintos dictadores para cada alternativa, por lo que existe un solo dictador para todas las alternativas.

Paso 2: sí una función de elección social $f : \mathcal{R}^I \rightarrow X$ es compatible con incentivos en estrategias dominantes y sobre, entonces $f(\cdot)$ es de Pareto y monótona.

Recordemos la Definición 5.9 para ver que $f(\cdot)$ sea honesta. Supongamos que la elección social es $f(\theta) = x$ y que para cada alternativa $y \in X$, la ordenación de preferencia θ' tiene que $x \succeq y$ siempre que pasa lo mismo para θ. Entonces, queremos mostrar que $f(\theta', \theta_{-i}) = x$. sí, por el contrario $f(\theta', \theta_{-i}) = y \neq x$, entonces la compatibilidad con incentivos implica que $x = f(\theta) \succeq f(\theta', \theta_{-i}) = y$ con respecto a θ. Como el lugar de x no cae con respecto al cambio a θ', entonces como $x = f(\theta) \succeq_i' y = f(\theta', \theta_{-i})$ de acuerdo a θ', pero esto es una contradicción a la compatibilidad con incentivos. Por lo tanto, $f(\theta', \theta_{-i}) = f(\theta) = x$.

Ahora, supongamos que $f(\theta) = x$ y para cada $i \in I$ y $y \in X$, θ' tiene que $x \succeq_i y$ siempre que pasa lo mismo para θ. Como sabemos que existe $\mu : \theta \to \theta'$ tal que $\mu(\theta_i) = \theta_i', \forall i \in I$, esto es claro simplemente moviendo la preferencia θ_i del agente i a θ_i' uno a la vez, y como ya mostramos que la elección social se tiene que mantener sin cambios para cada uno de estos cambios, entonces debemos tener que $f(\theta) = f(\theta')$. Por lo tanto, f es monótona.

Elijamos una $x \in X$. Como f es sobre, existe un $f(\theta) = x$ para algún $\theta \in \Theta$. Por monotonía, la elección social $f(\theta) = x$ siempre que $x \succeq_i y, \forall y \in X, \imath \in I$. Pero, otra vez, por la monotonía $f(\theta) = x$ sin importar la relación para todo $z \neq x \in X$, tal que $x \succeq_i x, \forall i \in I$. Consecuentemente, siempre que $x \succeq_i y, \forall y \in X, \forall i \in I$, entonces $f(\theta) = x$. Como $x \in X$ fue escogido arbitrariamente, entonces f es de Pareto.

De la Parte 1 y la Parte 2 podemos concluir la demostración del teorema. □

Se debe notar que la conclusión del Teorema 5.1 no se cumple sí X contiene dos elementos. Por ejemplo, una función de elección social de elección mayoritaria es a la vez dictatorial y compatible con incentivos en estrategias dominantes.

Notemos también que cuando $\mathcal{R}_\rangle = \mathcal{P}, \forall i$, cualquier función de elección social de Pareto *debe* cumplir que $f(\Theta) = X$. Por lo tanto, el teorema de Gibbard-Satterthwaite nos dice que cuando esto pasa y X contiene más de dos elementos, las únicas funciones de elección social que son compatibles con incentivos en estrategias dominantes son funciones de elección social dictatoriales. Dada esta conclusión negativa, sí queremos tener alguna esperanza de implementar funciones de elección social deseables, debemos aceptar una implementación por medio de conceptos de equilibrio menos robustos, de esta manera haciendo más débiles nuestros requisitos de implementación o enfocarnos en escenarios más restringidos. Lo que haremos en la siguiente sección será enfocarnos en lo primero, fijarnos en conceptos de solución menos robustos, en particular en equilibrios bayesianos de Nash.

Así, a partir de este resultado y las secciones siguientes tendremos el primer indicio de como Arrow y Myerson-Satterthwaite están más relacionados de lo que indican a primera vista.

Capítulo 6

Implementación Bayesiana

Ahora nos enfocaremos en estudiar la implementación de mecanismos en equilibrios bayesianos de Nash. Como ya vimos, un mecanismo $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ combinado con los posibles tipos $(\Theta_1, \ldots, \Theta_I)$, función de densidad $\phi(\cdot)$ y funciones de utilidad Bernoulli $(u_1(\cdot), \ldots, u_I(\cdot))$ define un juego bayesiano de información incompleta. También a menudo escribiremos $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_I)$, $s = (s_i, s_{-i})$ y $s(\cdot) = (s_i(\cdot), s_{-i}(\cdot))$ en donde $s_{-i}(\cdot) = (s_1(\cdot), \ldots, s_{i-1}(\cdot),$ $s_{i+1}(\cdot), \ldots, s_I(\cdot))$.

Comenzamos definiendo el concepto de un equilibrio bayesiano de Nash y modificando la Definición 5.4 para acoplarse a la noción de implementación en equilibrio bayesiano de Nash.

Definición 6.1. *El perfil de estrategias $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_I^*(\cdot))$ es un equilibrio bayesiano de Nash del mecanismo $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ sí, para toda i y toda $\theta_i \in \Theta_i$,*

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(\hat{s}_i, s_{-i}^*(\theta_i)), \theta_i)|\theta_i]$$

para toda $\hat{s}_i \in S_i$.

Definición 6.2. *El mecanismo $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ implementa la función de elección social $f(\cdot)$ en equilibrio bayesiano de Nash sí existe un equilibrio bayesiano de Nash de Γ, $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_I^*(\cdot))$ tal que $g(s^*(\theta)) = f(\theta)$ para toda $\theta \in \Theta$.*

A continuación veremos que una función de elección social es implementable en equilibrio bayesiano sí y solo sí es verázmente implementable en el sentido que denota la siguiente definición.

Definición 6.3. *La función de elección social $f(\cdot)$ es verazmente implementable en equilibrio bayesiano de Nash (o compatible con incentivos bayesianos) sí $s_i^*(\theta_i) = \theta_i$ para toda $\theta_i \in \Theta$ y $i = 1, \ldots, I$ es un equilibrio bayesiano de Nash del mecanismo de revelación directa $\Gamma = (\Theta_1, \ldots, \Theta_I, f(\cdot))$. Esto es, sí para toda $i = 1, \ldots, I$ y toda $\theta_i \in \Theta_i$,*

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i})\theta_i)|\theta_i] \tag{6.1}$$

para toda $\hat{\theta}_i \in \Theta_i$.

Podemos restringir nuestra atención, sin pérdida de generalidad, a que $f(\cdot)$ sea verazmente implementable gracias al principio de revelación para equilibrios bayesianos de Nash.

Proposición 6.1 (El Principio de Revelación en Equilibrio bayesiano de Nash)**.** *Supongamos que existe un mecanismo $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ que implementa la función de elección social $f(\cdot)$ en equilibrio bayesiano de Nash. Entonces $f(\cdot)$ es verazmente implementable en equilibrio bayesiano de Nash.*

Demostración. sí $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ implementa $f(\cdot)$ en equilibrio bayesiano de Nash, entonces existe un perfil de estrategias $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_I^*(\cdot))$ tal que $g(s^*(\theta)) = f(\theta)$ para toda θ y para toda $i\cdot$ y toda $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(\hat{s}_i, s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \tag{6.2}$$

para toda $\hat{s}_i \in S_i$.

La Condición 6.2 implica, en particular, que para todo i y $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \tag{6.3}$$

para toda $\hat{\theta}_i \in \Theta_i$. Como $g(s^*(\theta)) = f(\theta)$ para toda θ, la Condición 6.3 quiere decir que, para toda i y toda $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i})\theta_i)|\theta_i] \tag{6.4}$$

para toda $\hat{\theta}_i \in \Theta_i$. Pero esta es precisamente la condición 6.1, lo que nos dice que $f(\cdot)$ es compatible con incentivos en equilibrio bayesiano de Nash. $\square$

La idea básica detrás del principio de revelación en equilibrios bayesianos de Nash es que sí en un mecanismo $\Gamma = (S_1, \ldots, S_I, g(\cdot))$ cada agente revela que su tipo es θ_i, entonces elegir $s_i^*(\theta_i)$ es su mejor respuesta a las estrategias de los demás agentes, entonces sí introducimos un mediador que diga "Dime tu tipo, θ_i, y yo jugaré $s_i^*(\theta_i)$ por ti", cada agente verá que decir la verdad es una estrategia óptima dado que todos los otros agentes dicen la verdad, esto es, decir la verdad será un equilibrio bayesiano de Nash de este juego de revelación directa.

Como habíamos establecido, la implicación del principio de revelación es útil para identificar el conjunto de funciones de elección social implementables (ahora en equilibrios bayesianos de Nash), ya que ahora solamente necesitamos identificar aquellos que son verazmente implementables (o compatibles con incentivos).

Como bien sabemos de la teoría de juegos, el concepto de equilibrio en estrategias dominantes es más fuerte que aquel de equilibrio bayesiano de Nash. Como todo equilibrio en estrategias dominantes es necesariamente un equilibrio bayesiano de Nash, entonces cualquier función de elección social que sea implementable en estrategias dominantes es necesariamente implementable en equilibrio bayesiano de Nash. Visto intuitivamente, cuando comparamos los requerimientos para la implementación veraz de una función de elección social $f(\cdot)$ en estrategias dominantes[1] y en equilibrio bayesiano de Nash (6.1), podemos ver que, con la implementación Bayesiana, decir la verdad solo necesita dar al agente i su pago más alto *promediando sobre todos los posibles tipos θ_{-i} que puedan surgir para los demás agentes.* Comparando, el concepto de estrategia dominante requiere que decir la verdad sea la mejor estrategia posible para el agente i *para cada posible tipo θ_{-i}*. Por lo que esperamos razonablemente poder implementar un conjunto más grande de funciones de elección social en equilibrio bayesiano de Nash que en un equilibrio en estrategias dominantes. El inconveniente es que podemos estar menos seguros sobre esta implementación relativa a la de estrategias dominantes ya que depende de que los agentes (así como todo diseñador externo) conozca la densidad $\phi(\cdot)$ de los tipos de los agen-

[1] La función de elección social $f(\cdot)$ es verazmente implementable en estrategias dominantes (o compatible con incentivos en estrategias dominantes, o a prueba de estrategias) sí $s_i^*(\theta) = \theta_i$ para toda $\theta_i \in \Theta_i$ e $i = 1 \ldots, I$ es un equilibrio en estrategias dominantes del mecanismo de revelación directa $\Gamma = (\Theta_q, \ldots, \Theta_I, f(\cdot))$. Esto es, sí para toda i y toda $\theta_i \in \Theta_i$, $u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i), \forall \hat{\theta}_i \in \Theta_i, \theta_{-i} \in \Theta_{-i}$

tes, así como la plausibilidad de la suposición de Nash de que los agentes tienen esperanzas mutuamente correctas acerca de las elecciones de estrategias de cada quien.

6.1. El Mecanismo de Externalidad Esperada

Ahora veamos a una alternativa como un vector $x = (k, t_1, \ldots, t_I)$, en donde k es un elemento de un conjunto finito K y $t_i \in \mathbb{R}$ es una transferencia de un bien numerario ("dinero") al agente i. La función de utilidad tiene la forma cuasilineal

$$u_i(x, \theta_i) = v_i(k, \theta_i) + (\bar{m}_i + t_i), \tag{6.5}$$

en donde $\bar{m}_i$ es la aportación del agente i al capital total del intercambio.[2] Por simplicidad, de ahora en adelante normalizaremos $\bar{m}_i = 0$ para toda i. Asumimos que los I agentes no tienen fuentes externas de financiamiento y así $X = \{(k, t_1, \ldots, t_I) : k \in K, t_i \in \mathbb{R} \forall i \text{ y } \sum_i t_i \leq 0\}$. Una función de elección social en este ambiente toman la forma $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_I(\cdot))$. Notemos que sí $f(\cdot)$ es eficiente ex post entonces, para toda $\theta \in \Theta$,

$$\sum_i v_i(k(\theta), \theta_i) \geq \sum_i v_i(k, \theta_i) \ \forall k \in K \tag{6.6}$$

y

$$\sum_i t_i(\theta) = 0 \tag{6.7}$$

Ahora vamos a mostrar que es posible implementar en equilibrio bayesiano de Nash que cumpla las condiciones 6.6 y 6.7 siempre que los tipos de los agentes sean estadísticamente independientes uno de otros [i.e. cuando la densidad $\phi(\cdot)$ tiene la forma $\phi(\theta) = \phi_1(\theta_1) \times \cdots \times \phi_I(\theta_I)$].

Para verificar esto, hagamos que $k^*(\cdot)$ satisfaga la condición 6.6 y consideremos una función

[2] Este desarrollo depende solamente de las preferencia sobre ciertos resultados que tienen una forma cuasilineal pero también del hecho de que la función de utilidad es de la forma Bernoulli, es decir, cada agente i es neutral al riesgo con respecto a las loterías sobre su transferencia monetaria (sobre el pago que realiza al mecanismo)

de elección social $f(\cdot) = (k^*(\cdot), t_1(\cdot), \ldots, t_I(\cdot))$ en donde, para toda $i = 1, \ldots, I$,

$$t_i(\theta) = E_{\hat{\theta}_{-i}} \left[\sum_{j \neq i} v_j(k^*(\theta_i, \hat{\theta}_{-i}), \hat{\theta}_j) \right] + h_i(\theta_{-i}), \tag{6.8}$$

en donde, por ahora, tomamos a $h_i(\cdot)$ como una función arbitraria de θ_{-i}. Notemos que el término de la esperanza en la condición 6.8 representa el *beneficio esperado* de los agentes $j \neq i$ cuando el agente i anuncia que su tipo es θ_i y los agentes $j \neq i$ dicen la verdad. Como tal, es una función de únicamente los anuncios hechos del agente i, no es una función de los anuncios reales θ_{-i} de los agentes $j \neq i$. Por lo tanto, el cambio en la transferencia del agente i cuando cambia su tipo reportado es exactamente igual a la *externalidad esperada* de este cambio en los agentes $j \neq i$.

Procederemos a checar que cualquier función de elección social $f(\cdot)$ de la forma de la ecuación 6.8 es compatible con incentivos bayesianos. Para ver esto, notemos que cuando los agentes $j \neq i$ anuncian verázmente sus tipos, el agente i ve que decir la verdad es su estrategia óptima ya que (usando la independencia estadística de θ_i y θ_{-i})

$$\begin{aligned} E_{\theta_{-i}}[v_i(k^*(\theta), \theta_i) + t_i(\theta) | \theta_i] &= E_{\theta_{-i}} \left[\sum_{j=1}^{I} v_j(k^*(\theta), \theta_j) \right] + E_{\theta_{-i}}[h_i(\theta_{-i})] \\ &\geq E_{\theta_{-i}} \left[\sum_{j=1}^{I} v_j(k^*(\hat{\theta}_i, \theta_{-i}), \theta_j) \right] + E_{\theta_{-i}}[h_i(\theta_{-i})] \\ &= E_{\theta_{-i}}[v_i(k^*(\hat{\theta}_i, \theta_i), \theta_j) + t_i(\hat{\theta}_i, \theta_{-i}) | \theta_i] \end{aligned}$$

para toda $\hat{\theta}_i \in \Theta_i$, en donde se da la desigualdad ya que $k^*(\cdot)$ satisface la condición 6.6.

Lo que resta es mostrar que podemos elegir las funciones $h_i(\cdot)$ (para $i = 1, \ldots, I$) tal que satisfaga la condición de balance presupuestal dado en la condición 6.7. Para simplicidad notacional, definimos $\xi_i(\theta_i) = E_{\hat{\theta}_{-i}}[\sum_{j \neq i} v_j(k^*(\theta_i, \bar{\theta}_{-i}), \theta_j)]$. Ahora hacemos

$$h_i(\theta_{-i}) = -\left(\frac{1}{I-1} \right) \sum_{j \neq i} \xi_j(\theta_j), \tag{6.9}$$

para $i = 1, \ldots, I$. Con esta elección de $h_i(\cdot)$ tenemos

$$\begin{aligned}
\sum_i t_i(\theta) &= \sum_i \xi_i(\theta_i) + \sum_i h_i(\theta_{-i}) \\
&= \sum_i \xi(\theta_i) - \left(\frac{1}{I-1}\right) \sum_i \sum_{j \neq i} \xi_j(\theta_j) \\
&= \sum_i \xi_i(\theta_i) - \left(\frac{1}{I-1}\right) \sum_i (I-1) \xi_i(\theta_j) \\
&= 0.
\end{aligned}$$

Intuitivamente, podemos pensar la forma de las funciones $h_i(\cdot)$ en la condición 6.9 como sigue: Hemos visto que cuando los tipos de los agentes son $(\theta_1, \ldots, \theta_I)$, cada agente $i = 1, \ldots, I$ recibe un pago igual a $\xi_i(\theta_i)$ [el primer término en la condición 6.8]. Ahora, sí cada agente contribuye una parte $1/(I-1)$ igual de todos los pagos de los demás agentes, los pagos de un agente i dado a cada uno de los otros $I-1$ agentes en retribuciones que sumen $\xi_i(\theta_i)$. Por lo tanto la transferencia neta del agente i será $\xi_i(\theta_i) - (1/(I-1)) \sum_{j \neq i} \xi_j(\theta_j)$.

Este mecanismo de revelación directa es conocido como el *mecanismo de externalidad esperada*. En resumen, hemos mostrado que cuando las funciones de utilidad Bernoulli de los agentes tienen la forma de la condición 6.5 y los tipos de los agentes son estadísticamente independientes, existe una función de elección social eficiente ex post que es implementable en equilibrio bayesiano de Nash.

Aunque este es un resultado interesante, no hemos terminado del todo, aunque solo nos enfoquemos únicamente en funciones de utilidad Bernoulli que tengan la forma de la condición 6.5 y una distribución de tipos estadísticamente independiente. La razón por la que pasa esto es que, mientras el mecanismo de externalidad esperada implementa una función de elección social eficiente ex post, sus funciones de transferencia implican una distribución particular de utilidad sobre los varios tipos de los agentes y, nos gustaría considerar otros mecanismos, posiblemente algunos que involucren funciones de elección social que no sean eficientes ex post, que alteren esta distribución.

Una razón por la que esto es importante es que, en muchas aplicaciones de interés, los agentes son libres de salirse del mecanismo, así que cualquier mecanismo que queramos implementar no solamente tiene que ser compatible con incentivos en el sentido que hemos visto hasta ahora, sino que también debe satisfacer las *restricciones de racionalidad individual* (o de *participación*), las cuales aseguran que cada agente i realmente quiera participar en el mecanismo. sí el mecanismo de externalidad esperada no satisface estas restricciones, necesitamos considerar otros mecanismos que lo hagan. Por esta razón, es de nuestro interés identificar *todas* las funciones de elección social que sean implementables bayesianamente en este escenario.

Ahora, veremos una clase especial, pero a menudo estudiada, de casos en los cuales las preferencia de los agentes toman una forma lineal y sus tipos están distribuidos independientemente.

6.2. Compatibilidad con Incentivos Bayesianos con Utilidad Lineal

Supongamos ahora que cada agente i tiene una función de utilidad Bernoulli de la forma

$$u_i(x, \theta_i) = \theta_i v_i(k) + (\hat{m} + t_i).$$

Al igual que antes, normalizaremos $\hat{m} = 0$ para toda i. También suponemos que el tipo de cada agente i está en el intervalo $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i] \subset \mathcal{R}$ con $\underline{\theta}_i \neq \overline{\theta}_i$ y que los tipos de los agentes son estadísticamente independientes. Sea la función de distribución de θ_i, $\Phi_i(\cdot)$, y asumimos que tiene una densidad asociada $\phi_i(\cdot)$ que satisface $\phi_i(\theta_i) > 0$ para toda $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i]$.

Comenzamos obteniendo una condición necesaria y suficiente para la función de elección social $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_I(\cdot))$, para ser compatible con incentivos bayesianos. También definamos $\bar{t}_i(\hat{\theta}_i) = E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$ como la transferencia esperada del agente i dado que anuncia su tipo como $\hat{\theta}_i$ y todos los agentes $j \neq i$ dicen la verdad sobre sus tipos. Igualmente, denotamos como $\overline{v}_i(\hat{\theta}_i) = E_{\theta_{-i}}[v_i(k(\hat{\theta}_i, \theta_{-i}))]$ al "beneficio" esperado del agente i condicional a que anuncie $\hat{\theta}_i$. Debido a la forma de las funciones de utilidad de los agentes, podemos escribir la utilidad esperada del agente i cuando su tipo es θ_i y anuncia que su tipo es $\hat{\theta}_i$ (asumiendo que todos los agentes $j \neq i$ dicen la verdad) como

$$E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i] = \theta_i \overline{v}_i(\hat{\theta}_i) + \bar{t}_i(\hat{\theta}_i) \tag{6.10}$$

También es conveniente definir para cada i la función

$$U_i(\theta_i) = \theta_i \overline{v}_i(\theta_i) + \overline{t}(\theta_i),$$

que nos da la utilidad esperada del agente i en el mecanismo con la condición a que su tipo sea θ_i cuando él y todos los agentes reportan sus verdaderos tipos.

Proposición 6.2. *La función de elección social $f(\cdot) = (k(\cdot), t_i(\cdot), \ldots, t_I(\cdot))$ es compatible con incentivos bayesianos sí y solo sí para toda $i = 1, \ldots, I$,*

$$(i)\overline{v}_i(\cdot) \text{ es no decreciente} \tag{6.11}$$

$$(ii)U_i(\theta_i) = U_i(\underline{v}_i) + \int_{\underline{\theta}_i}^{\theta_i}, \forall\theta_i \tag{6.12}$$

Demostración. (i) *Necesidad.* La compatibilidad con incentivos bayesianos implica que para cada $\hat{\theta}_i > \theta_i$ tenemos

$$U_i(\theta_i) \geq \theta_i \overline{v}_i(\hat{\theta}_i) + \overline{t}_i(\hat{\theta}_i) = U_i(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i)\overline{v}_i(\hat{\theta}_i)$$

y

$$U_i(\hat{\theta}_i) \geq \hat{\theta}_i \overline{v}_i(\theta_i) + \overline{t}_i(\theta_i) = U_i(\theta_i) + (\hat{\theta}_i - \theta_i)\overline{v}_i(\theta_i).$$

Entonces,

$$\overline{v}_i(\hat{\theta}_i) \geq \frac{U_i(\hat{\theta}_i) - U_i(\theta_i)}{\hat{\theta}_i - \theta_i} \geq \overline{v}_i(\theta_i). \tag{6.13}$$

La expresión 6.13 inmediatamente implica que $\overline{v}_i(\cdot)$ debe ser no decreciente (recordemos que $\hat{\theta}_i > \theta_i$). Además, sí $\hat{\theta}_i \to \theta_i$ en 6.13 tenemos para toda θ_i

$$U_i'(\theta_i) = \overline{v}_i(\theta_i)$$

y así

$$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \overline{v}_i(s)ds, \forall\theta_i.$$

(ii) *Suficiencia.* Consideremos cualquier θ_i y $\hat{\theta}_i$ y supongamos sin pérdida de generalidad que $\theta_i > \hat{\theta}_i$. sí 6.11 y 6.12 se mantienen, entonces

$$\begin{aligned} U_i(\theta_i) - U_i(\hat{\theta}_i) &= \int_{\overline{\theta}_i}^{\theta_i} \overline{v}_i(s) ds \\ &\geq \int_{\overline{\theta}_i}^{\theta_i} \overline{v}_i(\hat{\theta}_i) ds \\ &= (\theta_i - \hat{\theta}_i)\overline{v}_i(\hat{\theta}_i). \end{aligned}$$

Por lo tanto,

$$U_i(\theta_i) \geq U_i(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i)\overline{v}_i(\hat{\theta}_i) = \theta_i \overline{v}_i(\hat{\theta}_i) + \overline{t}_i(\hat{\theta}_i).$$

Similarmente podemos obtener que

$$U_i(\hat{\theta}_i) \geq U_i(\theta_i) + (\hat{\theta}_i - \theta_i)\overline{v}_i(\theta_i) = \hat{\theta}_i \overline{v}_i(\theta_i) + \overline{t}_i(\theta_i).$$

Así que $f(\cdot)$ es compatible con incentivos bayesianos. □

La Proposición 6.2 muestra que para identificar todas la funciones de elección social compatibles con incentivos bayesianos en el modelo lineal podemos proceder de la siguiente manera: Primero, identificamos que funciones $k(\cdot)$ llevan a la función del beneficio esperado $\overline{v}_i(\cdot)$ de cada agente i son no decrecientes. Entonces, para cada una de estas funciones, identificamos las funciones de transferencia esperada $\overline{t}_1(\cdot), \ldots, \overline{t}_I(\cdot)$ que satisfagan la condición 6.12 de la proposición. sí sustituimos $U_i(\cdot)$, esto es precisamente la función de transferencia esperada que satisface, para cada $i = 1, \ldots, I$,

$$\overline{t}_i(\theta_i) = \overline{t}_i(\underline{\theta}_i) + \underline{\theta}_i v_i(\underline{\theta}_i) - \theta_i v_i(\theta_i) + \int_{\underline{\theta}_i}^{\theta_i} \overline{v}_i(s) ds$$

para alguna constante $\overline{t}_i(\theta_i)$. Finalmente, sea cualquier conjunto de funciones de transferencia $(t_i(\theta), \ldots, t_I(\theta))$ tal que $E_{\theta_i}[t_i(\theta_i, \theta_{-i})] = \overline{t}_i(\theta_i)$ para toda θ_i. En general, hay muchas funciones de este tipo $t_i(\cdot, \cdot)$; una por ejemplo, es $t_i(\theta_i, \theta_{-i}) = \overline{t}_i(\theta_i)$.

6.3. Subastas: el Teorema de Equivalencia de Ingresos

Ahora como una ilustración de la implicación de la caracterización de la sección anterior, consideremos el escenario de una subasta visto en el Ejemplo 5.2: el Agente 0 es el vendedor de un bien indivisible del cual no deriva ningún valor y los agentes $1, \ldots, I$ son compradores potenciales. Ahora veremos que es conveniente generalizar el conjunto de posibles alternativas relativas a las consideradas en el Ejemplo 5.2, permitiendo una asignación *aleatoria* del objeto. Por lo que, ahora tomamos $y_i(\theta)$ como la probabilidad de que el agente i obtenga el objeto cuando el vector de tipos anunciados es $\theta = (\theta_1, \ldots, \theta_I)$. La utilidad esperada del comprador i cuando el perfil de tipos para los I compradores es $\theta = (\theta_1, \ldots, \theta_I)$, es entonces $\theta_i y_i(\theta) + t_i(\theta)$. Notemos que el comprador i es neutral al riesgo con respecto a las loterías sobre las transferencias y sobre la asignación del bien.

El escenario corresponde al panorama estudiado en la Proposición 6.2 en el caso en donde tomamos $k = (y_1, \ldots, y_I), K = \{(y_1, \ldots, y_I) : y_i \in [0,1] \forall i = 1, \ldots, I$ y $\sum_i y_i \leq 1\}$ y $v_i(k) = y_i$. Entonces, para aplicar la Proposición 6.2, podemos escribir $\overline{v}_i(\hat{\theta}_i) = \overline{y}_i(\hat{\theta}_i)$ en donde $\overline{y}_i(\hat{\theta}_i) = E_{\theta_{-i}}[y_i(\hat{\theta}_i, \theta_{-i})]$ es la probabilidad de que i obtenga el objeto condicional a que anuncie que su tipo sea $\hat{\theta}_i$, cuando los agentes $j \neq i$ anuncian sus tipos verazmente y $U_i(\theta_i) = \theta_i \overline{y}_i(\theta_i) + \overline{t}_i(\theta_i)$.

Ahora podemos establecer un resultado notable, conocido como el *teorema de equivalencia de ingresos.*

Proposición 6.3 (Teorema de Equivalencia de Ingresos)**.** *Consideremos un escenario de subasta con I compradores neutrales al riesgo, en donde la valuación del comprador i se obtiene de un intervalo $[\underline{\theta}_i, \overline{\theta}_i]$ con $\underline{\theta}_i \neq \overline{\theta}_i$ y una densidad estrictamente positiva $\phi_i(\cdot) > 0$ y en donde los tipos de los compradores son estadísticamente independientes. Supongamos que un par dado de equilibrios bayesianos de Nash de dos procedimientos diferentes de subasta son tales que para cada comprador i: (i) Para cada posible realización de $(\theta_1, \ldots, \theta_I)$, el comprador i tiene una probabilidad idéntica de obtener el bien en las dos subastas; y (ii) el comprador i tiene el mismo nivel de utilidad en las dos subastas cuando su valuación por el objeto está en su más bajo nivel posible. Entonces, estos equilibrios de las dos subastas generan el mismo ingreso esperado para el vendedor.*

Demostración. Por el principio de revelación, sabemos que la función de elección social que es implementada por el equilibrio de cualquier procedimiento de subasta debe ser compatible con incentivos bayesianos. Entonces, podemos establecer el resultado mostrando que, sí dos funciones de elección social compatibles con incentivos bayesianos en este escenario de subasta tiene las mismas funciones $(y_1(\theta), \ldots, y_I(\theta))$ y los mismos valores de $(U_1(\underline{\theta}), \ldots, U_I(\underline{\theta}))$, entonces generan el mismo ingreso esperado para el vendedor.

Para mostrar esto, derivamos una expresión para el ingreso esperado del vendedor de un mecanismo arbitrario compatible con incentivos bayesianos. Notemos, primero, que el ingreso esperado del vendedor es igual a $\sum_{i=0}^{I} E[-t_i(\theta)]$. Ahora,

$$\begin{aligned}
E[-t_i(\theta)] &= E_{\theta_i}[-\hat{t}_i(\theta)] \\
&= \int_{\underline{\theta}_i}^{\overline{\theta}_i} [\overline{y}_i(\theta_i)\theta_i - U_i(\theta_i)]\, \phi_i(\theta_i) d\theta_i \\
&= \left[\int_{\underline{\theta}_i}^{\overline{\theta}_i} (\overline{y}_i(\theta_i)\theta_i - U_i(\theta_i)) - \int_{\underline{\theta}_i}^{\theta_i} \overline{y}_i(s) ds\right] \phi_i(\theta_i) d\theta_i \\
&= \left[\int_{\underline{\theta}_i}^{\overline{\theta}_i} \left(\overline{y}_i(\theta_i)\theta_i - \int_{\underline{\theta}_i}^{\theta_i} \overline{y}_i(s) ds\right) \phi_i(\theta_i) d\theta_i\right] - U_i(\underline{\theta}_i)
\end{aligned}$$

Y sí integramos por partes tenemos que

$$\begin{aligned}
\int_{\underline{\theta}_i}^{\overline{\theta}_i} \left(\int_{\underline{\theta}_i}^{\theta_i} \overline{y}_i(s) ds\right) \phi_i(\theta_i) d\theta_i &= \left(\int_{\underline{\theta}_i}^{\overline{\theta}_i} \overline{y}_i(\theta_i) d\theta_i\right) - \left(\int_{\underline{\theta}_i}^{\overline{\theta}_i} \overline{y}_i(\theta_i)\Phi(\theta_i) d\theta_i\right) \\
&= \int_{\underline{\theta}_i}^{\overline{\theta}_i} \overline{y}_i(\theta_i)(1 - \Phi_i(\theta_i)) d\theta_i.
\end{aligned}$$

Sustituyendo, vemos que

$$E[-\bar{t}(\theta_i)] = \left[\int_{\underline{\theta}_i}^{\overline{\theta}_i} \overline{y}_i(\theta_i)\Bigg(\theta_i - \frac{1 - \Phi_i(\theta_i)}{\phi_i(\theta_i)})\phi_i(\theta_i) d\theta_i] - U_i(\underline{\theta}_i).$$

o equivalentemente

$$E[-\bar{t}_i(\theta_i)] = \tag{6.14}$$

$$\left[\int_{\underline{\theta}_1}^{\overline{\theta}_1} \cdots \int_{\underline{\theta}_I}^{\overline{\theta}_I} y_i(\theta_1, \ldots, \theta_I) \left(\theta_i - \frac{1 - \Phi_i(\theta_i)}{\phi_i(\theta_i)}\right) \left(\Pi_{j=1}^{I} \Phi_j(\theta_j)\right) d\theta_1 \cdots d\theta_I\right] - U_i(\underline{\theta}_i). \tag{6.15}$$

Por lo que el ingreso esperado del vendedor es igual a

$$\left[\int_{\underline{\theta}_1}^{\bar{\theta}_1} \cdots \int_{\underline{\theta}_I}^{\bar{\theta}_I} y_i(\theta_1, \ldots, \theta_I) \left(\theta_i - \frac{1-\Phi_i(\theta_i)}{\phi_i(\theta_i)}\right) \left(\Pi_{j=1}^{I} \Phi_j(\theta_j)\right) d\theta_1 \cdots d\theta_I\right] - U_i(\underline{\theta}_i). \tag{6.16}$$

Inspeccionando 6.16 vemos que cualesquiera dos funciones de elección social compatible con incentivos bayesianos que generan las mismas funciones $(y_1(\theta), \ldots, y_I(\theta))$ y los mismos valores de $(U_1(\underline{\theta}_1), \ldots, U_I(\underline{\theta}_I))$ generan el mismo ingreso esperado para el vendedor. $\square$

Capítulo 7

El Teorema de Myerson - Satterthwaite

En los capítulos 4 y 6 notamos que existen restricciones dentro del conjunto de funciones de elección social implementables cuando hay información privada. Nuestro análisis hasta ese punto había asumido implicitamente que cada agente i no tiene otra opción más que participar en cualquier mecanismo que elija el diseñador del mecanismo. Esto es, la decisión del agente i se limitaba a escoger sus acciones óptimas dentro de las permitidas por el mecanismo.

Sin embargo, en muchas aplicaciones, la participación de los agentes en el mecanismo es *voluntaria*. Como resultado, la función de elección social que va a ser implementada por el mecanismo debe ser, no solo compatible con incentivos sino también debe satisfacer ciertas *restricciones de participación (o racionalidad individual)*, sí es que va a ser exitosamente implementada. Un importante resultado en este tema es el *Teorema de Myerson - Satterthwaite*, en donde se ilustran este tipo de limitaciones que pueden causar las restricciones de participación en el conjunto de funciones de elección social.

En el caso general, podemos distinguir entre tres etapas en las cuales las restricciones de participación pueden ser relevantes en cualquier aplicación particular. Primero, un agente i puede retirarse del mecanismo en la *etapa ex post* que surge después de que los agentes anuncian sus tipos y un resultado en X se ha escogido. Formalmente, supongamos que el agente i puede recibir una utilidad de $\overline{u}_i(\theta_i)$ con retirarse del mecanismo, siempre que su tipo sea θ_i. Entonces, para asegurar la participación del agente i debemos satisfacer la *restricción de participación ex post (o racionalidad individual)*,

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq \overline{u}_i(\theta_i), \forall(\theta_i, \theta_{-i})$$

En otras circunstancias, el agente i puede ser capaz de retirarse del mecanismo en la *etapa interim* que surge cuando los agentes han aprendido sus tipos pero antes de que hayan elegido sus acciones dentro del mecanismo. sí dejamos que $U_i(\theta_i|f) = E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i]$ denote la *utilidad esperada interim* de la función de elección social $f(\cdot)$ cuando su tipo es θ_i, el agente i participará en un mecanismo que implemente la función de elección social $f(\cdot)$, cuando tiene tipo θ_i sí y solo sí $U_i(\theta_i|f)$ no es menor que $\overline{u}(\theta_i)$. Por lo tanto, la *restricción de participación interim (o racionalidad individual)* para el agente i requiere que,

$$U_i(\theta_| f) = E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq \overline{u}_i(\theta_i), \forall\theta_i$$

En otros casos, el agente i solo puede rehusarse a participar en la *etapa ex ante* que surge antes de que los agentes conozcan sus tipos. Dejando que $U_i(f) = E_{\theta_{-i}}[U_i(\theta_i|f)] = E[u_i(f(\theta_i, \theta_{-i}), \theta_i)]$ denote la *utilidad esperada ex ante* del agente i de un mecanismo que implementa la función de elección social $f(\cdot)$ la *restricción de participación ex ante (o racionalidad individual)* para el agente i es

$$U_i(f) \geq E_{\theta_i}[\overline{u}_i(\theta_i)].$$

Las restricciones de participación son de la variedad *ex ante* cuando, los agentes pueden concordar en estar vinculados por el mismo mecanismo antes de conocer sus tipos. Cuando los agentes conocen sus tipos antes de que puedan acordar a la obligación del mecanismo, nos enfrentamos a restricciones de participación *interim*. Finalmente, sí no hay forma de amarrar contra su voluntad a los agentes a los resultados asignados, entonces nos enfrentamos a restricciones de participación ex post.

Notemos que sí $f(\cdot)$ satisface la restricción *ex post*, entonces satisface la interim y a su vez, sí satisface la interim entonces satisface la ex ante. Sin embargo, el inverso no es cierto. Por lo que las restricciones impuestas por la participación voluntaria son más severas cuando los agentes pueden retirarse en la etapa *ex post* y menos severas cuando se pueden retirar solo en la etapa *ex ante*.

Ahora, consideremos una vez más el escenario de comercio bilateral que introdujimos en el Ejemplo 5.3. El Agente 1 es el vendedor de un bien indivisible y tiene una valuación por el objeto que está en el intervalo $\Theta_1 = [\underline{\theta}_1, \overline{\theta}_1] \subset \mathbb{R}$; el agente 2 es el comprador y tiene una valuación que está en el intervalo $\Theta_2 = [\underline{\theta}_2, \overline{\theta}_2] \subset \mathbb{R}$. Las dos valuaciones son estadísticamente independientes y θ_i tiene una función de distribución $\Phi_i(\cdot)$ con una densidad asociada $\phi(\cdot)$ que satisface $\phi_i(\theta_i) > 0$ para toda $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i]$. Hacemos que $y_i(\theta)$ denote la probabilidad de que el agente i reciba el bien dados los tipos $\theta = (\theta_1, \theta_2)$ y así, la utilidad esperada del agente i dada θ es $\theta_i y_i(\theta) + t_i(\theta)$ (normalizamos $m_i = 0$ para toda i).

El mecanismo de externalidad esperada que mencionamos en la sección anterior, nos muestra que en este escenario podemos implementar bayesianamente una función de elección social eficiente ex post. Sin embargo, un problema surge con el mecanismo de externalidad esperada cuando el intercambio es voluntario. En este caso, cada tipo del comprador y del vendedor debe tener una ganancia esperada no negativa del intercambio sí es que va a participar. En particular, sí un vendedor de tipo θ_1 va a participar en un mecanismo que implementa una función de elección social $f(\cdot)$, esto es, sí la participación en el mecanismo es *individualmente racional* para este tipo del vendedor, entonces debe ser que $U_1(\theta_1|f) \geq \theta_1$, ya que este vendedor puede alcanzar una utilidad esperada de θ_1 sí no participa en el mecanismo simplemente consumiendo el bien. Igualmente, un comprador de tipo θ_2 siempre puede ganar cero simplemente rehusándose a participar y entonces, debemos tener que $U_1(\theta_2) \geq 0$. Desafortunadamente estas restricciones de participación no se satisfacen en el mecanismo de externalidad esperada.

Para probar esto consideremos una vez más un escenario de comercio bilateral en donde cada θ_i se obtiene independientemente de una distribución uniforme sobre $[0, 1]$. Supongamos que al rehusarse a participar en el mecanismo un vendedor con valuación θ_1 recibe una utilidad esperada de θ_1 (simplemente consume el bien), mientras que un comprador con valuación θ_2 recibe una utilidad esperada de 0 (simplemente consume su aportación al numerario, la cual normalizamos a 0).

Ahora, sabemos que el vendedor dirá la verdad y su utilidad esperada será dada por la función,

$$E_{\theta_2} u_1 = \theta_1^2 \cdot \frac{\theta_1^2}{2} + \frac{1}{2} \int_0^1 \theta_2^2 d\theta_2 = \frac{\theta_1^2}{2} + \frac{1}{6}$$

Por lo tanto, el vendedor ganará al no participar sí $\theta_1 > \frac{\theta_1^2}{2} + \frac{1}{6}$, esto es sí $\theta_1 > 1 - \sqrt{2/3}$.

Notemos que como la utilidad de reserva del comprador es 0, siempre le convendrá participar bajo la suposición de que el vendedor siempre participa.

El *Teorema de Myerson- Satterthwaite* nos indica el siguiente decepcionante resultado: Siempre que haya posibles ganancias del intercambio, pero que no sean seguras, no existe una función de elección social eficiente *ex post* que sea compatible con incentivos bayesianos y satisfaga las restricciones de participación. Entonces, bajo las condiciones del teorema, la presencia de información privada y participación voluntaria implica que es imposible alcanzar la eficiencia *ex post*.

Teorema 7.1 (Teorema de Myerson - Satterthwaite)**.** *Consideremos un escenario de comercio bilateral en donde el vendedor y el comprador son neutrales al riesgo, las valuaciones θ_1 y θ_2 son independientes sobre los intervalos $[\underline{\theta_1}, \overline{\theta_1}] \subset \mathcal{R}$ y $[\underline{\theta_2}, \overline{\theta_2}] \subset \mathcal{R}$ con densidades estrictamente positivas y $(\underline{\theta_1}, \overline{\theta_1}) \cap (\underline{\theta_2}, \overline{\theta_2}) \neq \emptyset$. Entonces, no existe una función de elección social compatible con incentivos bayesianos que sea eficiente ex post y que se obtengan ganancias esperadas no negativas de la participación para cada tipo del comprador y del vendedor.*

Demostración. El argumento consiste de dos pasos:

Paso 1: En cualquier función de elección social compatible con incentivos bayesianos y racional individualmente $f(\cdot) = [y_!(\cdot), y_2(\cdot), t_1(\cdot), t_2(\cdot)]$ en donde $y_1(\theta_1, \theta_2) + y_2(\theta_1, \theta_2) = 1$ y $t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) = 0$, debemos tener que

$$\int_{\underline{\theta}_1}^{\overline{\theta}_1} \int_{\underline{\theta}_2}^{\overline{\theta}_2} y_2(\theta_1, \theta_2) \left[\left(\theta_2 - \frac{1 - \Phi_2(\theta_2)}{\phi_2(\theta_2)} \right) - \left(\theta_1 + \frac{\Phi_1(\theta_1)}{\phi_1(\theta_1)} \right) \right] \phi_1(\theta_1) \phi_2(\theta_2) d\theta_2 d\theta_1 \geq 0 \qquad (7.1)$$

Para ver esto, primero notemos que el mismo argumento que nos lleva a la condición 6.14 lo podemos aplicar aquí para obtener [a lo largo de la demostración suprimimos el argumento f en $U_i(\theta_i|f)$ y simplemente escribimos $U_i(\theta_i)$]:

$$E[-\hat{t}_2(\theta_2)] = \left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} y_2(\theta_1,\theta_2)\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)}\right)\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1\right] - U_2(\underline{\theta}_2). \qquad (7.2)$$

También como la condición 6.12 implica que

$$U_1(\underline{\theta}_1) = U_1(\overline{\theta}_1) - \int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} y_1(\theta_1,\theta_2)\phi_2(\theta_2)d\theta_2 d\theta_1,$$

la condición 6.14 también implica que

$$E[-\hat{t}_1(\theta_1)] = \left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} y_1(\theta_1,\theta_2)\left(\theta_1 - \frac{\Phi_1(\theta_1)}{\phi_1(\theta_1)}\right)\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1\right] - U_1(\overline{\theta}_1). \qquad (7.3)$$

Entonces, como $y_1(\theta_1,\theta_2) = 1 - y_2(\theta_1,\theta_2)$ tenemos que

$$\begin{aligned} E[-\hat{t}_1(\theta_1)] = &\left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} \left(\theta_1 - \frac{\Phi_1(\theta_1)}{\phi_1(\theta_1)}\right)\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1\right] \\ &- \left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} y_2(\theta_1,\theta_2)\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)}\right)\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1\right] - U_1(\overline{\theta}_1). \end{aligned}$$

Pero

$$\begin{aligned} \left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} \left(\theta_1 - \frac{\Phi_1(\theta_1)}{\phi_1(\theta_1)}\right)\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1\right] &= \left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}[\theta_1\phi_1(\theta_!) + \Phi_1(\theta_1)]d\theta_1\right] \\ &= [\theta_1\Phi_1(\theta_1)]_{\underline{\theta}_1}^{\overline{\theta}_1} \\ &= \overline{\theta}_1 \end{aligned}$$

Entonces,

$$E[-\hat{t}_1(\theta_1)] = \overline{\theta}_1 - \left[\int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} y_2(\theta_1,\theta_2)\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)}\right)\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1\right] - U_1(\overline{\theta}_1). \quad (7.4)$$

Ahora el hecho de que $t_1(\theta_1,\theta_2) + t_2(\theta_1,\theta_2) = 0$ implica que $E[-t_1(\theta_1,\theta_2)] + E[-t_2(\theta_1,\theta_2)] =$

0. Así que, sumando 7.2 y 7.4 vemos que

$$[U_1(\overline{\theta}_1)-\overline{\theta}_1]+U_2(\underline{\theta}_2) = \int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2} y_2(\theta_1,\theta_2)\left[\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)}\right) - \left(\theta_1 + \frac{\Phi_1(\theta_1)}{\phi_1(\theta_1)}\right)\right]\phi_1(\theta_1)\phi_2(\theta_2)d\theta_2 d\theta_1$$

Pero la racionalidad individual implica que $U_1(\overline{\theta}_1) \geq \overline{\theta}_1$ y $U_2(\underline{\theta}_2) \geq 0$, lo que nos da la ecuación 7.1.

Paso 2: La condición 7.1 no puede satisfacerse sí $y_2(\theta_1,\theta_2) = 1$ *siempre que* $\theta_2 > \theta_1$ *y* $y_2(\theta_1,\theta_2) = 0$ *siempre que* $\theta_2 < \theta_1$.

Supongamos que sí se satisface, entonces el lado izquierdo de la condición 7.1 se puede escribir como,

$$\begin{aligned}
&\int_{\underline{\theta}_2}^{\overline{\theta}_2}\int_{\underline{\theta}_1}^{Min[\theta_2,\overline{\theta}_1]}\left[\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)} - \theta_1\right)\phi_1(\theta_1) - \Phi_1(\theta_1)\right]\phi_2(\theta_2)d\theta_1 d\theta_2 \\
&= \int_{\underline{\theta}_2}^{\overline{\theta}_2}\left[\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)} - \theta_1\right)\Phi_1(\theta_1)\right]_{\underline{\theta}_1}^{Min[\theta_2,\overline{\theta}_1]}\phi_2(\theta_2)d\theta_2 \\
&= \int_{\underline{\theta}_2}^{\overline{\theta}_2}\left[\left(\theta_2 - \frac{1-\Phi_2(\theta_2)}{\phi_2(\theta_2)} - Min\{\theta_2,\overline{\theta}_1\}\right)\Phi_1(Min\{\theta_2,\overline{\theta}_1\})\right]\phi_2(\theta_2)d\theta_2 \\
&= -\int_{\underline{\theta}_2}^{\overline{\theta}_1}[1-\Phi_2(\theta_2)]\Phi_2(\theta_2)d\theta_2 + \int_{\overline{\theta}_1}^{\overline{\theta}_2}[(\theta_2-\overline{\theta}_1)\phi_2(\theta_2) + (\Phi_2(\theta_2)-1)]d\theta_2 \\
&= -\int_{\underline{\theta}_2}^{\overline{\theta}_1}[1-\Phi_2(\theta_2)]\Phi_2(\theta_2)d\theta_2 + [(\theta_2-\overline{\theta}_1)(\Phi_2(\theta_2)-1)]_{\overline{\theta}_1}^{\overline{\theta}_2} \\
&= -\int_{\underline{\theta}_2}^{\overline{\theta}_1}[1-\Phi_2(\theta_2)]\Phi_2(\theta_2)d\theta_2 \\
&< 0
\end{aligned}$$

en donde la desigualdad se cumple porque $\overline{\theta}_1 > \underline{\theta}_2$ y $\underline{\theta}_1 < \overline{\theta}_2$. Esto contradice la condición 7.1 y completa el argumento. □

Recordando el principio de revelación para equilibrios bayesianos de Nash (Proposición 6.1), la implicación del teorema de Myerson - Satterthwaite se puede ver de la siguiente manera: Consideremos *cualquier* institución de intercambio voluntario que regule el intercambio entre el comprador y el vendedor. Esto incluye, por ejemplo, cualquier proceso de negociación en donde las partes pueden realizar ofertas y contraofertas entre ellas, así como cualquier mecanismo de arbitraje en donde las partes le reportan a un tercero sus tipos y este decide sí el intercambio ocurrirá y a qué precio. Por el principio de revelación, sabemos que la función de elección social que es indirectamente implementada en un equilibrio bayesiano de Nash de dicho mecanismo debe ser compatible con incentivos bayesianos. Más aún, como la participación en el mecanismo es voluntaria, la función de elección social $f(\cdot)$ debe satisfacer la restricción de racionalidad individual ínterin $U_1(\theta_1 \leq f) \geq \theta_1$ para toda θ_1 y $U_2(\theta_2|f) \geq 0$ para toda θ_2. Entonces, el teorema de Myerson - Satterthwaite nos dice que, bajo sus supuestos, no existen instituciones de intercambio voluntario que puedan tener un equilibrio bayesiano de Nash que de un resultado eficiente *ex post* para todas las valuaciones del vendedor y del comprador.

Podemos ver la demostración del teorema con un ejemplo sencillo para ver claramente de qué está hablando y las implicaciones que pudiera tener.

Ejemplo 7.1.

Supongamos el siguiente escenario de intercambio, un vendedor que poseé una unidad de un bien indivisible y un comprador que desea dicho bien. Sean $\theta_1 \to [0,1]$ la valuación del comprador y $\theta_2 \to [0,1]$ la valuación del vendedor y tenemos $\theta_1 = \{0{,}1, 1\}$ y $\theta_2 = \{0, 0{,}9\}$, ambos con la misma probabilidad para cada jugador. Podemos ver que en todas las combinaciones de (θ_1, θ_2) se da el intercambio, excepto en $(0{,}1, 0{,}9)$ ya que para las demás siempre $\theta_2 < \theta_1$ lo que significa que el valor que le da el comprador al bien siempre es menor que el precio que quiere el vendedor. Para facilidad de la explicación, podemos escribir los pagos como un solo precio $p(\theta_1, \theta_2)$ (el pago hecho por el comprador, recibido por el vendedor). Una vez más, vamos a suponer que en este escenario se cumple la racionalidad individual. Por lo que tenemos,

$$p(1, 0{,}9) \geq 0{,}9 \tag{7.5}$$

$$p(0{,}1, 0) \leq 0{,}1 \tag{7.6}$$

$$p(0{,}1, 0{,}9) = 0 \tag{7.7}$$

Lo que tenemos gracias al supuesto de racionalidad individual, 7.7, como habíamos visto, es claro ya que en ese caso no se realiza el intercambio. Ahora nos podemos preguntar. ¿Qué pasa con el caso restante, $p(1, 0)$? Para ver esto, debemos preguntarnos qué pasa para el vendedor y para el comprador en este caso.

Veamos el caso del vendedor. Por la compatibilidad de incentivos, él es de tipo $\theta_2 = 0$ y no le conviene pretender tener tipo $\theta_2 = 0{,}9$ tenemos,

$$\frac{p(1,0)}{2} + \frac{p(0{,}1,0)}{2} \geq \frac{p(1,0{,}9)}{2} + \frac{p(0{,}1,0{,}9)}{2}$$

lo que por 7.5, 7.6 y 7.7 implica que

$$p(1,0) + 0{,}1 \geq 0{,}9 + 0 \Rightarrow p(1,0) \geq 0{,}8.$$

Ahora, en el caso del comprador, por la compatibilidad de incentivos, le conviene ser de tipo $\theta_1 = 1$ y no tener tipo $\theta_1 = 0{,}1$ tenemos que,

$$\frac{1-p(1,0)}{2} + \frac{1-p(1,0{,}9)}{2} \geq \frac{1-p(0{,}1,0)}{2} + \frac{1-p(0{,}1,0{,}9)}{2}$$

lo que por las condiciones 7.5, 7.6 y 7.7 implican que,

$$1 - p(1,0) + 0{,}1 \geq 0{,}9 + 0 \Rightarrow 0{,}2 \geq p(1,0)$$

Por lo tanto, $0{,}2 \geq p(1,0)$ y $p(1,0) \geq 0{,}8$, que es imposible para cualquier precio, pero es la condición necesaria para que se realice el intercambio y se cumpla con la racionalidad individual de los jugadores. Lo que implica que es imposible cumplir con la eficiencia del mecanismo y al mismo tiempo con la compatibilidad de incentivos para los participantes.

Esto nos dice que, mientras exista información privada sobre los valores de los agentes, existirán ineficiencias en escenarios de intercambio voluntario y al mismo tiempo, se presentará una tensión entre los incentivos y la eficiencia.

Parte IV

Conclusiones

Tenemos que empezar por comparar el punto de vista de Arrow y el de Myerson - Satterthwaite. La teoría de elección social, y en particular el teorema de Arrow, se enfoca en cómo a partir de la agregación de preferencias, podemos obtener un resultado social a través de una función de elección social. Por otro lado, el diseño de mecanismos se preocupa por crear una institución mediante la cual, los participantes revelen su información privada para obtener un resultado social previamente esperado. A primera vista nos puede parecer que estas dos visiones son un tanto diferentes y se encargan de diferentes problemas. Ahora, después de saber qué y cómo es que funcionan ambos resultados, queremos ver cómo se relacionan.

¿Qué necesitamos para una función de elección social en el ambiente descrito por Arrow?, escrito fácilmente necesitamos un conjunto de tres o más elecciones, una regla de votación, un conjunto finito de agentes y un conjunto de soluciones sociales. Esto nos define, a través de una función de elección social el ambiente teórico que Arrow estructuró. Desde aquí, como ya vimos, se demuestra que no existe una función de elección social no dictatorial que nos de un resultado a través de una votación por mayoría. Comenzamos con un objeto algebraico, el conjunto de preferencias individuales y obtenemos un resultado dentro de otro objeto algebraico, el conjunto de soluciones sociales. Como mencionamos en la Introducción, el teorema de Arrow es un tema clásico dentro de la economía del bienestar, este sitio lo obtuvo por haber mostrado que, a partir de supuestos plausibles y de sentido común, no es posible construir un mapa de elección social que nos lleve de este conjunto de preferencias individuales acerca del conjunto social, sin pedir que algún individuo imponga sus preferencias sobre los demás. Es decir, de este objeto algebraico no existe una función que cumpla los supuestos descritos y que no sea dictatorial que tenga como imagen el conjunto de soluciones sociales.

Intuitivamente, podemos denotar que esto quiere decir que de ninguna manera (dentro del ambiente descrito), se puede lograr un "consenso" general en la población a menos que alguien se imponga, es decir, que no hay una manera de "ponernos de acuerdo".

Ahora pensemos un poco en qué implica el teorema de Myerson-Satterthwaite. Dentro de un escenario de intercambio bilateral tanto el vendedor como el comprador son incapaces de agotar las ganancias de dicho intercambio, mientras tengan información incompleta acerca del otro y exista una probabilidad positiva de que no existan ganancias externas. Una vez más, estamos diciendo que, mientras exista información privada entre los agentes no se puede realizar un intercambio bilateral que nos lleve a una situación eficiente (en el sentido de que el pago y la "satisfacción" por el intercambio son iguales entre ambos, que no estamos cumpliendo los principios de racionalidad individual y de compatibilidad con incentivos). Es decir, en otra teoría y en otro sentido, pero es una afirmación más de que no podemos "ponernos de acuerdo".

¿Por qué hacemos tanto hincapié en esto? Primeramente, si siempre existieran mecanismos para ponernos de acuerdo, entonces sería "trivial" el encontrarlos, precisamente es parte de lo que se expone en el capítulo 4 y, entonces podríamos predecir el comportamiento de los agentes para una cantidad considerable de situaciones sociales. También, podríamos encontrar procedimientos tanto de elección como de decisión que nos llevaran siempre a alguna solución deseada.

Sin mucho problema podemos ver que al comparar ambos teoremas, vemos que en general hablan de lo mismo, claramente no de la misma forma, pero sí tienen conclusiones muy similares. Además de ser parecidos matemáticamente hablando. Ambos parten de un conjunto algebraico que posee ciertas cualidades o supuestos, en el caso de Arrow, este es el conjunto o conjuntos de preferencias individuales y para Myerson-Satterthwaite el conjunto de tipos definidos por los intervalos de costo/valor del comprador/vendedor (que en sí mismos son conjuntos ordenados de preferencias definidos a partir de una distribución de probabilidad), y luego a través de una función de elección social se mapea a un elemento del conjunto de resultados o alternativas (que en el caso de Arrow es el llamado conjunto de soluciones sociales). Para consecuentemente, concluir -cada uno por su lado- que no hay una manera "eficiente" de ponernos de acuerdo.

Esto no es tan sombrío como podríamos pensar originalmente, ya que se ha hecho una gran cantidad de investigación posterior a estos dos resultados. En el caso de Arrow, existe una amplia literatura generada a partir de este resultado. Parte de esta literatura ha tratado con conjuntos de propiedades alternativos al propuesto dentro del modelo, manteniendo el marco conceptual dado por las funciones de elección social, otra parte de la literatura trata de salirse del contexto del teorema. Se ha aplicado este resultado a una variedad de problemas, como son: de agregación interpersonal, agregación de restricciones en la teoría de optimalidad en lingüística, fusión de evidencias, entre otros. Una gran porción de esta literatura tiene como objetivo la obtención de resultados positivos, es decir, la existencia de funciones de elección social no dictatoriales.

En general, si consideramos el ambiente de Arrow como adecuado y sus condiciones como indispensables, su teorema propone un serio obstáculo. Para evitarlo, necesitamos relajar alguna de las cinco hipótesis o quitar la restricción de las entradas de la regla de agregación como se hace en la llamada "agregación de bienestar", que es otra rama de investigación que se ha estado desarrollando en los últimos años.

En el caso de Myerson-Satterthwaite hay una gran variedad de avances, tanto en la teoría del diseño de mecanismos como en el resultado específico del teorema. Uno de los principales avances de los últimos años es el denominado "diseño de mecanismos combinatorios" o "diseño de mecanismos automatizados". Esta vertiente se enfoca en las diferentes consideraciones computacionales que existen a la hora de resolver o proponer un problema, el cual tiene como propósito tomar un modelo "computable" del mercado (i.e. una subasta), y encontrar una función de elección social maximizadora en un escenario discreto, utilizando procedimientos experimentales que den luz sobre el problema en cuestión. Una vertiente más en este tipo de problemas discretos son los resultados en los cuáles se elude la imposibilidad descrita por el teorema. Mientras que ésta parece mantenerse en general para escenarios con un gran número de posibles valuaciones (cuando se acerca al caso continuo), se ha encontrado que dominios con valuaciones discretas eluden en cierta medida la imposibilidad con sorprendente frecuencia (Myerson 2007).

En otros casos, aún si se mantiene la imposibilidad, a diferencia de lo que proponen Myerson y Satterthwaite (Myerson 1983), la cantidad de subsidio que se necesita para alcanzar la racionalidad individual y la compatibilidad con incentivos es relativamente pequeña (Nachbar 2011).

Como podemos ver, aunque existan tensiones entre la eficiencia y los incentivos necesarios para lograr un acuerdo, no es el final de la investigación. Existe una gran variedad de caminos que todavía falta por explorar, muchos de los cuales parten de estas ideas de imposibilidad.

Bibliografía

[1] Arrow, Kenneth, *Social Choice and Individual Values*, Wiley, 1951/1963

[2] Börgers, Tilman, *An Introduction to Mechanism Design*, Oxford University Press, 2015

[3] Cockshott, Paul, *Calculation in-Natura, from Neurath to Kantorovich*, University of Glasgow, 2008

[4] Healy, Paul J., *Institutions, Incentives and Behavior: Essays in Public Economics and Mechanism Design*, California Institute of Technology, 2005

[5] Hurwicz, Leonid, *But Who Will Guard the Guardians?*, Nobel Prize Lecture, 2007

[6] List, Christian, "Social Choice Theory", *The Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2013/entries/social-choice/

[7] Mas-Colell, Andreu, et.al, *Microeconomic Theory*, Oxford University Press, 1995

[8] Maskin, Eric S., *Mechanism Design: How to Implement Social Goals*, Nobel Prize Lecture, 2007

[9] Morreau, Michael, "Arrow's Theorem", *The Stanford Encyclopedia of Philosophy (Winter 2014 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2014/entries/arrows-theorem/

[10] Myerson, Roger B., *Game Theory: Analysis of Conflict*, Harvard University Press, 1991

[11] Myerson, Roger B., *Perspectives on Mechanism Design in Economic Theory*, Nobel Prize Lecture, 2007

[12] Myerson, Roger B., Satterthwaite, Mark A., "Efficient Mechanisms for Bilateral Trading", *Journal of Economic Theory 29*, 265-281, 1983

[13] Myerson, Roger B., "Optimal Auction Design", *Mathematics of Operation Research*, 6, 58-73, 1981

[14] Nachbar, John, *Notes on the Myerson-Satterthwaite Theorem*, University of Washington, 2011

[15] Palda, Filip, *The Apprentice Economist: Seven Steps to Mastery*, Cooper Wolfling, 2013

[16] Parkes, David C., *Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency*, University of Pennsylvania, 2001

[17] Rasmusen, Eric, *Games and Information: An Introduction to Game Theory*, Blackwell Publishing, 1989

[18] Reny, Philip J., "Arrow's Theorem and the Gibbard-Satterthwaite Theorem: a unified approach", *Economics Letters*, 70, 99-105, 2001

[19] Vaughn, Karen, "Economic Calculation under Socialism: The Austrian Contribution", *Economic Inquiry 18*, p. 537, 1980