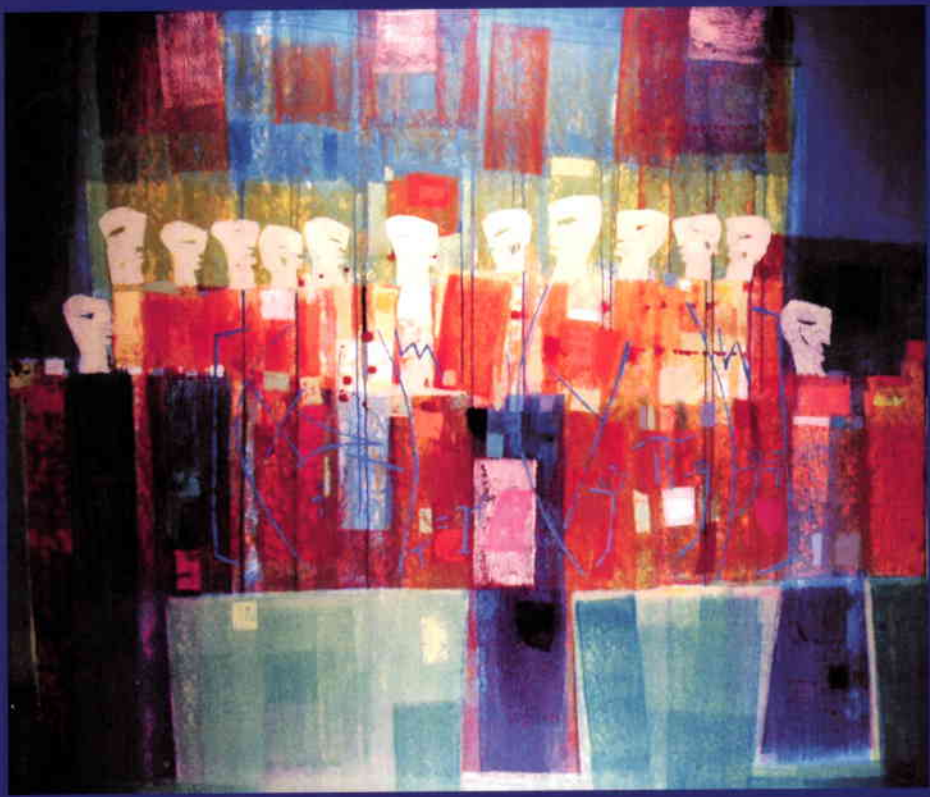

DECISIONES SOCIALES

ANTONIO VILLAR

DECISIONES SOCIALES

Antonio Villar

Mc Graw Hill

MADRID • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA • MÉXICO
NUEVA YORK • PANAMÁ • SAN JUAN • SANTIAGO • SÃO PAULO
AUCKLAND • HAMBURGO • LONDRES • MILÁN • MONTREAL • NUEVA DELHI • PARÍS
SAN FRANCISCO • SIDNEY • SINGAPUR • ST. LOUIS • TOKIO • TORONTO

CAPÍTULO

I

«The precision which allows us to get this also serves the usual function of theory in making precise what is not asserted as well as what is asserted»

R. CORNWALL
(Introduction to the Use of General Equilibrium Analysis)

I. Introducción

La racionalidad es un principio metodológico para el análisis de los comportamientos de individuos, instituciones y grupos sociales. Con él tratamos de comprender la lógica de sus decisiones, valorar sus conductas y, en ciertos casos, anticipar las acciones que emprenderán en diferentes circunstancias. El punto de partida es la consideración de que los comportamientos son intencionales (derivan de la existencia de ciertos fines) o responden a la aplicación de ciertos principios (éticos y operativos). Hablamos de «decisiones racionales» cuando existe coherencia entre las finalidades de los agentes y las decisiones que adoptan. Un agente racional actúa escogiendo de forma sistemática aquellas alternativas que mejor se ajustan a sus propósitos, en cada contexto en que debe tomar una decisión. Por supuesto las cualificaciones «de forma sistemática», «que mejor se ajustan a sus propósitos» y «en cada contexto», son claves para el desarrollo de una teoría relevante al tiempo que permiten toda una serie de formulaciones alternativas. Las páginas que siguen se ocupan, precisamente, de dar un contenido más preciso a estos aspectos.

En todo problema de decisión encontramos siempre tres ingredientes básicos. El primero, un ***agente*** (un individuo o un colectivo) que tiene que elegir un curso de acción. El segundo, una colección de alternativas que constituye su ***conjunto de elección*** y que puede estar sujeto a ciertas *restricciones* (normas, renta disponible, habilidades requeridas, información accesible, etc.). Por último, algún ***criterio de valoración*** que permite al agente comparar las diferentes alternativas.

Existe una enorme variedad de problemas de decisión y éstos aparecen en contextos muy diversos. Efectuar una inversión, escoger un sombrero o elegir un alcalde son problemas de decisión muy diferentes, pero en todos ellos encontramos los tres ingredientes básicos señalados: uno o varios agentes que toman decisiones (una empresa, un comprador, los ciudadanos de un municipio), un conjunto de alternativas donde elegir (diferentes proyectos de inversión, distintos modelos y tallas de sombreros, una colección de candidatos), un criterio de valoración (ordenación de las alternativas), y tal vez algunas restricciones (las inversiones que podemos pagar, los sombreros de la talla del comprador, los candidatos de la orientación política del votante).

Las diferencias relevantes entre los distintos tipos de problemas de elección no tienen tanto que ver con los objetos entre los que tenemos que elegir (elementos del conjuntos de elección), como con *el contexto* del problema. En este contexto incluimos aspectos como los siguientes: si el agente decisor es un individuo o un colectivo, si hay incertidumbre o no sobre la naturaleza de las alternativas o las consecuencias de nuestras decisiones, si el problema tiene aspectos dinámicos (consecuencias sobre el futuro o implicaciones del pasado), si las decisiones del agente generan interacciones estratégicas con otros agentes, etc.

La diferencia contextual en la que haremos hincapié en esta monografía se refiere a los **problemas de decisión individual** frente a los **problemas de decisión colectiva**.

I.1 Racionalidad individual, racionalidad colectiva y racionalidad de sistemas sociales

En el problema estándar de **decisión individual** encontramos un agente que actúa de forma aislada aplicando un criterio de valoración bien definido sobre el universo de alternativas a su disposición. La racionalidad está asociada a que este criterio de elección sea aplicable de forma efectiva y sistemática a los diversos problemas de decisión con los que se enfrenta. La idea de elección racional refleja una concepción de *racionalidad como procedimiento* y no como esquema de valores; es decir, la racionalidad se identifica con la coherencia entre las decisiones que adopta un individuo y su forma de valorar las alternativas con las que se enfrenta. No hay pues juicio acerca de los criterios valorativos de un agente, que se suponen datos del problema y habitualmente ajenos al análisis. Nada hay de «irracional» en que uno prefiera las hamburguesas del MalcDonald's a la lubina a la sal de un restaurante de lujo. Lo que sería irracional sería tomar lubina cuando a uno le gustan más las hamburguesas, si ambas opciones están disponibles y no hay otros condicionantes. La racionalidad de un agente decisor consiste simplemente en *escoger en cada circunstancia la mejor opción accesible de acuerdo con su propia valoración de las alternativas y la información disponible.*

Los problemas de **decisión colectiva** involucran a una colección de individuos que deben elegir entre un conjunto de alternativas que les afectan como grupo. Decidir el alcalde en un municipio, elegir una nueva vivienda en el seno de una familia o repartir los fondos de ayuda al desarrollo entre los países necesitados, son ejemplos de estas situaciones. En este tipo de problemas se supone habitualmente que cada uno de los individuos involucrados tiene una valoración completa y sistemática de las opciones del grupo. La aproximación más general al tema parte de la idea de agregar esos criterios individuales en un criterio social que permita hacer también comparaciones sistemáticas de alternativas desde un punto de vista colectivo. Aquí la idea de racionalidad colectiva pasa a tener un *componente normativo* vinculado a los principios que permiten obtener una valoración social de las diferentes alternativas a partir de las valoraciones individuales. Veremos que el paso de las preferencias individuales a las preferencias sociales está plagado de dificultades.

Estas dificultades han llevado a una aproximación algo diferente en la resolución de una amplia familia de problemas de decisión colectiva (en particular en problemas de repartos de ganancias o pérdidas conjuntas). Se trata de abordar el problema de decisión desde la perspectiva de un juez imparcial que recurre a la aplicación de ciertos principios éticos y operativos. De este modo se persigue identificar de forma unívoca el criterio de decisión colectiva con el conjunto de juicios de valor que consideramos deseables.

Un tercer nivel de análisis sobre las decisiones racionales, que en cierto modo es un híbrido de los dos anteriores, se refiere al caso en que consideramos la actuación individual dentro de un ***sistema social***. Se trata de un contexto en el que un conjunto de agentes toman decisiones de naturaleza esencialmente individual, pero cuyas acciones resultan interdependientes porque las decisiones privadas de un individuo afectan, directa o indirectamente, a otros. Pensemos en el caso de la actividad económica de un país, en el que familias, empresas y administraciones públicas toman decisiones independientes en busca de sus propios objetivos, pero que terminan afectando a los otros agentes. También podemos interpretar como un sistema social un municipio, una empresa, una comunidad de vecinos o, en el caso más simple, los intervinientes en un contrato bilateral. En todos estos escenarios observamos la presencia de individuos que persiguen sus propias metas pero cuyas acciones condicionan en parte (y son condicionadas por) las posibles acciones de los demás.

Así pues, aunque en algunos casos el problema de decisión individual puede formularse como un problema específico de un agente, aislado del resto de la sociedad, no siempre es así. Muchas de las decisiones individuales se toman en contextos en los que los agentes interactúan sobre un mismo escenario (lo que hemos llamado un *sistema social*). Esta interacción puede ser explícitamente reconocida y, por consiguiente, formar parte de los datos con los que el agente formula sus decisiones (típicamente cuando se toman decisiones en grupos relativamente pequeños). El elemento nuevo aquí es que las consecuencias de la elección de un individuo dependen directamente de las decisiones adoptadas por los demás de modo que las alternativas entre las que el individuo elige pueden interpretarse como «estrategias» (cursos de acción condicionales a las posibles acciones de los demás).

La valoración de la racionalidad de un sistema social se refiere al resultado de la acción de un conjunto de individuos que toman decisiones sobre variables que le conciernen en exclusiva, aunque puedan generar efectos indirectos sobre los demás. Aquí los dos elementos esenciales que son objeto de análisis se refieren a la posible compatibilidad de todas estas acciones individuales tomadas de forma independiente, por una parte (lo que los economistas llamarían «equilibrio»), y a la racionalidad social de los resultados globales cuando miramos el problema desde el punto de vista colectivo (típicamente consideraciones relativas a la eficiencia o la equidad).

I.2 La información y el tiempo

Otro aspecto fundamental que distingue los tipos de problemas de decisión que podemos encontrar se refiere a la ***información***, tanto en lo relativo a su disponibilidad como a la capacidad de procesamiento de los agentes.

El caso más sencillo se refiere a un problema de decisión individual donde el agente decisor dispone de toda la información relevante y es capaz de procesarla de forma completa, rápida y sin costes. En este contexto adoptar una decisión equivale a determinar un resultado. Pero muchos de los problemas de decisión individual, especialmente aquellos que involucran varios periodos de tiempo, se caracterizan por la existencia de *incertidumbre*.

Cuando hay incertidumbre adoptar una decisión y obtener un resultado son dos cosas distintas. Al tomar decisiones como la adquisición de valores en bolsa, la elección de una carrera universitaria o votar el Gobierno de un país, uno escoge con precisión la alternativa pero las consecuencias que se derivan de esa elección son aleatorias, es decir, dependen en parte de una serie de sucesos incontrolables. Adoptar una decisión en un contexto incierto equivale a elegir una «lotería», en el sentido de escoger una colección de posibles resultados con ciertas probabilidades. Adviértase que en este caso la racionalidad es un concepto «ex ante», es decir, se refiere a adoptar la mejor decisión posible antes de conocer la consecuencia exacta que en el futuro se derivará de la acción que emprendemos hoy. Que luego las cosas no sean como esperábamos no significa que hiciéramos mal en el momento en que tomamos la decisión (¡todo el mundo sabría qué número de lotería comprar el día después del sorteo!).

En los problemas de decisión colectiva encontramos una dificultad adicional: la necesidad de recoger y procesar la información sobre las valoraciones individuales de las alternativas sociales. Esto supone incurrir en costes que pueden ser sustantivos a la hora de elaborar criterios de valoración social a partir de los criterios individuales. Pero puede suceder, además, que los individuos no quieran suministrar la información necesaria o la falseen si con ello pueden beneficiarse. Surge así una nueva dimensión en los problemas de decisión colectiva: la posibilidad de que los individuos se comporten estratégicamente en la manifestación de sus valoraciones (lo que se conoce como *manipulación de las preferencias*).

Estos problemas de manipulación de las preferencias también pueden aparecer relacionados con la racionalidad de los sistemas sociales. Pero en este contexto hay un nuevo elemento relevante que debemos considerar: la posibilidad de que los agentes que interactúan tengan distinta información sobre los objetos de su relación (por ejemplo las partes de un contrato, Hacienda y los contribuyentes, los Gobiernos autonómicos y el Gobierno central, etc.). Los problemas de *información asimétrica* afectan de forma sustantiva a los resultados del funcionamiento de los sistemas sociales.

La dimensión temporal de los problemas de decisión es en general un aspecto relevante. Podemos encontrar así tanto **problemas estáticos** como **problemas dinámicos**. Los primeros son aquellos en que la decisión se refiere a un momento concreto del tiempo, sin tomar en cuenta las implicaciones futuras. En muchos casos estos problemas estáticos constituyen una drástica simplificación de la realidad que resulta útil para hacer más manejable el análisis. En los problemas dinámicos las decisiones en un determinado momento del tiempo pueden tener consecuencias sobre periodos posteriores y también estar condicionadas por decisiones precedentes. Pero, sobre todo, implican la concepción de las decisiones racionales como un *proceso* en el que la nueva información generada en cada momento del tiempo se va incorporando al mecanismo de toma de decisiones.

I.3 Avance de contenidos

En esta monografía hemos seleccionado un subconjunto de los temas que configuran el campo de la teoría de la decisión, con el énfasis puesto en la racionalidad colectiva considerada en un contexto estático y en condiciones de información completa.

El Capítulo II proporciona un repaso a la teoría de la decisión individual. Partimos de la modelización de un agente cuyo criterio de valoración viene descrito por una relación de preferencias (una relación binaria transitiva y completa que permite comparar alternativas en términos de «ser mejor o igual que»). Repasamos a continuación la formulación del problema de decisión en términos de la maximización de una función objetivo (la *función de utilidad*) y extendemos el análisis al caso de problemas de decisión sin certidumbre. Nos ocupamos también, aunque muy brevemente, de presentar la idea de equilibrio de un sistema social (como el equilibrio de Nash de un juego no cooperativo). Termina el capítulo con una discusión acerca del principio de racionalidad, a partir de algunos ejemplos concretos en los que dicho principio no parece guiar la acción.

Los Capítulos III y IV abordan el problema de racionalidad colectiva en el contexto más general: cómo realizar valoraciones sociales a partir de la agregación de las valoraciones individuales de los miembros de la sociedad. Comenzamos por ilustrar las dificultades que presenta este programa a partir de un sencillo ejemplo. Luego presentamos el conocido como «Teorema de Imposibilidad de Arrow» que establece que no existe ninguna forma general, eficiente y democrática de agregar las preferencias individuales en preferencias sociales. A partir de este resultado tan negativo procedemos a explorar las vías de escape al mismo merced a ciertas modificaciones en la forma de plantear el problema por parte del Premio Nobel Keneth Arrow. Comprobaremos que la forma efectiva de escapar al Teorema de Imposibilidad es mediante la realización de comparaciones interpersonales de utilidad. Pero también veremos que éste es un terreno conceptualmente resbaladizo y que, aun así, no son muchas las formas de agregar las preferencias individuales en preferencias sociales. Cerramos el capítulo con algunas ilustraciones y una discusión sobre la naturaleza de los resultados presentados.

El Capítulo V se ocupa de un problema de decisión colectivo particular: el reparto de las ganancias que se pueden derivar de la cooperación entre los miembros de un grupo. La teoría de los juegos cooperativos es el instrumento básico con el que abordamos esta problemática. Se analizan dos escenarios diferenciados. El primero, al que denominamos cooperación simple, se refiere al caso en que la cooperación se deriva del acuerdo unánime de la partes; si no hay unanimidad no hay cooperación alguna. Este contexto puede mirarse como una generalización de la negociación bilateral en la que éstos son los únicos resultados posibles. El segundo escenario considerado es el de la cooperación multilateral. La diferencia es que en este contexto también son posibles acuerdos parciales entre coaliciones (subgrupos). En el escenario de cooperación simple discutimos las soluciones arbitradas de Nash y Kalai-Smorodinsky. En el caso de cooperación multilateral analizamos los conceptos de núcleo, nucleolo y valor de Shapley. El capítulo termina con un par de aplicaciones.

El Capítulo VI se ocupa de un problema todavía más particular: Cómo repartir una cantidad dada de un bien entre un conjunto de individuos cuyos derechos totales exceden la cantidad disponible. Se trata de una generalización de problema de bancarrota en el que el valor de liquidación de una empresa es insuficiente para cubrir las deudas contraídas. Estudiamos aquí, desde un punto de vista axiomático, las diferentes soluciones a este tipo de problemas (reparto proporcional, reparto de ganancias o pérdidas iguales, o la regla del Talmud).

En el Capítulo VII se estudia cómo los diferentes criterios de valoración solucionan un problema convencional consistente en sufragar el coste de una obra pública que, una vez realizada, puede ser usada libremente por todos los ciudadanos. Tomando como referencia un ejemplo extremadamente sencillo procedemos a estudiar las recomendaciones de las distintas soluciones y a comparar los criterios que aplican en un caso concreto.

Termina el libro con un capítulo en el que discutimos las dificultades de poner en práctica las decisiones sociales cuando éstas dependen de una información que suministran los individuos y no puede contrastarse. Analizamos en este capítulo tanto las posibilidades de que existan reglas de decisión que induzcan a los agentes a decir la verdad, como el diseño de mecanismos que generen los resultados adecuados, tomando en cuenta explícitamente el comportamiento estratégico de los agentes.

CAPÍTULO

II

«Quiero vengarme lo antes posible,
antes de que se me pasen las ganas,
antes de que se me alivie la saña»

RAMÓN BUENAVENTURA
(El último Negro)

II. Racionalidad individual

El tipo de situación que queremos analizar se refiere a un ***agente individual*** que se enfrenta a un problema consistente en escoger alguna alternativa entre una serie de opciones disponibles. Llamaremos ***conjunto de elección*** a esta colección de opciones y lo denominaremos genéricamente como X. Los elementos de este conjunto (las alternativas entre las que tiene que elegir) serán nombrados como x, y, z,...

La naturaleza de estas alternativas x, y, z,... variará dependiendo del problema de elección de que se trate. Pueden referirse a opciones unidimensionales o multidimensionales. Un ejemplo del primer caso es el de la elección de un sombrero en una tienda, o la decisión sobre qué candidato votar como presidente de la comunidad de vecinos. En otros casos, sin embargo, cada opción está constituida por varios objetos o personas. Así en la decisión sobre la compra semanal de una familia las alternativas están compuestas por «cestas de la compra» (vectores de mercancías diversas). También en la decisión sobre el voto en unas elecciones nacionales las alternativas vienen generalmente dadas por colecciones de candidatos que presenta cada partido.

En ocasiones el problema de elección se formula no sobre un conjunto X fijo sino sobre una parte de X que resulta dependiente de ciertas variables. Por ejemplo si se trata de elegir un sombrero en una tienda, podemos definir X como el conjunto de todos los sombreros disponibles en esa tienda que no cuesten más de una cierta cantidad. En tal caso el conjunto X depende del precio de los sombreros que haya fijado el vendedor. Si cambian los precios cambiará este conjunto. En situaciones de este tipo suele distinguirse entre el universo de alternativas disponibles (todos los sombreros de la tienda en este ejemplo) y las opciones accesibles (el conjunto alcanzable). La diferencia entre uno y otro viene dada por la presencia de ciertas restricciones (la cantidad máxima de dinero que queremos gastarnos, en este caso).

Para simplificar la discusión en lo que sigue tomaremos X como un conjunto dado, que se interpreta como el conjunto de alternativas en el que efectivamente podemos elegir (un conjunto en el que ya hemos incorporado las restricciones pertinentes). No hay nada sustancial del siguiente análisis que dependa de esta simplificación.

Hay varias formas de aproximarse al estudio de la racionalidad de las decisiones individuales. Una, suponer que el individuo es capaz de ordenar todas las alternativas de su conjunto de elección y asociar la racionalidad a la elección de la alternativa que ocupa la primera posición del ranking. Una variante de esta formulación consiste en suponer que el individuo busca alcanzar la mayor satisfacción posible de modo que en su elección se guía por el principio de maximizar su utilidad o bienestar personal. En ambos casos la noción de elección racional se vincula al principio de elegir aquella opción que resulte «la mejor de las alcanzables». La diferencia es que en el segundo caso se supone que el agente es capaz de dar valores numéricos a las diferentes alternativas. Una aproximación diferente consiste en estudiar las decisiones que realiza el agente y relacionar la racionalidad con la consistencia de sus acciones, en el sentido de que sus elecciones no resulten erráticas (por ejemplo en un problema elige x frente a y pero en otro elige y cuando x está también disponible).

Repasaremos brevemente estas aproximaciones al estudio del comportamiento racional del individuo, considerando primero que el problema de decisión se plantea en condiciones de certidumbre. Luego nos ocuparemos de reformular el problema de decisión en un contexto incierto.

II.1 Preferencias

Una forma general de establecer un criterio de comparación entre alternativas consiste en la introducción de algún tipo de ordenación sobre los elementos de X. Aludiremos a este criterio de comparación con el nombre genérico de ***preferencias*** e interpretaremos que reflejan la forma en que el individuo valora las diferentes alternativas.

Podemos formalizar las preferencias del individuo como una relación binaria $\succsim$ definida sobre el conjunto de alternativas X, que nos permite comparar pares de alternativas en términos de ***ser mejor o igual que***. Es decir, dados dos elementos x, y de X, la expresión $x \succsim y$ significa que este individuo estima que la alternativa x es *al menos tan buena como* (mejor o igual que) la alternativa y.

Para conseguir una modelización operativa de los agentes racionales debemos suponer que las preferencias cumplen unas mínimas regularidades. Con objeto de simplificar la discusión al máximo nos centraremos en dos propiedades de esta relación de preferencias, la *completitud* y la *transitividad*, que resultan de la aplicación con independencia de la naturaleza del conjunto de alternativas X. Estas propiedades reflejan la idea de un agente que es capaz de valorar de forma coherente cualquier par de opciones.

Observación.- *Hay muchas otras propiedades interesantes que tienen sentido para conjuntos de elección con una cierta estructura formal*[1]. *Si el lector así lo prefiere puede tomar como referencia intuitiva problemas de decisión en los que X está constituido por un conjunto finito de alternativas.*

[1] Entre ellas destacan las propiedades de *continuidad*, *convexidad* y *monotonía*, que tienen sentido cuando el conjunto de elección tiene una estructura matemática particular (cada alternativa viene dada por un vector de números reales). Para una discusión de estas propiedades véase Villar (1999, Cap. 2). Presentamos un esbozo informal de estas propiedades en el Apéndice a este capítulo.

La completitud establece que la relación de preferencias es aplicable a cualquier par de alternativas que consideremos. Es decir, descarta la posibilidad de que existan opciones incomparables (dicho informalmente, cuando preguntamos a un individuo sobre su valoración de dos alternativas, no puede acogerse al «no sabe / no contesta»).

Formalmente[2]:

Completitud: *Para todo par de alternativas x, y del conjunto X, se verifica $x \succsim y$ o bien $y \succsim x$.*

La transitividad, por su parte, postula la coherencia en el comportamiento del agente y garantiza la ordenación sistemática de las alternativas. Establece que si una opción es mejor o igual que otra, y esta última es mejor o igual que una tercera, entonces la primera será mejor o igual que la tercera.

Formalmente:

Transitividad: *Para todo x, y, z de X, $x \succsim y$ & $y \succsim z$ implica $x \succsim z$.*

Las propiedades de completitud y transitividad de las preferencias implican que el criterio valoración del agente resulta exhaustivo y sistemático. Lo importante para la teoría es que las preferencias del sujeto tengan estas propiedades sin entrar en el análisis de los motivos que le llevan a valorar las cosas de una determinada forma.

A partir de la relación $\succsim$ podemos definir una nueva relación sobre los elementos del conjunto de alternativas. Se trata de la ***relación de indiferencia***, que representamos por el símbolo $\approx$, y que se define como sigue: dados x, y de X, $x \approx y$ si y sólo si $x \succsim y$ al tiempo que $y \succsim x$. Esta expresión, que se lee como x es *indiferente* a y, nos indica que las opciones x e y son igualmente valoradas por el sujeto.

Observación.- *Adviértase que los conceptos de «indiferente» e «incomparable» son distintos en términos lógicos. En efecto, el primero nos dice que $x \succsim y$ al tiempo que $y \succsim x$, mientras que el segundo nos dice que ni x es mejor o igual que y, ni lo contrario.*

Puede comprobarse fácilmente que, cuando la relación «ser mejor o igual que» es completa y transitiva la relación de indiferencia es *reflexiva* (toda alternativa es tan buena como sí misma), *simétrica* (es decir, si x es tan buena como y, entonces y es tan buena como x), y *transitiva* (x indiferente a y e y indiferente a z implica que x es indiferente a z). Por tanto la relación $\approx$ constituye una *relación de equivalencia*, lo que comporta que podemos clasificar (en sentido estricto) todas las alternativas en grupos que contienen colecciones de alternativas indiferentes[3].

[2] Obsérvese que el axioma de completitud, tal y como se define a continuación, implica que la relación de preferencias es *reflexiva* (dado que podríamos tomar $x = y$).

[3] Cuando hablamos de «clasificar en sentido estricto» queremos decir que cada elemento de X está en una y sólo una de estas categorías de suerte que la unión de todas ellas es precisamente el conjunto X.

Así, dada una alternativa x de X definimos como $I(x)$ el conjunto de todas las opciones de X que resultan indiferentes a x. Como la relación $\approx$ es reflexiva, x siempre pertenece a $I(x)$, de modo que cada uno de estos conjuntos tiene al menos un elemento. Los conjuntos $I(x)$ se conocen como ***clases de indiferencia***. Una clase de indiferencia contiene todas las opciones que resultan igualmente apreciadas (indiferentes) por el individuo a una alternativa x dada.

Observación.- *A veces la transitividad de la relación de indiferencia resulta cuestionable, porque tiene implicaciones difícilmente aceptables. Un ejemplo tradicional es el siguiente: Supongamos que le pedimos a un individuo amante del café que escoja entre dos tazas de café con azúcar. La única diferencia entre ambas es que una tiene una millonésima de gramo más de azúcar que la otra. Cabe esperar que el individuo sea incapaz de distinguir entre ambas tazas de café y se declare indiferente entre ellas. Si repetimos la prueba un número suficiente de veces, variando por millonésimas de gramo de azúcar resultará, por la transitividad de la indiferencia, que le da lo mismo el café sin azúcar que con 20 cucharadas.*

A partir de las relaciones y $\approx$ podemos a su vez definir una nueva, la relación de ***preferencia estricta***, que simbolizaremos por . Se define como sigue: una opción x es *estrictamente mejor* que una opción y, lo que escribimos como x y, si y sólo si, x y pero no es cierto que $x \approx y$. Diremos en este caso que x es *preferido* a y.

La relación es transitiva y verifica la propiedad de ***asimetría*** (si una alternativa x es mejor que otra y, entonces y no puede ser mejor que x). Obsérvese que podemos ordenar de forma estricta y completa las clases de indiferencia mediante la relación .

Es fácil comprobar que cuando el conjunto de alternativas es finito y la relación de preferencias de un agente verifica las propiedades de completitud y transitividad, siempre existirá una opción que sea mejor o igual que todas las demás. Y que si existe más de una, todas ellas serán indiferentes. Por tanto, el problema de elección del agente se resuelve escogiendo un elemento máximo de esta relación de preferencias.

Hay dos propiedades de racionalidad más débiles que la transitividad de la relación de preferencias que, no obstante, garantizan la existencia de solución al problema de decisión del agente individual. Se trata de la *casi-transitividad* y de la *aciclicidad*.

Una relación de preferencias se dice ***casi-transitiva*** cuando la relación de preferencia estricta cumple la transitividad (pero la relación de indiferencia no necesariamente la cumple).

Cuando una relación de preferencias es casi-transitiva se cumple que si x es mejor que y e y es mejor que z entonces x es mejor que z. Pero no se sigue necesariamente que si x es indiferente a y e y es indiferente a z resulte que x sea indiferente a z).

Una relación de preferencias se dice ***acíclica*** cuando se cumple la siguiente condición: Para toda colección (finita) de alternativas, $x_1, x_2, ..., x_k$, si resulta que x_1 es mejor que x_2, x_2 es mejor que x_3, ..., y así sucesivamente hasta llegar a que x_{k-1} es mejor que x_k, entonces x_k no puede ser mejor que x_1.

Una relación de preferencia acíclica es aquella en la que no hay ciclos de preferencia estricta. Así, para el caso de tres alternativas, x, y, z, si x es mejor que y e y es mejor que z la aciclicidad establece que z no puede ser mejor que x. Pero no impide que x sea indiferente a z (frente a la conclusión de que x sea mejor que z, que se derivaría tanto de la transitividad como de la casi-transitividad de las preferencias).

Cuando no se cumple la aciclicidad de las preferencias se pierde la lógica del criterio de ordenación y pueden aparecer «ciclos» de preferencia que hagan imposible tomar decisiones. Si tomamos, a modo de ejemplo, un conjunto con tres elementos, $X = \{x, y, z\}$, donde resulta que x es mejor que y, y es mejor que z, al tiempo que z resultara ser mejor o igual que x, nunca sabríamos con qué opción quedarnos.

Es fácil comprobar que la transitividad implica la casi-transitividad y que la casi-transitividad implica la aciclidad.

En buena parte de la discusión que desarrollamos a continuación supondremos que el agente individual posee una relación de preferencias que satisface estas dos propiedades básicas (preferencias transitivas y completas), de modo que siempre será capaz de escoger alguna opción frente a cualquier conjunto (finito) de alternativas con el que se enfrente.

En este contexto la decisión racional del agente individual que se enfrenta con un problema de elección consiste en escoger una opción que sea mejor o igual que todas las demás, según su relación de preferencias. El ***comportamiento*** de un agente podemos así identificarlo como las distintas decisiones que adopta cuando se enfrenta a diferentes conjuntos de alternativas.

En este enfoque del problema partimos de las preferencias como concepto primitivo y explicamos las decisiones a partir de la selección de elementos máximos de esta relación de preferencias. Es decir, la alternativa x^* es elegida en el conjunto X si se cumple que x^* y para todos los elementos y de X.

Como hemos señalado hay otras dos aproximaciones diferentes al problema de elección racional de un individuo, si bien están estrechamente relacionadas con la que acabamos de presentar. La primera, que tiene sus orígenes en el utilitarismo clásico desarrollado en los siglos XVIII y XIX, consiste en definir el comportamiento de los individuos como la maximización de una función de utilidad o bienestar personal. También aquí el comportamiento de los agentes se obtiene como la resultante de una función de valoración de alternativas. La segunda, toma el camino inverso; es decir, parte de observar las decisiones de los agentes y trata de entender la lógica de las mismas.

II.2 La utilidad

Una aproximación ligeramente distinta al comportamiento racional de los individuos es aquella que se basa en la idea de utilidad. El agente actúa movido por la búsqueda de su propia satisfacción, que se puede describir mediante una función que asocia valores numéricos a las diferentes alternativas. Esta función se conoce como ***función de utilidad***.

Hay dos concepciones diferentes de la utilidad, a las que podemos denominar *utilidad cardinal* y *utilidad ordinal*. La utilidad cardinal está vinculada al movimiento utilitarista de los siglos XVIII y XIX y parte de la noción de utilidad como concepto primitivo. La utilidad ordinal, iniciada a principios del siglo XX por Vilfredo Pareto, identifica utilidad como una mera representación de la ordenación de preferencias. Ambas concepciones son compatibles, útiles y frecuentes. En favor de la utilidad ordinal está la «parsimonia». En favor de la utilidad cardinal la precisión de los resultados y su aplicabilidad a contextos específicos.

La concepción clásica de la utilidad: utilidad cardinal

Los primeros modelos de comportamiento individual partían de la idea de que los agentes elegían sus acciones con el fin de *maximizar su utilidad*. Esta noción de utilidad se identificaba con conceptos como bienestar, satisfacción, felicidad, beneficio propio, etc. La idea subyacente era que el individuo actuaba movido por la búsqueda de su mayor satisfacción o felicidad[4]. A partir de aquí surgía la pregunta de si las instituciones sociales y económicas eran capaces de canalizar estos comportamientos individuales hacia un resultado social satisfactorio o por el contrario abocaban a la sociedad al caos.

La idea de concebir el comportamiento como resultado de la búsqueda de la mayor felicidad se formulaba a partir de la consideración de que cada individuo poseía una función de utilidad que describía numéricamente el valor asociado a cada posible alternativa. Así, dado un conjunto de alternativas X el agente asociaba a cada opción x de este conjunto el número $u(x)$ que constituía una medida de la satisfacción que esta alternativa le proporcionaba. En cada problema de decisión el individuo escoge la alternativa que le da mayor satisfacción, es decir, la opción que tiene asociado el valor más alto posible de la función $u(.)$.

Dado que la satisfacción individual es un criterio de valoración eminentemente subjetivo, aparecen inmediatamente algunas preguntas en relación a la medición de la utilidad. ¿En qué unidades se mide la satisfacción de un individuo? ¿Es importante el punto de referencia desde el que se mide la satisfacción? ¿Todos los individuos miden la satisfacción del mismo modo?

En la concepción clásica se suponía que la utilidad era una variable «cardinal», es decir, susceptible de medición en un sentido similar a la distancia o la temperatura. El caso de la temperatura es especialmente ilustrativo, porque los valores de nuestra medición dependen de las *unidades de medida* y el *origen*. Los sistemas de medición de temperatura Celsius (es decir, medida en grados centígrados como es más común en Europa) y Farenheit (el más usado en Estados Unidos) son un ejemplo de ello. Podemos cambiar la forma en que expresamos el valor de la temperatura sin que la temperatura cambie. Las unidades de medida de los diferentes sistemas nos dicen qué magnitud de cambio en la temperatura vale una unidad, mientras que fijar el origen equivale a determinar el valor que damos a una temperatura singular (típicamente el valor que asociamos al punto de congelación del agua, que juega el papel del «cero» de la escala en el sistema Celsius)[5].

En términos de la utilidad las unidades de medida nos dicen también para qué incremento de satisfacción diremos que la utilidad ha aumentado en una unidad. Y fijar el «origen» equivale a dar un valor concreto a una alternativa distinguida (por ejemplo decir que la utilidad asociada a la peor opción disponible es igual a cero).

Obsérvese que los posibles cambios de origen y de unidades en la medición de la utilidad no afectan a las alternativas que resultan mejores. Así, dadas dos alternativas x, y del conjunto X, tendremos que $u(x) > u(y)$ (es decir, la alternativa x da más utilidad que la alternativa y), si y sólo si:

$$t\,u(x) + a > t\,u(y) + a$$

para cualquier número positivo t que representa un cambio de unidades, y cualquier número a que representa un cambio de origen[6].

Es también interesante advertir que con este tipo de utilidad uno puede hacer comparaciones de «intensidad de satisfacción» en el siguiente sentido: la alternativa x me gusta más que la alternativa z, y *mucho más* que la alternativa y». Este tipo de comparaciones podemos expresarlas del siguiente modo:

(a) $u(x) > u(z)$
(b) $u(x) > u(y)$
(c) $u(x) - u(y) > u(x) - u(z)$

Ya hemos visto que un cambio de origen y de unidades no afecta a la relación x mejor que z ni a la relación x mejor que y, recogidas en (a) y (b). Veamos ahora que la expresión (c) tampoco se ve alterada por estos cambios. En efecto, cambiando el origen y las unidades en que medimos la utilidad tendremos:

$$(t\,u(x) + a) - (t\,u(y) + a) > (t\,u(x) + a) - (t\,u(z) + a)$$

[4] Lo que no necesariamente implica comportamientos egoístas, dado que parte de la satisfacción de un individuo puede derivar de la felicidad de los demás.

[5] En la escala Fahrenheit el agua hierve a 212 grados (equivalente a 100 grados Celsius) y se congela a 32 grados (equivalente a cero grados Celsius). Cada grado Celsius se transforma así en grados Fahrenheit mediante una doble operación: cambiando el origen y cambiando la escala. En concreto, F = 32 +1,8C (donde F son los grados Fahrenheit y C los grados Celsius).

[6] Obsérvese que, si restamos en la expresión $t\,u(x) + a > t\,u(y) + a$, el número a en los dos miembros de la desigualdad tendremos $t\,u(x) > t\,u(y)$. Si ahora dividimos ambos términos por t, que es un número mayor que cero, llegamos a: $u(x) > u(y)$.

Operando y reordenando términos:

$$t\,u(x) + a - t\,u(y) - a > t\,u(x) + a - t\,u(z) - a$$

eliminando la constante a que suma y resta en cada lado de la desigualdad y sacando factor común t, obtenemos:

$$t[\,u(x) - u(y)] > t[\,u(x) - u(z)]$$

de donde, dividiendo ambos términos por t, que es un número positivo, llegamos a

$$u(x) - u(y) > u(x) - u(z)$$

que es precisamente la expresión (c). Por tanto, la magnitud de las diferencias de utilidad se preserva frente a cambios en las unidades de medida y en el origen.

Obsérvese que cuando asociamos valoraciones numéricas a las distintas alternativas estamos de hecho generando una ordenación completa de las mismas. Es decir, la función de utilidad induce una relación de preferencias que resulta completa y transitiva. La completitud y transitividad se deducen simplemente de las propiedades de los números reales. Si un número a es mayor o igual que otro número b, y b es mayor o igual que un tercer número c, entonces a es necesariamente mayor o igual que c (transitividad). Por otra parte, dados dos números reales cualesquiera siempre podemos decir que uno es mayor o igual que otro (completitud).

El utilitarismo clásico y utilidad social

El utilitarismo clásico consideraba que las utilidades podían medirse en unidades diversas, pero también que eran *las mismas para todos los individuos*. El «origen» desde el que cada individuo medía su utilidad, sin embargo, podía ser diferente (cada individuo podía asociar un número diferente a su peor opción). Este planteamiento permitía evaluar socialmente las alternativas sumando las utilidades de todos los individuos. Con ello el sistema de valoración individual tenía una extensión inmediata a la valoración social: si un individuo escogía la opción que le proporcionaba mayor utilidad individual, una sociedad debería elegir aquella opción que le proporcionaba mayor *utilidad social*, entendida como la suma de las utilidades de los miembros del colectivo.

Es fácil comprobar que, dada una sociedad compuesta por m individuos que está valorando dos opciones alternativas $\boldsymbol{x}$ e $\boldsymbol{y}$, si la suma de utilidades para una alternativa $\boldsymbol{x}$ es mayor que para una alternativa $\boldsymbol{y}$, esta relación se mantiene si cambiamos las unidades de medida de las utilidades de la misma forma para todos los individuos, o si cambiamos el origen de las utilidades de manera diferente para cada uno de ellos. Así, tomando el caso más sencillo de dos individuos a los que identificaremos con los subíndices 1 y 2, supongamos que:

$$u_1(x) + u_2(x) > u_1(y) + u_2(y)$$

Esta expresión nos dice que la alternativa $\boldsymbol{x}$ es mejor que la alternativa $\boldsymbol{y}$, para la sociedad compuesta por estos dos individuos, porque proporciona mayor utilidad total. Si ahora cambiamos la forma de medir las utilidades variando las unidades en la misma forma para ambos agentes y los orígenes de manera diferente, tendremos lo siguiente. La valoración social de $\boldsymbol{x}$ vendrá dada ahora por:

$$(t\,u_1(x) + a_1) + (t\,u_2(x) + a_2)$$

mientras que la valoración social de $\boldsymbol{y}$ será:

$$(t\,u_1(y) + a_1) + (t\,u_2(y) + a_2)$$

Haciendo unas sencillas operaciones comprobamos que:

$$(t\,u_1(x) + a_1) + (t\,u_2(x) + a_2) = t[\,u_1(x) + u_2(x)] + t(a_1 + a_2)$$

$$(t\,u_1(y) + a_1) + (t\,u_2(y) + a_2) = t[\,u_1(y) + u_2(y)] + t(a_1 + a_2)$$

Por consiguiente, la valoración relativa de ambas opciones no depende de los números a_1, a_2, dado que entran del mismo modo en la valoración de $\boldsymbol{x}$ y en la valoración de $\boldsymbol{y}$. Y tampoco depende de las unidades t, puesto que resulta un factor común positivo en la evaluación de $\boldsymbol{x}$ e $\boldsymbol{y}$.

En consecuencia, si $u_1(x) + u_2(x) > u_1(y) + u_2(y)$ también es cierto que:

$$(t\,u_1(x) + a_1) + (t\,u_2(x) + a_2) > (t\,u_1(y) + a_1) + (t\,u_2(y) + a_2)$$

Por ello la valoración social consistente en la suma de las utilidades individuales está bien definida y es independiente de cambios comunes en las unidades y cambios diferenciados en los orígenes desde los que cada individuo mide su utilidad.

La utilidad entendida como una función objetivo genérica

La noción de utilidad cardinal tiene aplicación a algunos contextos que van más allá de la idea de satisfacción personal y pueden identificarse con una «función objetivo» genérica de un agente que puede ser un individuo, una empresa o una institución, a veces denominada ***función de pago*** (una traducción poco afortunada del término *payoff function*). La idea es que el comportamiento de este agente puede describirse como la maximización de esta función objetivo.

Consideremos primero el caso de una empresa en una economía de mercado. El comportamiento de la empresa está guiado por el principio del beneficio. Es decir, un plan de producción es considerado mejor que otro si produce más beneficios (donde los beneficios no son más que la diferencia entre ingresos y costes). Podemos decir que la función que asocia a cada alternativa (plan de producción) de la empresa el número que describe los beneficios que genera es su función de utilidad. Aquí los beneficios son cantidades monetarias y no tienen por tanto nada de valoraciones subjetivas. En este caso los cambios de unidades corresponderían a variaciones en la moneda en que se expresan los beneficios (euros, dólares, yenes, etc.). El origen natural de estas «utilidades de empresa» sería el cero, aunque podemos pensar que cada empresa puede considerar un mínimo de beneficios como el nivel desde el cual medir sus ganancias, en cuyo caso las utilidades expresarían los incrementos de beneficio por encima del mínimo considerado.

Un caso diferente sería aquel en que el problema de elección se refiere a la selección de un candidato para un determinado puesto de trabajo mediante la aplicación de un cierto baremo. Aquí la «utilidad» que el agente decisor asocia a cada alternativa será el valor del indicador que se deriva de las características del candidato considerado. Podemos imaginar que el baremo valora cinco aspectos diferentes de cada candidato y da una puntuación de 0 a 20 puntos a cada uno de estos aspectos. Así la escala del baremo es de 0 a 100. Un cambio de unidades consistiría en modificar esta escala (por ejemplo de 1 a 10, dando como máximo 2 puntos por aspecto considerado). El origen utilizado es el cero, pero podría considerarse también cualquier otro. Así por ejemplo una variante del baremo propuesto sería dar de entrada 20 puntos a cada candidato y un máximo de 16 puntos a cada uno de los cinco aspectos considerados. En este caso el baremo variaría entre 20 y 100.

La concepción moderna de la utilidad: utilidades ordinales

A principios del siglo XX Vilfredo Pareto comienza a formular el análisis del comportamiento del consumidor tomando la utilidad como «un índice» y no como una magnitud. Dicho de otro modo, la utilidad empieza a ser considerada como una variable ordinal: podemos decir que una alternativa nos da más utilidad que otra, pero no hace falta saber «cuánta utilidad más». No es que la pregunta de «cuánta más utilidad proporciona una alternativa que otra» no tenga sentido. Lo que Pareto muestra es que no hace falta para poder construir una teoría sistemática de la demanda a partir del análisis de las decisiones del consumidor basado en unas preferencias puramente ordinales. Así pues la aplicación del principio metodológico de *parsimonia* ha dado primacía a la concepción ordinal en una gran parte del análisis económico contemporáneo.

La interpretación de la utilidad como un índice propuesta por Pareto es la base de la moderna teoría de la utilidad. Según esta formulación la utilidad no será más que una forma de representar las preferencias de los individuos, sin que sus magnitudes tengan relevancia[7].

Al analizar el problema de decisión individual en términos de preferencias hemos identificado el comportamiento racional del agente con la elección en cada circunstancia de la mejor opción accesible. Es decir, el agente elige una alternativa que sea un elemento máximo de su relación de preferencias. Esta formulación del problema de decisión se puede transformar en otra equivalente, de más fácil manejo, describiendo la valoración de alternativas del agente mediante una función que representa sus preferencias (la función de utilidad) asociando números a las alternativas, de modo que un número más alto indica una opción mejor.

La siguiente definición especifica a qué nos referimos cuando hablamos de «representación» de las preferencias:

[7] Digamos, no obstante, que en determinados contextos la utilidad cardinal sigue siendo un instrumento de análisis muy útil y difícilmente sustituible (en particular en las decisiones con incertidumbre).

Definición.- *Se dice que una función* ***u*** *que asocia números a las alternativas del conjunto* ***X*** ***representa*** *las preferencias cuando para todo par de opciones* ***x****,* ***y*** *de* ***X****, se verifica:*

$$u(x) \geq u(y) \text{ si y sólo si } \quad x \succsim y$$

La función u se conoce como la ***función de utilidad ordinal****.*

Por construcción, la función $u(.)$ no es más que una forma de representar la relación de preferencias sin que sus magnitudes concretas posean significación. Más precisamente: al comparar dos opciones ***x*, *y*** del conjunto ***X*** la única información relevante se refiere a saber si $u(x)$ es mayor o menor que $u(y)$ (lo que nos dice si una alternativa es mejor que otra), sin que la magnitud de la diferencia $u(x) - u(y)$ podamos interpretarla como una medida de «cuánto mejor».

De la definición se deduce que si $u(.)$ es una forma de representar las preferencias de un agente, cualquier transformación creciente de $u(.)$ también será una representación equivalente de las mismas preferencias. Así, ahora no solo podemos cambiar el origen y las unidades en las que medimos la utilidad, sino hacer transformaciones más complejas (como elevar al cuadrado o tomar logaritmos, pongamos por caso). Las opciones escogidas como «las mejores» no cambiarán con estas transformaciones.

Observación.- *La función u(.) no nos da más información que las preferencias que representa, aunque en muchas ocasiones nos facilita la discusión de los problemas de decisión. Cuando el conjunto de alternativas no es finito la existencia de tal función no es un problema trivial, aunque puede demostrarse que existe en condiciones muy generales.*

II.3 Decisiones y preferencia revelada

Consideraremos ahora el enfoque que parte del análisis de las decisiones de los individuos y no de un criterio apriorístico de valoración de alternativas. La cuestión ahora es determinar cuándo podemos hablar de comportamiento racional de un agente a la vista de las decisiones que adopta.

Imaginemos un individuo que se enfrenta a diferentes problemas de elección donde los conjuntos sobre los que tiene que elegir son, ***X*, *T*, *S***, ... Para cada uno de estos problemas este individuo ha escogido una o más alternativas. Denominamos $C(X)$, $C(T)$, $C(S)$, ... a las *alternativas elegidas* en los problemas ***X*, *T*, *S***,... respectivamente. La función $C(.)$ que asocia los elementos elegidos a cada conjunto de alternativas disponible, se denomina ***función de elección***.

Para entender la lógica que explica el comportamiento de este individuo podemos proceder a comparar sus decisiones en problemas de elección que tengan elementos comunes y analizar si estas decisiones resultan mínimamente consistentes.

El requisito básico de consistencia en la elección individual se denomina ***consistencia en las contracciones***. Puede expresarse en los siguientes términos: Si un individuo elige una opción x^* en un conjunto X y ahora se enfrenta a un nuevo problema de elección definido por un conjunto T que es una parte de X (es decir, en la nueva situación las alternativas disponibles se han reducido), pero x^* sigue estando disponible, entonces x^* también deberá ser elegida en el problema T[8].

En principio, la formulación de la racionalidad en términos de preferencias parece más exigente que la que parte de las decisiones del individuo, porque en el primer caso pedimos que el agente ordene completamente todas las alternativas de su conjunto de elección mientras que en el segundo únicamente exigimos que en cada problema elija alguna alternativa.

Sin embargo ambas aproximaciones no son tan distintas como parece. En realidad podemos deducir unas preferencias del individuo viendo cómo elige en diferentes situaciones. Así por ejemplo, si x, y son dos alternativas del conjunto X y observamos que el individuo ha elegido x pero no y, podemos decir que sus decisiones *revelan* que x es mejor que y. Y si ha elegido tanto x como y, podemos decir que sus decisiones revelan que x e y son igualmente valoradas (indiferentes). Podemos construir así una ***preferencia revelada*** a partir de las decisiones de los individuos.

Puede comprobarse que cuando las decisiones del agente verifican el principio de consistencia en las contracciones que hemos enunciado, esta relación de preferencia revelada resulta completa y acíclica, de suerte que reproduce en buena medida la formulación de las preferencias de la que habíamos partido. Por ello, cuando se cumple la propiedad de consistencia en las contracciones, se suele decir que una función de elección $C(.)$ se puede ***racionalizar*** mediante una relación de preferencia revelada.

Por consiguiente, no hay una diferencia sustancial entre la formulación del modelo de comportamiento racional como la elección de aquellas alternativas que son mejores con respecto a una relación de preferencias que ordena todos los elementos del conjunto de alternativas, y aquella otra que se identifica con la decisión consistente de los individuos[9].

II.4 Incertidumbre

El problema de decisión racional de un agente individual consiste en determinar cuál es el mejor curso de acción en cada circunstancia. Bajo las condiciones de certidumbre consideradas hasta ahora, tomar una decisión equivale a escoger un resultado perfectamente determinado. Decir que el agente elige una alternativa x del conjunto X equivale a decir que *obtiene* efectivamente x. Hay así una relación

[8] Formalmente: Si $x^* \in C(X)$, $T \subset X$, & $x^* \in T$, entonces $x^* \in C(T)$.

[9] Hay también otro requisito frecuente que resulta complementario al anterior. Se trata de la ***consistencia en las expansiones,*** que se formula como sigue: Si x^* es una alternativa elegida en un conjunto X y x^* es también una alternativa elegida en un conjunto T, entonces x^* será también elegida en el conjunto formado por la unión de los dos anteriores. Cuando las decisiones del agente verifican los principios de consistencia en las contracciones y en las expansiones la relación de preferencia revelada resulta completa y transitiva.

biunívoca entre *acciones* (decisiones) y *consecuencias* (resultados). Cuando no existe certidumbre esta relación biunívoca entre acciones y consecuencias se pierde. La falta de certidumbre significa que *algunas de las acciones disponibles para el agente pueden generar dos o más resultados alternativos sin que sea posible asegurar a priori cuál de ellos se obtendrá*. Identificaremos la falta de certidumbre con la falta de información, en el momento de tomar la decisión, del resultado exacto que se deriva de nuestra elección.

Veamos dos casos comunes en los que el individuo se enfrenta a un problema de decisión sin certidumbre:

Ejemplo 1.- *(SALUD Y TRABAJO)* Consideremos un mundo con dos periodos de tiempo, a los que llamaremos «hoy» y «mañana» por simplicidad. Un individuo tiene que decidir *hoy* su demanda de bienes y su oferta de trabajo para *hoy y mañana*. A la hora de tomar su decisión no sabe si mañana tendrá buena salud o padecerá alguna enfermedad, lo que afecta a su capacidad de trabajo (y por tanto a su renta) de mañana. La decisión del individuo deberá ser una decisión contingente, ya que tendrá que escoger entre planes de consumo condicionales a su estado de salud mañana. Aunque sólo una de estas situaciones alternativas ocurrirá efectivamente, su decisión hoy consiste en escoger un vector de posibles resultados condicionales a lo que suceda mañana.

Ejemplo 2.- *(EL AHORRO DE LA FAMILIA)* Una familia planea invertir sus ahorros en diversos tipos de activos financieros. Supongamos que hay simplemente dos opciones: a) *Deuda pública*, que ofrece una rentabilidad fija del 3 % ; y b) *Acciones*, que ofrecen una rentabilidad media del 10 %, pero con una gran variabilidad. Para tomar una decisión esta familia debe anticipar la evolución del mercado de valores, estimando las probabilidades de los distintos rendimientos. Obsérvese que si escoge invertir todo en deuda pública el resultado de su acción está perfectamente determinado. Pero si invierte en acciones su decisión tiene un componente aleatorio. Ello ocurre también cuando la decisión es mixta, es decir, cuando esta familia invierte una parte sus fondos en renta fija y destina la otra parte a renta variable (con lo que el rendimiento esperado es menor que si lo destinara todo a renta variable, pero la varianza de los ingresos esperados también se reduce).

Estos ejemplos indican que en un problema de decisión sin certidumbre las alternativas entre las que debemos elegir son en realidad colecciones de resultados contingentes, a los que podemos asociar ciertas probabilidades derivadas de nuestra percepción del mundo. Así que mientras los únicos elementos a tener en cuenta en el problema de elección con certidumbre eran el conjunto de alternativas, las restricciones y las preferencias, ahora la situación es más compleja. Cada opción consiste en un conjunto de posibles resultados alternativos que pueden darse con ciertas probabilidades. La pregunta que surge es cómo pasar de la valoración de los resultados ciertos (consecuencias) a la valoración de las opciones consistentes en vectores de resultados probables (acciones).

Obsérvese que tomar una decisión en un contexto incierto exige la resolución de dos tipos de problemas diferentes:

(i) En primer lugar *un problema de información*: el agente debe disponer de una estimación de las probabilidades con las que ocurrirán los distintos resultados asociados a cada opción (como paso previo a determinar un curso de acción).

(ii) En segundo lugar *un problema de decisión*: el agente necesita un criterio de ordenación de estos objetos que son vectores de posibles consecuencias condicionadas por sus probabilidades (una teoría en la que la evaluación de las opciones se relacione con la distribución de probabilidad de los resultados).

En lo que sigue nos limitaremos a discutir problemas de decisión en los que las probabilidades están dadas, es decir, supondremos que el agente ha resuelto previamente el problema de la búsqueda de información para la estimación de la distribución de probabilidad que rige la aparición de los resultados.

Loterías

Hay varias formas distintas de plantear el problema de decisión sin certidumbre. Nosotros lo abordaremos aquí identificando los elementos del conjunto de elección con *loterías*, es decir, el agente tiene que elegir entre vectores de consecuencias alternativas con sus respectivas probabilidades. Cada opción se identifica así con un conjunto bien definido de posibles resultados cuyas probabilidades son «conocidas» por el agente. En el momento de tomar su decisión el agente no sabe qué consecuencia concreta resultará de su acción, pero conoce tanto los posibles resultados que pueden darse como sus respectivas probabilidades. La decisión racional consistirá en elegir aquella opción que lleve aparejados «los mejores resultados posibles».

Extendiendo la acepción que tiene en el lenguaje corriente, llamaremos *lotería* a una colección de posibles resultados con sus correspondientes probabilidades. Suele distinguirse entre loterías simples y compuestas, según que cada uno de estos resultados alternativos sean deterministas o no. Así por ejemplo, una lotería que consiste en ganar 10 euros si al lanzar una moneda al aire sale cara y nada si sale cruz, es una lotería simple. Los dos resultados posibles son deterministas. Pero si la lotería consiste en ganar 10 euros si sale cara, y volver a lanzar otra vez la moneda si sale cruz, y ganar 10 euros si sale cara o perder 10 si sale cruz, es una lotería compuesta. En este caso la lotería que se juega tiene dos resultados alternativos, uno de los cuales es determinista (ganar 10 euros si sale cara), y otro es una nueva lotería (volver a lanzar la moneda y ganar o perder 10 euros).

Para precisar estas ideas, denominamos X al conjunto de todos los posibles *resultados* (consecuencias) que pueden ser obtenidos por el agente como producto de sus decisiones en un problema de elección sin certidumbre. Por simplicidad supondremos que X es un conjunto finito, compuesto por k elementos. Como antes, los elementos de X serán designados por x, y, z, ... Ahora el agente no elige directamente entre los elementos de X, que corresponden a las «consecuencias». Las alternativas entre las que debe escoger, que denotaremos por A, B, C, etc., son ahora las *loterías*: colecciones de resultados que se pueden obtener con ciertas

probabilidades. Así, por ejemplo, si una alternativa A tiene tres posibles resultados, x, y, z con probabilidades 2/6, 1/6, 3/6, respectivamente, podemos escribir:

$$A = [(x, y, z), (2/6, 1/6, 3/6)]$$

De forma más general podemos definir la noción de lotería como sigue:

Definición.- *Una* ***lotería simple,*** *definida sobre el conjunto de posibles resultados X, es un elemento de la forma $A = (Y, \pi_Y)$, donde Y es un subconjunto de X, compuesto por un número $n \leq k$ de resultados, y π_Y es un vector que describe las probabilidades de cada uno de estos resultados, es decir, $\pi_Y = (\pi_1, \pi_2, \ldots \pi_n)$, con $\sum_{i=1}^{n} \pi_i = 1$, $\pi_i \geq 0$, para todo i.*

En el ejemplo anterior, $Y = \{x, y, z\}$, $\pi_Y = (2/6, 1/6, 3/6)$.

El conjunto de acciones sobre las que el agente tiene que elegir ahora son pues todas las loterías definidas sobre el conjunto X. Llamaremos $L(X)$ al conjunto de elección cuyos elementos A, B, C, etc. son las loterías de la forma $A = (Y, \pi_Y)$, siendo Y un subconjunto de X y π_Y la distribución de probabilidad de los elementos de Y.

Dos observaciones resultan pertinentes. La primera, que también podemos llamar *lotería* a una alternativa consistente en obtener un resultado x con seguridad (probabilidad igual a 1). En dicho caso tendremos una lotería (degenerada) de la forma: $A = [\{x\}, 1]$. Por tanto los «sucesos ciertos» también están incluidos en $L(X)$.

La segunda, que cuando una lotería conste únicamente de dos posibles resultados alternativos, x, y, podemos describirla sencillamente como $A = [\{x, y\}, \pi]$, en el entendimiento de que al elegir A se obtiene el resultado x con probabilidad π y el resultado y con probabilidad $(1 - \pi)$.

A partir del conjunto de loterías simples podemos generar nuevas loterías en la forma siguiente. Dadas dos loterías B, C y un número real π comprendido entre 0 y 1, definimos una ***lotería compuesta*** como una nueva lotería A, cuyos resultados son la lotería B con probabilidad π, y la lotería C con probabilidad $(1 - \pi)$. Es decir, $A = [\{B, C\}, \pi]$.

La idea de lotería compuesta sirve para describir aquellas situaciones en las que las consecuencias de las acciones puedan ser a su vez variables aleatorias. Un caso muy frecuente es el que ocurre con la lotería de Navidad: muchos jugadores deciden adquirir un número de esta lotería y jugar de nuevo a la lotería de «El Niño» si obtienen el reintegro del billete que jugaban. Las decisiones de reinversión de las ganancias derivadas de la adquisición de activos financieros pueden mirarse también como «loterías compuestas». Además, la composición de loterías permite establecer equivalencias entre loterías cuyos resultados son probabilísticamente indistinguibles.

Para simplificar la discusión consideraremos únicamente loterías simples y, en la mayor parte de los casos, con dos únicos resultados alternativos. Así, describiremos en general una lotería A como $A = [\{x, y\}, \pi]$ que nos dice que si elegimos esta alternativa incierta obtendremos la consecuencia x con probabilidad π y la consecuencia y con probabilidad $(1 - \pi)$.

Observación.- *En muchas ocasiones el problema de decisión sin certidumbre se plantea con algo más de estructura, tomando las probabilidades sobre los resultados de una lotería como probabilidades inducidas y no como primitivos del problema. La idea es que existe una variable aleatoria que gobierna las circunstancias, o «estados del mundo», que conducen a cada posible resultado.*

La función de utilidad esperada

En principio podemos modelizar el problema de decisión con incertidumbre en los mismos términos de antes, tomando ahora como conjunto de elección el conjunto de loterías $L(X)$. Como en la discusión desarrollada no establecíamos ninguna característica sobre los objetos que constituían el conjunto de elección, podemos aquí partir de que el agente tiene definida una relación de preferencias sobre el conjunto de loterías $L(X)$. Cuando estas preferencias son completas y transitivas su decisión racional consiste en elegir un elemento máximo de esa relación de preferencias.

Pero esta aproximación es insatisfactoria por dos razones complementarias. La primera, que no parece tan fácil admitir que un individuo pueda ordenar de forma completa y sistemática las loterías como puede hacerlo cuando las alternativas son vestidos o sombreros. La segunda, que nos gustaría tener una teoría de la decisión que nos permitiera relacionar la ordenación de las loterías (acciones) con la valoración de los resultados (consecuencias) y con su plausibilidad (probabilidades).

La forma más ampliamente utilizada para relacionar la valoración de una lotería con la valoración de los resultados y las probabilidades que la definen es la conocida como ***función de utilidad esperada***, propuesta en 1944 por J. Von Neumann y O. Morgenstern. Dada una lotería A que consta de dos resultados alternativos, x que puede obtenerse con probabilidad π, e y que puede obtenerse con probabilidad $(1 - \pi)$, tenemos:

$$U(A) = \pi\, U(x) + (1 - \pi)\, U(y)$$

donde $U(x)$, $U(y)$ son las utilidades que nos proporcionaría conseguir con certeza los resultados x e y, respectivamente[10].

Es decir, la utilidad de una lotería A no es más que el *valor esperado* de la utilidad de los resultados que puede proporcionar: la suma ponderada de las utilidades de los distintos resultados alternativos, cuyos coeficientes de ponderación son las respectivas probabilidades.

Pero no es sólo que esta forma de valorar las loterías sea razonable. En realidad von Neumann y Morgenstern demostraron que esta forma de valoración resultaba equivalente al cumplimiento de unos simples axiomas que reflejan propiedades intuitivas sobre la relación de preferencias definida sobre las loterías. Estos axiomas son los siguientes:

[10] En realidad, para ser precisos, debiéramos escribir $U[\{x\}, 1]$, $U[\{y\}, 1]$, en lugar de $U(x)$, $U(y)$, para indicar que nos referimos a los resultados x e y como loterías degeneradas con probabilidad 1.

Ordenación: *El individuo posee una relación de preferencias completa y transitiva sobre el conjunto de loterías.*

Continuidad: *Si una lotería $A = [\{x, y\}, \pi]$ es mejor que otra $B = [\{x', y'\}, \pi']$, entonces cambios «muy pequeños» en las probabilidades π, π' no alteran la ordenación de A con respecto a B.*

Independencia: *Sean x, y, z tres posibles resultados y supongamos que y es indiferente a z. Entonces, la lotería $A = [\{x, y\}, \pi]$ es indiferente a la lotería $B = [\{x, z\}, \pi]$, para cualquier probabilidad π.*

El axioma 1 nos dice que el agente es capaz de comparar consistentemente cualquier par de loterías. El axioma 2 establece que pequeños cambios en las probabilidades no modifican la ordenación entre cualquier par de loterías dado[11]. El axioma de *independencia* puede considerarse como el supuesto central en la teoría de la utilidad esperada. Establece que la valoración de una lotería no se ve alterada si cambiamos uno de sus posibles resultados por otro que sea indiferente, con las mismas probabilidades. Sucede así que la ordenación de dos loterías cualesquiera no se ve alterada por (es independiente de) la sustitución de un resultado posible por otro que sea igualmente valorado.

Estos tres axiomas caracterizan la existencia de una función de utilidad esperada, es decir:

Teorema II.1: [von Neumann & Morgenstern (1944)]

Los axiomas de Ordenación, Continuidad e Independencia se cumplen si y sólo si existe una función U definida sobre L(X) que representa las preferencias y tal que, si una lotería A proporciona el resultado x con probabilidad π y el resultado y con probabilidad $(1 - \pi)$, asocia el valor:

$$U(A) = \pi\, U(x) + (1 - \pi)\, U(y)$$

Además, U es una función de utilidad cardinal.

Este teorema asegura que podemos encontrar una forma de representar las preferencias sobre las loterías que consiste en asociar a cada opción el valor esperado (en términos de utilidad) de sus componentes.

La última parte del teorema nos da una característica distintiva de la función de utilidad esperada: se trata de una función cardinal. En efecto, sea U una función de utilidad esperada que representa las preferencias sobre las loterías. Entonces una función U' es una representación alternativa de U si y sólo si:

$$U' = tU + a \qquad (\text{con } t > 0)$$

[11] Formalmente: los conjuntos de loterías «mejores» y «peores» que una lotería dada, son *abiertos* en [0,1].

Ello significa que podemos cambiar el origen y las unidades en que medimos la utilidad (pero también que no cualquier transformación monótona creciente es admisible, como ocurría con la utilidad ordinal en condiciones de certidumbre)[12].

Cuando A es una lotería que tienen un número cualquiera k de resultados posibles, la utilidad esperada viene dada por:

$$U(A) = \sum_{i=1}^{k} \pi_i U(x_i)$$

Una nota histórica

La primera formulación del principio de la utilidad esperada aparece ya en el siglo XVIII de la mano de Daniel Bernoulli (1700-1782), quien fuera un notable médico, físico y matemático. Bernoulli introduce la idea de «valoración moral» del premio esperado de una lotería como algo diferente del valor esperado del premio. Con ello trata de resolver la conocida como «Paradoja de San Petesburgo», propuesta por su primo Nicholas Bernoulli[13].

La estimación del premio asociado a una lotería monetaria mediante el cálculo de su valor esperado era ya bien conocida en el siglo XVII. Pero Nicholas Bernoulli apunta un aspecto poco satisfactorio de esta forma de valoración mediante un sencillo ejemplo. Se trata de un juego consistente en tirar una moneda al aire de forma repetida hasta que salga «cara». El participante en el juego recibe dos euros si sale cara en la primera tirada, cuatro euros si sale cara en la segunda tirada, ocho si se necesitan tres tiradas, etc. Es decir, este juego da una ganancia de 2^n euros si sale cara en la n-ésima tirada. Como la probabilidad de que salga cara después de n lanzamientos es $\left(\frac{1}{2}\right)^n$ el valor esperado del premio que proporciona este juego, que denominamos VE, viene dado por:

$$VE(A) = \frac{1}{2}2 + \frac{1}{4}4 + \frac{1}{8}8 + +\ldots = 1 + 1 + 1 + \ldots = +\infty$$

En consecuencia, cualquier individuo debería estar dispuesto a pagar cualquier cantidad finita para tener derecho a participar en este juego. Pero en realidad, y ésta es la paradoja, ningún jugador está dispuesto a pagar grandes sumas para participar en este juego. Ni siquiera sumas apreciables.

[12] El siguiente ejemplo ilustra esta característica. Sean A, A' dos loterías dadas por: $A = [\{x, y\}, 0.5]$, $A' = [\{x', y'\}, 0.5]$. La función de utilidad esperada nos dice que $U(A) = 0.5U(x) + 0.5U(y)$, y que $U(A') = 0.5U(x') + 0.5U(y')$. Supongamos que $U(x) = U(y) = 10$, $U(x') = 19$, $U(y') = 1$. Es obvio que $U(A) = U(A') = 10$ (es decir, que las loterías A y A' son indiferentes), y que cualquier transformación lineal de U mantiene esta igualdad. Veamos qué ocurre si calculamos la utilidad esperada cuando admitimos una transformación monótona creciente arbitraria sobre U. En particular tomemos $U' = (U)^2$. Se sigue que $U'(x) = U(y) = 100$, de modo que $U'(A) = 100$; pero encontramos ahora que $U'(x') = 361$, $U'(y') = 1$, de modo que $U'(A') > U'(A)$, lo que ciertamente contradice que A y A' sean indiferentes.

[13] El nombre deriva de que Daniel Bernoulli propone la solución al problema suscitado por su primo en un escrito de la Academia de San Petesburgo. Aunque lo cierto es que una solución similar ya había sido propuesta por Gabriel Cramer en la correspondencia que mantuvo con Nicholas Bernoulli.

La solución a este enigma fue propuesta por Daniel Bernoulli mediante el argumento de que la gente no tiene en cuenta simplemente el valor esperado de las ganancias $\sum_j \pi_j x_j$, sino la «valoración moral» de las mismas. Esta «utilidad» del dinero puede medirse mediante una cierta función $U(x)$ que crece con el valor x del premio pero lo hace menos que proporcionalmente. Es decir, una cantidad de dinero $2x$ es valorada menos del doble de la cantidad x. Así, la valoración de una lotería hay que hacerla mediante el valor esperado de su utilidad: $\sum_j \pi_j U(x_j)$.

En realidad, esta diferencia entre lo que supone el *valor esperado* y la *utilidad esperada* de una lotería permiten definir un concepto importante en la teoría de la decisión sin certidumbre: la aversión al riesgo. Imaginemos que a un individuo le ofrecemos la siguiente alternativa: ganar una cantidad de dinero Y con seguridad o jugar una lotería cuyo valor esperado es precisamente Y (en el caso más sencillo podemos pensar que ofrecemos al agente que elija entre 100 € «en la mano» o tirar una moneda al aire y ganar 200 € si sale cara y nada si sale cruz). Los individuos que prefieren el dinero en mano se denominan ***aversos al riesgo***, mientras que los que prefieren jugársela se denominan ***amantes del riesgo*** (y llamamos «neutrales frente al riesgo» a aquellos que son indiferentes entre ambas alternativas).

II.5 Interacción estratégica

Hasta ahora hemos considerado que la decisión de un agente equivale a la determinación de un resultado o bien a la elección de una lotería, según hubiera o no certidumbre en el problema considerado. En ambos casos el agente se enfrenta al problema de elección de forma aislada, independientemente de las decisiones de otros agentes. Pero hay muchos problemas de decisión donde no es éste el caso. Se trata de problemas de decisión individual en los que las posibilidades de elección y los resultados que un agente puede obtener están condicionados por las decisiones de otros agentes.

Imaginemos una persona que está planeando sus vacaciones de verano. Considera varios destinos alternativos, dada su capacidad de gasto y las fechas disponibles para sus vacaciones. En principio éste podría ser un problema de elección típico de los que hemos analizado hasta ahora. El conjunto X de alternativas viene dado por aquellos programas de vacaciones que esta persona puede pagar y que resultan compatibles con las fechas que tiene disponibles. Esta persona elegirá aquel programa de vacaciones que más le satisface de entre los que se puede permitir. Podríamos incluso añadir un elemento de incertidumbre en la descripción, según la meteorología con la que se vaya a encontrar en el lugar de destino de sus vacaciones.

Pero hay un aspecto adicional que, razonablemente, esta persona debe tomar en consideración: las decisiones de otras personas que también salen de vacaciones. Eso afectará tanto a sus posibilidades de elección (plazas disponibles en los hoteles o medios de transporte) como al propio disfrute de sus vacaciones (mayores o menores aglomeraciones de turistas en el destino elegido). Por ello en su problema de decisión aparece ahora un componente nuevo que debe tomar en cuenta: las decisiones de los demás. Para tomar una decisión deberá por tanto incorporar la información de que disponga sobre las acciones decididas por los demás. Estamos en un contexto de interacción estratégica, aunque sea de naturaleza muy simple.

Los problemas de decisión individual en un contexto estratégico se analizan desde la perspectiva de la ***Teoría de Juegos***. La Teoría de Juegos es una técnica de análisis que se ocupa de modelar la interacción entre diversos agentes. Su nombre deriva de la analogía de las situaciones de interacción estratégica con los juegos de salón (ajedrez, dominó, juegos de cartas, etc.) en los que los resultados de cada jugador no sólo dependen de las jugadas que él sea capaz de realizar sino de las que realicen los otros. En estos juegos cada jugador decide cuál es su siguiente jugada en función de la situación en que se encuentra (posición en el tablero, fichas o cartas en la mano) y de lo que estime que el otro o los otros jugadores pueden hacer.

Consideremos el caso más sencillo posible, aquel en que existen únicamente dos agentes que interactúan. Llamaremos Alberto y Beatriz a estos dos agentes. Alberto elige sus acciones en el conjunto X_A que por simplicidad supondremos dado e independiente de las acciones de Beatriz. Por su parte, Beatriz elige sus acciones en el conjunto X_B que también supondremos está dado y es independiente de las acciones de Alberto. Denominaremos $x_A, y_A, z_A, \ldots$ a las alternativas disponibles para Alberto (los elementos de X_A), y $x_B, y_B, z_B, \ldots$ a las opciones a disposición de Beatriz (los elementos de X_B).

La utilidad que obtiene Alberto depende ahora tanto de su decisión como de la acción escogida por Beatriz. Así, si Alberto elige x_A y Beatriz elige x_B la utilidad de Alberto será $U_A(x_A, x_B)$, que es un expresión que pone de manifiesto la dependencia de los resultados para un agente de las acciones de los dos. Análogamente, los resultados que obtiene Beatriz, en términos de utilidad, dependen tanto de su elección x_B, como de la elección de Alberto. Es decir, $U_B(x_A, x_B)$.

En este contexto el comportamiento racional de los agentes se formula en los siguientes términos. Alberto elige, para cada opción escogida por Beatriz, la alternativa del conjunto X_A que maximiza su utilidad. Es decir, si Beatriz escoge x_B, entonces Alberto elige aquella opción x_A* tal que:

$$U_A(x^*_A, x_B) \geq U_A(x_A, x_B)$$

para todo x_A del conjunto X_A. Decimos que x_A* es la ***mejor respuesta*** de Alberto a la acción x_B de Beatriz.

Del mismo modo Beatriz elige la alternativa en X_B que maximiza su utilidad, para cada acción x_A decidida por Alberto. Es decir, si Alberto ha elegido la acción x_A, entonces Beatriz escoge la opción x_B* tal que:

$$U_B(x_A, x^*_B) \geq U_B(x_A, x_B)$$

para todo x_B del conjunto X_B. La acción x_B* es la ***mejor respuesta*** de Beatriz a la acción x_A de Alberto.

La resolución del problema de decisión en este contexto involucra a los dos agentes simultáneamente y requiere que cada uno elija una acción que sea la mejor respuesta a la acción elegida por el otro. En efecto, imaginemos que x_A*, x_B* son dos acciones elegidas por Alberto y Beatriz tales que, x_A* es la mejor respuesta por parte de Alberto a la acción x_B*; y x_B* es la mejor respuesta por parte de Beatriz a la acción x_A*. En tal caso ninguno de los dos agentes tiene incentivos a cambiar su decisión y el problema se habrá resuelto.

Así pues ahora el problema de decisión racional de un agente no consiste en tomar una única decisión, sino en planear una decisión para cada posible acción decidida por los otros. Estas acciones de un agente condicionales a las decisiones de los demás se denominan ***estrategias***. El problema de decisión en el que existe este tipo de interacción estratégica se denomina ***juego***, y se alude a los agentes que toman decisiones como ***jugadores***. Una situación en la que cada agente ha elegido una estrategia que es la mejor respuesta a las estrategias elegidas por los demás se denomina un ***equilibrio del juego***[14].

Una situación de juego (un problema de decisión con interacción estratégica) puede tener un equilibrio, varios o ninguno, dependiendo de la estructura del problema. Cuando existe un único equilibrio el conocimiento de los elementos que componen el juego nos puede permitir predecir su resultado haciendo los cálculos adecuados. Cuando existen varios equilibrios debemos enfrentarnos además al problema de determinar si hay motivos para singularizar alguno de ellos y poder así establecer cuál se alcanzará (es el conocido como *problema de selección* de equilibrios). Finalmente si no existe equilibrio en el juego no cabe esperar que se produzca ningún resultado estable derivado de la interacción de los agentes.

Un ejemplo clásico de este tipo de situación con dos únicos agentes es la de un *duopolio*, es decir, la competencia estratégica entre dos empresas que surten un mercado de un cierto producto homogéneo (por ejemplo discos compactos para grabar en el ordenador). El precio al que se vende este producto depende de la cantidad total que llevan al mercado las dos empresas. Las estrategias de cada empresa son pues los niveles de producción de este producto. Es decir, x_A representa ahora la cantidad de producto ofrecida por la empresa A y x_B la cantidad de producto ofrecido por la empresa B. Dada la demanda existente, el precio p al que se vende este producto dependerá de las decisiones de ambas empresas respecto a la cantidad puesta en el mercado, es decir, $p = f(x_A + x_B)$. Los beneficios de cada empresa, que podemos identificar con sus «utilidades» y vienen dados por la diferencia entre los ingresos y los costes, dependen de la cantidad que vende individualmente y del precio de venta (que es igual para las dos, por ser un producto homogéneo). Un equilibrio es una decisión x_A* de la empresa A y una deci-

[14] Esta es la noción de equilibrio de un juego no cooperativo debida a John Nash. Por ello también suele denominarse frecuentemente «equilibrio de Nash».

sión x_B^* de la empresa B tales que, cada una de las empresas maximiza sus beneficios, dado el nivel de producción elegido por la otra empresa[15].

En el problema de decisión individual con certidumbre el resultado de la elección de un agente era la consecución de la alternativa escogida. En una situación con incertidumbre el agente escogía una lotería (un conjunto de posibles resultados con ciertas probabilidades) y el azar determinaba qué resultado concreto se obtenía. En un contexto de interacción estratégica el resultado que obtiene un agente depende de su elección y de cuál sea la acción decidida por los demás agentes. La diferencia con el caso de la decisión sin certidumbre es que ahora lo que condiciona el resultado que se obtiene no es el azar, imprevisible por definición, sino la acción de otros agentes que sí puede ser anticipada por el agente decisor. Como en el ajedrez, cada agente debe pensar cuál será la respuesta del otro jugador a cada posible movimiento propio a la hora de elegir su juego.

Desde luego estas situaciones de juego difieren según que los agentes involucrados tengan o no toda la información sobre las preferencias y posibilidades de los demás, si pueden observar o no lo que han hecho los otros antes de elegir la propia acción, si el problema se resuelve secuencialmente o no, y de si el juego se juega una sola vez o se va a repetir. Los aspectos dinámicos introducen toda una serie de temas en el proceso de decisión de gran relevancia y complejidad: la información disponible en cada momento por un agente sobre los demás y su actualización a lo largo del juego, fenómenos de «señalización» (el uso de ciertas estrategias que revelan características de un jugador que quiere que los otros tomen en cuenta), la aparición de estrategias de amenaza o represalia, la construcción de una «reputación» por parte de los jugadores, etc.

No podemos más que dejar aquí apuntados estos temas, que nos alejan del objeto central de este libro. Comentaremos sin embargo un caso especial que tendrá su continuidad más adelante, al hablar de la implementación de las decisiones sociales.

Decisiones estratégicas... pero menos

Hay algunos contextos en los que las decisiones estratégicas de los agentes resultan especialmente sencillas. Son aquellos en que existen acciones a disposición de un agente que son «las mejores», cualesquiera que sean las decididas por los demás. Estas estrategias, cuando existen, se conocen como ***estrategias dominantes*** y tienen la ventaja de permitir al agente decidir sin tener que tomar en cuenta las decisiones de los demás. Retomando el ejemplo de Alberto y Beatriz, presentado antes, diremos que la opción x^*_A es una estrategia dominante para Alberto si

$$U_A(x^*_A, x_B) \geq U_A(x_A, x_B)$$

para todo x_A del conjunto X_A y para cualquier opción x_B que escoja Beatriz. Es decir, x_A^* es una estrategia dominante cuando es la *mejor respuesta* de Alberto a cualquier acción x_B que pueda emprender Beatriz. En este caso el problema de elección de Alberto puede resolverse como si no dependiera de las acciones de Beatriz. Esto podemos simbolizarlo como $U_A(x^*_A, .) \geq U_A(x_A, .)$ donde el punto indica que da igual cuál sea la estrategia decidida por Beatriz.

Un equilibrio en estrategias dominantes es una situación en la que cada agente dispone de una estrategia dominante. Es decir, un par (x^*_A, x^*_B) tal que $U_A(x^*_A, .) \geq U_A(x_A, .)$, para todo x_A del conjunto X_A y $U_B(.,x^*_B.) \geq U_B(.,x_B.)$, para todo x_B de X_B.

Desde luego no es un tipo de equilibrio que quepa esperar en gran parte de las situaciones estratégicas. Pero cuando existe permite una predicción muy robusta del resultado de la interacción, dado que las acciones óptimas de los agentes resultan independientes de las decisiones de los demás.

[15] Formalmente: $U_A(x_A, x_B) = p\, x_A - C_A(x_A) = f(x_A + x_B)x_A - C_A(x_A)$, donde $C_A(x_A)$ es el coste de producir la cantidad x_A. Análogamente: $U_B(x_A, x_B) = f(x_A + x_B)x_B - C_B(x_B)$. Un equilibrio es un par (x_A^*, x_B^*) tal que: $f(x^*_A + x^*_B)x^*_A - C_A(x^*_A) \geq f(x_A + x^*_B)x_A - C_A(x_A)$ para cualquier producción x_A distinta alcanzable para la empresa A. Y al tiempo, $f(x^*_A + x^*_B)x^*_B - C_B(x^*_B) \geq f(x^*_A + x_B)x_B - C_B(x_B)$ para cualquier otra producción x_B alcanzable para la empresa B.

II.6 ¿Son racionales los agentes?

Presentamos a continuación una serie de ejemplos típicos de situaciones en las que el comportamiento racional parece brillar por su ausencia. Son interesantes porque alertan sobre las limitaciones de un uso simplista de este tipo de modelización y también porque pueden ayudarnos a comprender a qué se debe que en muchos casos los comportamientos observados no parezcan explicables.

La fiesta de Gómez

El Sr. Gómez ha organizado una pequeña fiesta en su casa a la que ha invitado a unos cuantos amigos. Pasea con una bandeja de dulces ofreciéndolos a sus invitados. Cuando llega el turno a la Srta. Martínez quedan en la bandeja dos dulces, uno grande y otro pequeño. Aunque la Srta. Martínez es una amante de los dulces y no tiene motivos para hacer dieta, elige el dulce pequeño «por educación».

El Sr. Gómez vuelve a la cocina, llena de nuevo la bandeja y hace una segunda ronda repartiendo dulces. Cuando llega el turno a la Srta. Martínez quedan en la bandeja tres dulces. Uno pequeño y uno grande, como antes, y un tercero mucho mayor. Ahora la Srta. Martínez elige el dulce de tamaño intermedio (el «grande» de antes).

¿Ha actuado racionalmente la Srta. Martínez?

Uno podría argumentar que la Srta. Martínez no actúa racionalmente porque si le gustan los dulces debería elegir el más grande en cada ocasión, dado que está a su disposición. También podría ponerse en cuestión su racionalidad apelando no tanto a sus preferencias sino a la inconsistencia de sus decisiones. Si en la segunda situación elige el dulce grande cuando tenía disponible el pequeño, no parece coherente que en la situación inicial, donde ambas alternativas también estaban disponibles, haya elegido el dulce pequeño.

Pero muchos de nosotros hemos tomado en nuestra vida este tipo de decisiones, sin que nos parezca que ponen en cuestión nuestro sano juicio. Lo que este ejemplo tiene de interesante es que nos advierte frente a determinadas simplificaciones. En muchas ocasiones las preferencias de los agentes no se explican por la aplicación de un único principio («cuanto más, mejor», en el caso de los dulces), sino que incluyen otros aspectos relacionados con valores éticos, reglas sociales, convenciones, etc. La «buena educación» es una de ellas y ciertamente condiciona nuestras decisiones sin que ello suponga una contradicción con la decisión racional. Sólo parece inconsistente la elección del pastel pequeño en el primer caso si ignoramos que las preferencias de los individuos pueden incluir aspectos que van más allá de la naturaleza de las opciones que estamos escogiendo. Así que si la Srta. Martínez prefiere comer cuantos dulces sea posible pero también actuar de acuerdo con las normas de urbanidad (ya sea porque las ha asumido como propias o porque quiere emitir el mensaje de que es una persona educada), parece totalmente razonable que haya escogido el pastel pequeño.

En la segunda elección aparece además un elemento nuevo. La opción escogida puede cambiar dependiendo del contexto. Aunque parezca absurdo que si en la primera situación escogió el dulce pequeño ahora no lo haga cuando también está a su disposición y elija en cambio aquel que descartó antes, los motivos también son fácilmente comprensibles. La «buena educación» le recomienda no elegir el pastel más grande de todos. Pero ahora que hay más de dos alternativas, la Srta. Martínez elige el dulce mayor de los que no son *el más grande.*

Tampoco aquí parece que el comportamiento de la Srta. Martínez tenga nada de irracional, aunque no encaje con una aplicación simplista de los principios de racionalidad presentados. La cuestión es que la alternativa «dulce pequeño» no está definida por sí misma. «Dulce pequeño acompañado de un dulce grande» es una cosa, y «dulce pequeño acompañado de un dulce grande y otro muy grande» es una cosa distinta.

Incertidumbre y paradojas

La teoría de la utilidad esperada es un poderoso instrumento para analizar el problema de decisión sin certidumbre. Nos proporciona un método operativo de evaluar opciones inciertas, que se caracteriza a partir de unas pocas propiedades cuyo significado es fácilmente comprensible.

Pero no todos los comportamientos son explicables mediante esta teoría. Hay evidencia empírica que muestra cómo los individuos realizan elecciones inconsistentes con la teoría de la utilidad esperada. Quizás los más célebres de estos

ejemplos sean las denominadas Paradoja de Allais y Paradoja de Ellsberg, que describimos a continuación.

La Paradoja de Allais

Supongamos que damos a elegir a un grupo de individuos entre dos loterías A, B definidas como sigue: la lotería A proporciona mil euros con seguridad, mientras que la lotería B nos da cinco mil euros con probabilidad 0.1, un millón con probabilidad 0.89, y cero con probabilidad 0.01. Es decir, midiendo la cantidad de dinero en miles de euros y usando la forma de expresar las loterías desarrollada anteriormente, tendremos:

$$A = [\{\mathbf{1}\}, 1], B = [\{\mathbf{5}, \mathbf{1}, \mathbf{0}\}, (0.1, 0.89, 0.01)].$$

Es frecuente encontrar que una parte sustancial de los individuos consultados prefieren la opción A sobre la B, es decir, $U(A) > U(B)$.

Ahora ofrecemos al mismo grupo de individuos las dos alternativas siguientes: una lotería A' que proporciona mil euros con probabilidad 0.11, o cero con probabilidad 0.89; y una lotería B' que proporciona cinco mil euros con probabilidad 0.1 y cero con probabilidad 0.9. Es decir,

$$A' = [\{\mathbf{1}, \mathbf{0}\}, (0.11, 0.89)] \quad B' = [\{\mathbf{5}, \mathbf{0}\}, (0.1, 0.9)]$$

La evidencia empírica muestra que muchos de los individuos que prefirieron en el primer caso A sobre B prefieren ahora B' sobre A'.

Pero estas opciones son inconsistentes con la teoría de la utilidad esperada. Para verlo comencemos por escribir la relación entre las utilidades de las loterías A y B. Puesto que A es mejor que B tendremos:

$$U(\mathbf{1}) > 0.1U(\mathbf{5}) + 0.89U(\mathbf{1}) + 0.01U(\mathbf{0})$$

Si ahora sumamos $0.89U(\mathbf{0}) - 0.89U(\mathbf{1})$ en ambos términos de la desigualdad, obtenemos:

$$0.11U(\mathbf{1}) + 0.89U(\mathbf{0}) > 0.1U(\mathbf{5}) + 0.9U(\mathbf{0})$$

Pero $0.11U(\mathbf{1}) + 0.89U(\mathbf{0}) = U(A')$, y $0.1U(\mathbf{5}) + 0.9U(\mathbf{0}) = U(B')$. Es decir, la teoría de la utilidad esperada nos dice que si A es preferida a B entonces A' debe ser preferida a B', contra lo que se observa empíricamente.

La paradoja de Ellsberg

Esta paradoja se refiere a la inconsistencia que suele presentarse como resultado de una primaria aversión a la antigüedad.

La situación con la que nos enfrentamos en este caso es la siguiente. Se trata de un juego de dos personas en el que gana 100 euros aquel que acierte el color de una bola extraída al azar de una urna. Las bolas pueden ser rojas o negras y hay dos urnas de las que pueden ser extraídas, con 100 bolas cada una. El primer jugador escoge el color (rojo o negro). El segundo jugador ya no puede escoger qué color prefiere (le toca el que no ha escogido el primer jugador), pero en cambio puede elegir la urna de la que se extrae la bola (urna A o urna B). En la primera urna hay 50 bolas rojas y 50 bolas negras. En la segunda urna hay 100 bolas rojas y negras, pero no sabemos en qué proporciones.

Supongamos ahora que el jugador 1 ha escogido el color negro, de modo que el jugador 2 apuesta sobre la bola roja. En general se observa empíricamente que los jugadores tipo 2 eligen la urna A. Pero también se observa que el jugador 2 sigue prefiriendo la urna A cuando el jugador 1 eligió el color rojo. Y esto resulta inconsistente con la utilidad esperada. Dado que el premio es el mismo, si eligió en el primer caso la urna A es porque atribuye mayor probabilidad a obtener una bola roja en esa urna que en la urna B; pero entonces debería atribuir menor probabilidad a obtener una bola negra en esa misma urna.

Lo que estas paradojas ponen de relieve es que los individuos son menos precisos a la hora de calcular las probabilidades de las loterías compuestas de lo que establece la teoría.

El caso de la paradoja de Ellsberg parece ilustrar un fenómeno de «terror a lo desconocido», de modo que nos resulta preferible elegir una urna en la que sabemos que tenemos el 50 de bolas de cada clase, a otra en la que ignoramos por completo su distribución (¡podría ocurrir que las 100 bolas que contiene la segunda urna fueran del color sobre el que apostamos!).

En la paradoja de Allais los agentes parecen sufrir una «ilusión óptica» que hace que no perciban con precisión la relación entre la variación del premio y la variación de la probabilidad. En la primera alternativa el agente parece percibir que los premios que ofrecen las loterías A y B son muy parecidos en cuanto a su valor esperado, con la diferencia de que $A = [\{1\}, 1]$ nos da un millón de pesetas con seguridad. Es como si la utilidad experimentara una discontinuidad al pasar de «lo seguro» a «lo casi seguro». En la segunda alternativa nos encontramos con dos loterías en las que todos los resultados son probabilísiticos. En este caso la decisión parece guiada por la magnitud del mayor premio que ofrece, dado que las probabilidades son «parecidas». Si calculamos el valor esperado de los premios en los dos casos vemos que el valor esperado del premio en la lotería B es $0.5 + 0.89 = 1.39$, que es superior al de la lotería A (que obviamente es la unidad). Aquí nos comportamos de forma aversa al riesgo, ya que escogemos una alternativa con menor valor esperado, pero cuyo resultado obtenemos con seguridad. En el segundo caso tenemos un valor esperado de la lotería A' igual a 0.11, y un valor esperado de la lotería B' igual a 0.5. En este caso estamos escogiendo la lotería con mayor valor esperado y con menor probabilidad asociada al premio mayor.

Sobre presos y helados

La Teoría de Juegos es un modelo matemático que trata de predecir el resultado de la interacción entre individuos, cada uno de los cuales actúa buscando su propio interés. El contexto en el que se aplica es similar al de los juegos de mesa, como el ajedrez o muchos de los juegos de cartas, donde el resultado que obtiene cada jugador depende tanto de la estrategia que siga el jugador como de las que sigan los demás jugadores. Como en los juegos de mesa el aspecto interesante es que la acción mejor de cada jugador depende de cómo piense que va a comportarse su adversario. La teoría de juegos no pretende mostrar cómo ganar al ajedrez o hacer rentable la afición al póker, sino que trata de explicar algunos comportamientos económicos, como los de las grandes empresas que compiten entre sí. Una de las cosas que esta teoría pone de manifiesto es que en muchas ocasiones la interacción de individuos o instituciones que se comportan guiados por sus intereses individuales generan «malos resultados», ya sea para los propios jugadores, ya sea para la sociedad en su conjunto. Por «malos resultados» queremos decir que actuando cooperativamente todos podrían haber terminado mejor que haciéndolo de modo independiente.

El dilema del prisionero

La situación en la que los agentes involucrados en un juego obtienen un resultado peor para todos ellos de lo que hubiera sido posible mediante su cooperación se conoce, en la jerga de los teóricos de juegos, como «el dilema del prisionero». Su nombre deriva del ejemplo convencional que se refiere al caso de dos detenidos por la policía, Alberto y Beatriz, que son interrogados separadamente. La policía puede probar un delito menor a los imputados, que les llevaría a la cárcel por unos meses, pero saben –aunque no pueden probarlo– que son los autores de un importante robo con violencia, que les reportaría de 10 a 15 años de cárcel. La policía ofrece a cada detenido, al que no se permite comunicar con el otro, un pacto del siguiente tipo: «Si tú confiesas que vosotros dos sois los autores del robo con violencia, cargaremos las culpas sobre tu compañero y tú te podrás beneficiar como colaborador de la justicia, con una sentencia reducida que podrás cumplir con libertad condicional. Pero si confiesa primero tu colega, entonces las cosas serán al revés. Piénsalo y mañana a primera hora nos dices cuál es tu decisión».

La solución mejor para los delincuentes conjuntamente es, desde luego, la de «no confesar» porque así no les pueden imputar el delito mayor. Sin embargo la teoría de juegos nos dice que la solución que debemos esperar es que al día siguiente ambos confiesen simultáneamente de modo que nadie podrá aprovechar los beneficios planteados en la propuesta de confesión. ¿Por qué? Pensemos en cómo razonará aquella noche el delincuente Alberto. «Si yo no confieso y Beatriz sí lo hace, entonces me cargarán a mí con todo el peso de la ley por el robo que hicimos juntos, que es lo peor que me puede ocurrir. De modo que en este caso me conviene confesar a mí antes para así evitar la cárcel. Pero si Beatriz no confiesa, entonces a mí también me interesa confesar porque de ese modo no cumpliré ninguna pena de prisión». Y Beatriz, obviamente, llega inmediatamente a la misma conclusión. Con el resultado de que ambos terminan con penas de 10 a 15 años.

¡Al rico helado!

El Ayuntamiento de Alicante ha decidido conceder dos licencias municipales para la instalación de dos puestos de helados en la playa de la ciudad, que tiene un kilómetro de larga. Los puestos de helados se sitúan al borde del agua y deben ofrecer los mismos servicios (el Ayuntamiento no quiere que los ciudadanos sean discriminados en los servicios que reciben porque vayan a una u otra zona de la playa).

Los ganadores de las licencias municipales son la empresa A (El Helado de Alicante) y la empresa B (El Buen Helado). Una vez obtenidas sus licencias las empresas deben decidir dónde situar sus quioscos. Lo ideal, desde el punto de vista de los consumidores (los bañistas de la playa), es que un puesto de venta se situara a 250 mts. del comienzo de la playa y el otro a 250 mts. del final de la playa. ¿Por qué? Pues porque así cualquier bañista no tendrá que andar más de 250 mts. para ir a comprar un helado. En efecto, todos los bañistas que están en los primeros 500 mts. de la playa irán al puesto que se sitúa al principio de la playa y nadie tiene que recorrer más de 250 mts. porque este puesto está justo en la mitad de esta primera parte de la playa. Y lo mismo ocurre para los bañistas que están en los últimos 500 mts. de la playa, que acudirán al puesto de venta que está a 250 mts. del final.

Pero no debemos esperar que sea éste el resultado de la acción independiente de las empresas. La empresa A, que se ubica en la primera parte de la playa, sabe que todos los bañistas que se sitúen entre el inicio de la playa y su puesto de helados «son suyos», porque su quiosco está más cerca que el otro. Y además captará aquellos bañistas que aunque estén del otro lado les quede más cerca su puesto. Por tanto tendrá incentivos a desplazarse desde el punto ideal de 250 mts. del inicio de la playa, hacia el centro de la misma, para ganar más clientes.

Pero la empresa B que cubre la parte final de la playa realiza el mismo tipo de razonamiento. ¿Con qué resultado? Pues, si nadie lo remedia, los dos puestos de helados estarán pegados uno a otro en la mitad exacta de la playa. Con el consiguiente perjuicio para los bañistas que se sitúan en los extremos de la misma.

El inefable Sam Spade

Sam Spade, de Spade & Archer, Detectives, antes de que Archer fuera asesinado, es el único que conoce dónde está el tesoro (en forma de pájaro negro). Pero ha caído en una encerrona organizada por El Gordo Gutman y sus secuaces. Trata de negociar la entrega del tesoro a cambio de cierta cantidad de dinero y de dejar en manos de la policía a alguno de los esbirros de Gutman como culpable de los crímenes cometidos. Gutman y los suyos no están por labor y le recuerdan a Spade que está en sus manos y que pueden matarle. El bueno de Sam discute con sus captores en los siguientes términos:

– ... Si me matan ustedes, ¿cómo van a saber dónde está el pájaro? Y si yo sé que no pueden permitirse el lujo de matarme hasta tener el pájaro en su poder, ¿me quieren decir cómo creen que me pueden atemorizar para que ceda ante sus pretensiones?

Gutman ladeó la cabeza y consideró estas preguntas. Brillaron sus ojos por entre los párpados fruncidos. Y, al cabo, dio la solución en su acostumbrado tono amistoso.

– Verá usted, caballero, existen otros métodos de persuasión, además de matar o amenazar de muerte.

– Es verdad –asintió Spade–, pero a no ser que estén respaldados por una amenaza de muerte, no sirven de gran cosa para convencer a la víctima. ¿Comprende? Si ustedes tratan de hacer algo que no me plazca, me negaré a aceptarlo. Y les presentaré el dilema de renunciar a ello o matarme, sabiendo que no pueden matarme.

– Comprendo su punto de vista –dijo Gutman riendo entre dientes-. Es una actitud que exige ser juzgada con muy minucioso cuidado por las dos partes pues, como bien sabe usted, señor mío, ocurre que los hombres, a veces, en medio del ardor de la acción, olvidan lo que real y verdaderamente les conviene y permiten que las emociones los arrastren.

Spade también se deshizo en amables y dulces sonrisas:

– En eso consiste el quid de mi estrategia, en actuar con suficiente firmeza para estorbarles a ustedes la libertad de movimientos pero sin llegar a enfurecerlos de tal manera que me manden al otro barrio muy en contra de lo que les conviene.

– ¡Caramba, señor mío! –dijo Gutman con admiración–: ¡Es usted todo un carácter!

Dashiell Hammet, *El Halcón Maltés* (1929) (traducción de F. Calleja).

El principio de racionalidad

El modelo de racionalidad individual parte del supuesto primitivo de que los individuos están dotados de voluntad, discernimiento y capacidad de elección (los requisitos que se asocian a la *responsabilidad* de los actos). La capacidad de elección significa simplemente que el individuo tiene a su disposición más de un curso de acción posible. El discernimiento se asocia a la idea de que es capaz de comparar las diversas alternativas que se le ofrecen. Y la voluntad equivale a la capacidad de tomar una decisión, es decir, elegir con un propósito. La racionalidad puede formularse entonces como el principio por el que el individuo elige aquella alternativa disponible que más valora o que se adapta mejor a sus propósitos.

El modelo de agente racional presupone que las acciones de un individuo son *intencionales* y reflejan su valoración de las diferentes alternativas y de los condicionamientos con los que se enfrenta (posibilidades de elección, información disponible, condiciones del entorno). Esta formulación permite anticipar el comportamiento esperable de los individuos a partir del conocimiento de su valoración de las alternativas y de las restricciones con que se enfrenta. Y también reconstruir el tipo de valoración del individuo a partir de la observación de sus acciones y del conocimiento de sus condicionamientos.

Por supuesto los atributos de voluntad, discernimiento y capacidad de elección admiten toda una gradación. En los modelos más estilizados se supone que se cumplen «en grado sumo». Así un individuo tiene siempre un conjunto de opciones entre las que elegir, conoce perfectamente la naturaleza de las alternativas con las que se enfrenta y es capaz de valorarlas de forma completa, sistemática y sin costes, por más numerosas y complejas que sean éstas. Por último, se supone que el individuo afronta el problema de elección con la finalidad de alcanzar algún resultado.

Aunque éste es un buen punto de partida para comprender los comportamientos individuales, no debemos perder de vista que en la vida real difícilmente se cumplirán en su totalidad los postulados característicos de estos modelos estilizados. En particular es frecuente que nos encontremos con situaciones en las que el individuo no conoce con todo detalle la naturaleza de las alternativas de que dispone, o bien no es capaz de valorarlas sistemáticamente. Los agentes pueden cometer errores en la recogida y el procesamiento de la información, pueden resolver problemas de decisión sin la suficiente concentración o esfuerzo, pueden estar cansados, irritados, o verse influidos por los comportamientos de otros, o resultar sensibles a la existencia de normas sociales implícitas, a la forma de presentación del problema (lo que se conoce como *«framing effects»*), o de criterios valorativos múltiples. Todos estos elementos darán lugar a la aparición de comportamientos erráticos o que contravienen las predicciones de la teoría.

¿Se deduce de la observación de estos comportamientos contrarios a las predicciones que el enfoque de la racionalidad es incorrecto? No necesariamente. Puede muy bien suceder que hayamos elegido un modelo de racionalidad inadecuado para explicar el problema de decisión concreto. No es que la teoría que predica el comportamiento racional del agente «sea mala», sino de que los postulados sobre los que se asienta (conocimiento completo, capacidad de valoración y cálculo ilimitados, valoración de las alternativas independiente del contexto, existencia de un único criterio valorativo para cada agente, etc.) no se corresponden con la situación observada. En estos casos se requiere una concepción de la racionalidad menos extrema (v.g. racionalidad limitada) o bien más sofisticada (v.g. decisión multicriterio), que no obstante sigue formando parte del paradigma que vincula las acciones de los individuos al *comportamiento racional*.

Los ejemplos anteriores ilustran diversos aspectos de esta falta de racionalidad aparente. En el ejemplo de los dulces la incoherencia surge cuando tratamos de analizar el comportamiento en términos de un único criterio valorativo. Lo que el ejemplo pone de manifiesto es que a veces los individuos actúan con motivaciones múltiples lo que da lugar a aparentes intransitividades o falta de coherencia en los comportamientos observados. Es la «confederación de almas» de la que habla Tabucchi en la cita que hemos seleccionado para abrir este texto.

Las paradojas de Allais y Ellsberg ponen de manifiesto otro tipo de dificultad práctica. A veces los agentes no consiguen diferenciar con absoluta precisión la naturaleza de las alternativas, sufren «ilusiones ópticas» o se ven influenciados por el contexto del problema. También podemos encontrar aquí que los criterios de valoración no son unidimensionales (por ejemplo la valoración de una alternativa incierta puede depender del punto de referencia desde el que se la compara, o el agente puede tomar en cuenta cómo se sentirá *después* según el resultado que obtenga, y no sólo cuál es la alternativa con mayor utilidad esperada)[16].

El caso del dilema del prisionero plantea una problemática diferente. Aquí las acciones de los individuos son perfectamente racionales, pero el resultado de su interacción es el peor posible desde el punto de vista de los individuos involucrados. Lo que este conocido ejemplo pone de manifiesto es que no hay ninguna garantía de que la acción racional de un conjunto de individuos que interaccionan dé lugar a resultados deseables para los propios participantes en «el juego».

Por último, en el caso de los puestos de venta de helados en la playa también vemos que las acciones racionales de las empresas llevan a un resultado indeseable. En este caso no indeseable para ellas (las «jugadoras»), sino para los bañistas, que es la sociedad afectada por los servicios proporcionados por estas empresas. Aquí parece claro que una intervención pública que forzara la ubicación de los puestos de venta conseguiría una solución socialmente mejor, sin perjudicar a las empresas.

En resumen: el principio de racionalidad proporciona una buena aproximación al estudio del comportamiento de los individuos, siempre que lo interpretemos con la suficiente flexibilidad. Distintas situaciones pueden reclamar diferentes vías de modelizar el «comportamiento racional». El modelo convencional de un agente decisor perfectamente racional, capaz de valorar de forma instantánea y sin coste todas las alternativas, calcular cuál es la mejor de ellas y tomar decisiones precisas, no pasa de ser un esquema de referencia que nos sirve para comprender los comportamientos en un mundo estilizado. Entender cómo varían los comportamientos esperables cuando consideramos situaciones que se separan de ese modelo canónico es también una fuente de conocimiento que nos ayuda a comprender los problemas reales.

[16] Existen diversas teorías alternativas a la utilidad esperada que son capaces de explicar gran parte de las incongruencias observadas. Entre ellas destacan diversas generalizaciones de la utilidad esperada, como la *prospect theory* de Khaneman y Twersky (1979), la *regret theory* de Loomes y Sudgen (1982), o la *skwen symmetric bilinear theory* de Fishburn (1988).

Apéndice al capítulo II

A.II.1: Relaciones binarias

Una ***relación binaria*** definida sobre un conjunto A es una forma de asociar entre sí los elementos de A. Formalmente, una relación binaria sobre A es un subconjunto G del producto cartesiano $A \times A$, es decir, un conjunto de pares $(\boldsymbol{a}, \boldsymbol{b})$ donde tanto $\boldsymbol{a}$ como $\boldsymbol{b}$ son elementos de A que «están relacionados». En muchas ocasiones esta relación se expresa a través de algún criterio R, de modo que dos elementos están relacionados si satisfacen dicho criterio. En este caso escribimos $\boldsymbol{a}\ R\ \boldsymbol{b}$ para indicar que «$\boldsymbol{a}$ está relacionado con $\boldsymbol{b}$».

El siguiente ejemplo ilustra estas ideas. Sea $A = \{1,2,3,4,5,6\}$ y consideremos la relación R definida como sigue: un elemento $\boldsymbol{a}$ del conjunto A está relacionado con otro $\boldsymbol{b}$ también en A si y sólo si $\boldsymbol{a}$ es el doble de $\boldsymbol{b}$. En este caso el conjunto G viene dado por:

$$G = \{(2, 1), (4, 2), (6, 3)\}$$

Consideremos ahora las siguientes propiedades que puede cumplir una relación binaria R:

Reflexividad: $\boldsymbol{a}\ R\ \boldsymbol{a}$, para todo $\boldsymbol{a}$ de A.

Transitividad: Para todo $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ de A, si se cumple que $\boldsymbol{a}\ R\ \boldsymbol{b}$ y que $\boldsymbol{b}\ R\ \boldsymbol{c}$, entonces también se cumple que $\boldsymbol{a}\ R\ \boldsymbol{c}$.

Simetría: Para todo $\boldsymbol{a}, \boldsymbol{b}$ de A, $\boldsymbol{a}\ R\ \boldsymbol{b}$ implica $\boldsymbol{b}\ R\ \boldsymbol{a}$.

Antisimetría: Para todo $\boldsymbol{a}, \boldsymbol{b}$ de A, si se cumple que $\boldsymbol{a}\ R\ \boldsymbol{b}$ y que $\boldsymbol{b}\ R\ \boldsymbol{a}$, entonces $\boldsymbol{a} = \boldsymbol{b}$.

Asimetría: Para todo $\boldsymbol{a}, \boldsymbol{b}$ de A, si se cumple que $\boldsymbol{a}\ R\ \boldsymbol{b}$ entonces no puede darse que $\boldsymbol{b}\ R\ \boldsymbol{a}$.

Completitud: Para todo $\boldsymbol{a}, \boldsymbol{b}$ de A, siempre se cumple que $\boldsymbol{a}\ R\ \boldsymbol{b}$ o bien que $\boldsymbol{b}\ R\ \boldsymbol{a}$ (o ambos).

Una relación binaria R que verifica las propiedades de reflexividad, simetría y transitividad se denomina ***relación de equivalencia***. Suele designarse mediante el símbolo $\approx$. Un ejemplo de relación de equivalencia es la definida por el criterio «tener la misma edad que», sobre un conjunto dado de personas.

Sea A un conjunto dado y sea $\approx$ una relación de equivalencia definida sobre A. Dado un elemento cualquiera $\boldsymbol{a}$ de A, se denomina ***clase de equivalencia*** de $\boldsymbol{a}$ al conjunto $C(\boldsymbol{a})$ de los elementos $\boldsymbol{b}$ de A tales que $\boldsymbol{b} \approx \boldsymbol{a}$. Desde luego cualquier elemento de $C(\boldsymbol{a})$ constituye un representante de la clase $C(\boldsymbol{a})$.

La familia $C(\boldsymbol{a})$, para cada $\boldsymbol{a}$ perteneciente a A, de las clases de equivalencia generadas por una relación de equivalencia $\approx$ determinan una ***partición*** del conjunto A. Es decir, la relación $\approx$ genera una familia $C(\boldsymbol{a})$, con $\boldsymbol{a}$ perteneciente a A, de subconjuntos de A tales que:

a) $C(\boldsymbol{a})$ es distinto del vacío para todo $\boldsymbol{a}$;
b) La unión de todos los $C(\boldsymbol{a})$ es precisamente el conjunto A;
c) Dos conjuntos $C(\boldsymbol{a})$, $C(\boldsymbol{b})$ distintos no pueden tener ningún elemento en común.

La familia de clases de equivalencia de A definidas por la relación $\approx$ se denomina *conjunto cociente*, y se designa por $A/\approx$.

Una relación binaria definida sobre un conjunto A que verifica las propiedades reflexiva y transitiva se llama ***preorden***. Cuando la relación es además completa, entonces se dice que es un ***preorden completo***[17].

La relación «costar igual o menos que», definida sobre un conjunto de mercancías es un preorden[18].

Llamamos ***orden estricto*** a una relación que verifica las propiedades transitiva y asimétrica. La relación «tener más edad que» en el conjunto de ciudadanos de un país es un preorden estricto.

Sea A un conjunto preordenado por una relación R. Se llama ***elemento maximal*** de A a un punto $\boldsymbol{a}$ de A que verifica la siguiente propiedad: Para todo $\boldsymbol{b}$ de A, si $\boldsymbol{b}\ R\ \boldsymbol{a}$ entonces $\boldsymbol{b} \approx \boldsymbol{a}$. Es decir, $\boldsymbol{a}$ es un elemento maximal cuando no hay ningún elemento en A que esté relacionado con él, excepto los que son equivalentes.

Análogamente podemos definir el ***elemento minimal***, como aquel punto $\boldsymbol{c}$ de A tal que: Para todo $\boldsymbol{b}$ de A, si $\boldsymbol{c}\ R\ \boldsymbol{b}$ entonces $\boldsymbol{b} \approx \boldsymbol{c}$. Es decir, $\boldsymbol{c}$ es un elemento minimal cuando no está relacionado con ningún elemento de A, excepto los que son equivalentes.

Sea R un preorden definido sobre un conjunto A. Se dice que $\boldsymbol{a} \in A$ es un ***elemento máximo*** si $\boldsymbol{a}\ R\ \boldsymbol{b}$, para todo $\boldsymbol{b}$ de A. Análogamente, se dice que $\boldsymbol{c} \in A$ es un ***elemento mínimo*** si $\boldsymbol{b}\ R\ \boldsymbol{c}$ para todo $\boldsymbol{b}$ de A.

Puede comprobarse que si la relación R es un preorden completo, entonces todo maximal es un máximo (y todo minimal es un mínimo).

A.II.2: Más sobre preferencias

Como mencionamos en la sección II.2, hay tres propiedades habituales con respecto a las preferencias que tienen sentido cuando las alternativas sobre las que tenemos que elegir pueden describirse mediante vectores de números reales. El ejemplo inmediato es el relativo a escoger una cesta de mercancías compuesta de una serie de bienes perfectamente divisibles.

Para apoyar la discusión recurriremos a ilustrar los conceptos presentados tomando como referencia un ejemplo concreto, extremadamente sencillo.

---

[17] Adviértase que la forma de definir completitud implica la reflexividad, de modo que un preorden completo es una relación completa y transitiva.

[18] Dado un preorden definido sobre un conjunto A, podemos establecer de forma natural una relación de equivalencia $\approx$ sobre dicho conjunto como sigue: $\boldsymbol{a} \approx \boldsymbol{b}$ y sólo si se cumple que $\boldsymbol{a}\ R\ \boldsymbol{b}$ y que $\boldsymbol{b}\ R\ \boldsymbol{a}$. Es el caso de la ***relación de indiferencia*** que analizamos cuando R era una relación de preferencias.

Ejemplo: ***La decisión sobre el tiempo de ocio.*** Se trata de decidir cómo distribuimos nuestro tiempo libre de la próxima semana entre dos posibles actividades: la lectura y las salidas con los amigos. En este caso cada una de las alternativas entre las que tiene que elegir el agente consta únicamente de dos componentes, cuyas magnitudes podemos representarlas por medio de números reales. Es decir, una opción vendrá dada por un vector de la forma $x = (x_1, x_2)$, donde la primera componente representa el tiempo dedicado a la lectura y la segunda el tiempo dedicado a salir con los amigos. Supondremos además, para hacer más concreto el ejemplo, que tenemos 40 horas disponibles para ocio la próxima semana. Entonces, el conjunto de elección X viene dado por todos los vectores de la forma $x = (x_1, x_2)$ tales que $x_1 + x_2 \leq 40$, donde tanto x_1 como x_2 son números no negativos.

La primera de estas propiedades relevantes que queremos presentar, relativa al caso en que las alternativas vienen descritas por vectores de números reales, es la ***continuidad***. La continuidad de las preferencias establece que si una opción x es mejor que otra opción y, cambios «suficientemente pequeños» en x darán lugar a otra alternativa que siga siendo mejor que y. Intuitivamente la continuidad nos dice que si una alternativa x es preferida a otra y, las alternativas que sean «casi idénticas» a x, también serán mejores que y.

En el ejemplo de la decisión sobre el tiempo libre la continuidad de las preferencias significa que si una opción x es considerada mejor que otra opción y, siempre podemos modificar ligeramente los tiempos dedicados a la lectura y a las salidas con los amigos en x para obtener una alternativa diferente x', que seguirá siendo mejor que y[19].

La continuidad de las preferencias tiene muchas implicaciones analíticas relevantes que aquí no abordaremos. Nos limitaremos a señalar dos consecuencias particulares. La primera, que cuando las preferencias son continuas la función de utilidad que las representa es una función continua. La segunda, que si una relación de preferencias es transitiva y continua entonces es también completa [Schmeidler (1971)].

Conviene señalar que hay relaciones de preferencia definidas mediante vectores de números reales que son completas y transitivas pero que no son continuas. El ejemplo característico es el *orden lexicográfico*, que es un principio de valoración que ordena las alternativas consistentes en vectores de números reales en modo similar a cómo el diccionario ordena las palabras. En particular, el orden lexicográfico da prioridad a aquellas alternativas con un mayor valor en una componente predeterminada (por ejemplo la primera).

Tomando el ejemplo de la distribución del tiempo de ocio, un orden lexicográfico que prioriza la lectura sería aquel que ordena las alternativas de acuerdo con los siguientes principios. Dadas dos alternativas cualesquiera, damos siempre preferencia a aquella que tiene un mayor valor en la primera componente (sean

[19] Por supuesto aquí habría que precisar qué significa "cambios suficientemente pequeños" o una alternativa "casi idéntica" a otra. Cuando las alternativas están compuestas por vectores de números reales estos conceptos pueden definirse con precisión en términos de *entornos* de un punto. La continuidad de las preferencias equivale a suponer que los conjuntos de alternativas mejores y peores a una dada son abiertos en X.

cuales sean los valores de la segunda componente). Si dos alternativas tienen la primera componente igual, entonces ponemos delante a aquella que tiene un mayor valor de la segunda componente.

Puede comprobarse que, con este principio de ordenación, dos alternativas casi idénticas pueden estar muy lejos en la valoración del individuo.[20] Y también que no es posible representar estas preferencias mediante una función de utilidad (porque, dicho informalmente, «no tenemos bastantes números»).

La segunda propiedad relevante en este contexto es la ***convexidad***. Esta propiedad establece que si una alternativa x es mejor que otra alternativa y, entonces las alternativas intermedias entre x e y son también mejores que y. La idea de convexidad introduce la noción de «gusto por la variedad» que tiene sentido cuando las opciones que comparamos son planes de consumo. También refleja la idea de que un estímulo repetido reduce la intensidad de la respuesta, lo que se traduce en que estímulos variados (combinaciones intermedias de alternativas) tienden a resultar más valorados.

En el ejemplo de la distribución del tiempo de ocio la convexidad nos diría que siempre resultan mejores las opciones que reparten el tiempo libre entre la lectura y los amigos, antes que las alternativas que se centran en una u otra actividad.

En un contexto de decisión sin certidumbre la convexidad de las preferencias incorpora la idea de ***aversión al riesgo*** y tiene como resultado la elección de alternativas en las que el riesgo no se concentra en una consecuencia particular sino que se distribuye entre varias. Así, en las decisiones de inversión esta propiedad tiene como consecuencia que los agentes *diversifiquen los riesgos* escogiendo carteras mixtas compuestas por varios tipos de activos.

La convexidad de las preferencias también tiene muchas implicaciones relevantes que tienen que ver con las buenas propiedades que se derivan para la función de utilidad con respecto a los problemas de optimización mediante los que podemos modelizar el comportamiento racional del agente.

La última propiedad que suele aparecer en este contexto es la de ***monotonía***, que dice que cuando comparamos dos alternativas y una de ellas tiene «más de todo», entonces resulta preferida. Así que la monotonía traduce simplemente el principio de «cuanto más, mejor», que tiene un sentido muy claro en determinados contextos. En el ejemplo de la distribución del tiempo de ocio nos diría que siempre estaríamos dispuestos a disfrutar de más de 40 horas de ocio la próxima semana.

La consecuencia más relevante de esta propiedad es que garantiza que el agente siempre elegirá alternativas que agoten todas sus posibilidades (el agente de nuestro ejemplo gastará efectivamente las 40 horas de ocio que tiene disponibles).

[20] Lo mismo ocurre en el diccionario con dos palabras que difieren en una sola letra. Si esta letra es la primera pueden estar muy lejos una de otra, a pesar de que sean palabra "casi iguales". Pensemos por ejemplo en las palabras "bendición" y "rendición", que sólo difieren en una letra pero pueden estar en tomos distintos de nuestro diccionario.

CAPÍTULO

III

«Lo que no puede ser no puede ser, y además es imposible»

POPULAR

III. Racionalidad colectiva (1)

III.1 Introducción

El modelo de elección racional de un agente individual formula el problema de decisión a partir de la especificación de dos elementos básicos: el *conjunto de elección* (que interpretamos incluyendo las posibles restricciones que determinan el conjunto de opciones efectivamente disponibles), y sus *preferencias* (la ordenación de alternativas que el individuo realiza). La ***decisión racional*** del agente consiste en la elección de un elemento «mejor» para cada conjunto de oportunidades. Nos ocuparemos aquí del problema de modelar la decisión colectiva. El tema central es el estudio de mecanismos de decisión cuyo sujeto es el colectivo y no el individuo. Como en el caso del problema de decisión individual hay un genérico conjunto de elección, al que se suele aludir como conjunto de *alternativas sociales.*

Pero hay dos problemas sustancialmente nuevos que aparecen al analizar las decisiones colectivas:

(i) La determinación del óptimo social.

(ii) La implementación de las decisiones sociales.

El primer problema con que nos enfrentamos al hablar de decisión colectiva se refiere a cómo modelar la ***racionalidad social***, para poder así determinar las decisiones óptimas para la sociedad. La determinación del óptimo social se refiere al estudio de diversos procedimientos que permiten seleccionar cuáles son las mejores alternativas sociales, para una amplia familia de problemas de decisión colectiva.

El estudio de la ***implementación de las decisiones sociales*** analiza las dificultades para la puesta en práctica de las decisiones sociales que se derivan del comportamiento estratégico de los agentes, cuando parte de la información que define el óptimo social es privada. El tema relevante aquí es el diseño de mecanismos de asignación que hagan compatibles los incentivos individuales con los objetivos sociales. La forma de abordar esta problemática, a la que nos referiremos más adelante, es en términos de la teoría de juegos no cooperativos (con la idea de que el óptimo social se alcance como un equilibrio de Nash de un juego no-cooperativo convenientemente diseñado).

Nuestra discusión se centrará en el primer aspecto, es decir, en la forma de determinar lo que resulta socialmente «mejor». La forma más general de abordar este problema es en términos de la denominada ***teoría de la elección social.*** Esta teoría toma como punto de partida las preferencias individuales y trata de construir, a partir de ellas, una relación de preferencia social (o función de bienestar social), que permita formular el problema de decisión colectiva en modo análogo al problema de decisión individual. Se trata de construir una regla de valoración colectiva a partir de la agregación de las preferencias individuales en preferencias sociales. Tiene sus orígenes en los estudios relacionados con el bienestar personal (asociados a nombres como Bentham o Mill), por una parte, y a la teoría de las votaciones (Condorcet, Borda), por otra. Su formulación moderna arranca del planteamiento del problema efectuado por Bergson (1938), en términos de una función de bienestar social. Es sobre todo el trabajo de Arrow (1951) el que marca la evolución de este campo, a partir del sorprendente resultado de que al exigir unas pocas condiciones al proceso de valoración colectiva, todas ellas razonables, no resultaba posible encontrar ningún procedimiento capaz de satisfacerlas. Estudiaremos el «teorema de imposibilidad» de Arrow y discutiremos su alcance y significado. Analizaremos también los resultados de posibilidad que se obtienen cuando permitimos comparaciones interpersonales de utilidad. Veremos que sólo en condiciones especiales es posible obtener una agregación satisfactoria de preferencias individuales en preferencias sociales.

Los ***juegos cooperativos*** constituyen una técnica de análisis complementaria que es aplicable a una familia particular de problemas de decisión colectiva. Estos problemas se refieren a la distribución de las ganancias que surgen de la cooperación entre un grupo de agentes con intereses parcialmente contrapuestos. Una aproximación diferente la proporciona la ***axiomatización de reglas de asignación.*** Como en el caso de los Juegos Cooperativos, también aquí se parte de una concepción apriorística del problema mediante la definición de ciertos principios de deseabilidad social, basados en criterios éticos y operativos. Se trata de establecer cuáles son las condiciones ideales que una decisión social debe satisfacer para familias de problemas concretas.

Dedicamos este capítulo y el próximo al tema de la agregación de preferencias y los dos siguientes a las extensiones relacionadas con los juegos cooperativos y la axiomatización de reglas de decisión.

III.2 Las tribulaciones del Presidente

Comencemos por discutir un sencillo problema de decisión social que servirá de referencia para entender la naturaleza de muchas de las dificultades que encontraremos.

El Presidente del Gobierno de un país quiere tomar alguna medida de choque para reducir el déficit público. Para ello convoca a sus tres Vicepresidentes y les plantea la siguiente cuestión. «Quiero tomar una medida drástica que ejemplifique el deseo del Gobierno de reducir el déficit público. He considerado las tres alternativas siguientes:

(I) Reducir sustancialmente los altos cargos de la Administración y amortizar una proporción de las plazas de los funcionarios que se jubilen.

(II) Reducir el presupuesto para obras públicas y privatizar las empresas públicas que sean rentables.

(III) Introducir un control estricto del déficit de las Administraciones locales y regionales.

Quiero que cada uno de vosotros me diga cuál de estas medidas es la mejor opción, cuál la opción intermedia, y cuál la peor. Cuando tenga vuestras opiniones tomaré una decisión».

La idea del Presidente es elegir aquella alternativa que consiga un apoyo mayoritario entre sus Vicepresidentes.

Una vez estudiadas las diferentes propuestas por cada uno de los Vicepresidentes, éstos remiten un informe al Presidente del Gobierno en el que le comunican sus conclusiones. Éstas se resumen en el siguiente cuadro:

Cuadro III.1

Opciones	Vicepresidente 1°	Vicepresidente 2°	Vicepresidente 3°
Mejor opción	I	II	III
Opción intermedia	II	III	I
Peor opción	III	I	II

El Presidente comienza a hacer las cuentas en su despacho. Observa que la mayoría (Vicepresidentes 1° y 3°) prefiere la opción I a la opción II; y también que la mayoría (Vicepresidentes 1° y 2°) prefiere la opción II a la opción III. En este punto el Presidente sonríe y piensa «Ya está todo claro: I es mejor que II y II es mejor que III, de modo que I también será mejor que III y por tanto adoptaremos la alternativa I». Pero al momento su decisión se tambalea: ¡Hay también una mayoría (Vicepresidentes 2° y 3°) que prefiere la alternativa III a la alternativa I!

Esta es una primera ilustración de las dificultades que aparecen cuando agregamos preferencias individuales en preferencias sociales. Aquí la *sociedad* está compuesta simplemente por tres individuos (los tres Vicepresidentes), cada uno de los cuales tiene perfectamente definida su ordenación del conjunto de alternativas. El criterio de agregación de preferencias también está perfectamente definido: una alternativa es declarada socialmente mejor que otra si es preferida por la mayoría. Y sin embargo las preferencias sociales que se derivan del criterio mayoritario *no consiguen ordenar las alternativas*, puesto que no se cumple la propiedad de transitividad de la preferencia social. Esta falta de transitividad tiene como consecuencia la incapacidad de llegar a una decisión sobre qué alternativa elegir, puesto que se genera un *ciclo* en la valoración social de las alternativas: I mejor que II, II mejor que III y III mejor que I.

Para salir de este atolladero el Presidente decide consultar a las bases y pide que manifiesten su opinión sobre cuál de las tres ordenaciones de alternativas prefieren, la propuesta por el Vicepresidente 1°, la propuesta por el Vicepresidente 2° o la propuesta por el Vicepresidente 3°. La idea del Presidente es escoger la opción preferida por la mayoría. Los datos de la encuesta entre los militantes se recogen en la siguiente tabla, que es idéntica a la anterior excepto que hemos incluido en la cabecera de cada columna el porcentaje de militantes que apoya la valoración de cada Vicepresidente:

Cuadro III.2

Opciones	33 % (Vice. 1°)	32 % (Vice. 2°)	35 % (Vice. 3°)
Mejor opción	I	II	III
Opción intermedia	II	III	I
Peor opción	III	I	II

La conclusión parece obvia: la mayoría de los militantes valora las alternativas como el Vicepresidente 3°, de modo que la opción III debe ser la elegida. Pero el Jefe del Gabinete del Presidente le hace notar que en realidad un 65 % de los militantes prefiere la opción II a la opción III, de modo que no parece que eligiendo III se represente bien la voluntad mayoritaria del partido. Y la cosa no acaba ahí, un 68 % de los militantes prefiere la opción I a la opción II y un 67 % la opción III a la opción I.

El recurso a la votación entre los militantes de las distintas ordenaciones propuestas por los Vicepresidentes no es una solución al problema.

Al Presidente, recordando con añoranza los concursos de Eurovisión de años atrás, se le ocurre una vía alternativa para llegar a tomar una decisión. Cada Vicepresidente contará con 10 puntos que debe repartir entre las diferentes alternativas. Luego se suman los puntos obtenidos y se adopta la alternativa con mayor puntuación.

El Presidente se encuentra con la distribución de estos puntos recogida en la siguiente tabla. En cada una de las columnas del cuerpo de la tabla se recoge la puntuación de cada Vicepresidente a las opciones I, II y III planteadas por el Jefe de Gobierno. En la última columna se suman los puntos concedidos por cada Vicepresidente a cada una de las opciones.

Cuadro III.3

Alternativas	Vice. 1°	Vice. 2°	Vice. 3°	Total puntos
I	6	2	0	8
II	3	5	0	8
III	1	3	10	14

Ya esquilmado por las dificultades encontradas con anterioridad nuestro Presidente no acepta tan fácilmente la idea obvia de que la opción III debe ser la elegida. Algo no le huele bien en esta distribución de puntos. De nuevo su Jefe de Gabinete le apunta el problema que subyace: mientras que los Vicepresidentes 1° y 2° seguramente han declarado honestamente la valoración que conceden a las distintas alternativas, el Vicepresidente 3°, más pillo y experimentado, ha dado todos los puntos a su alternativa preferida para hacerla ganar, dando 0 a las otras dos. Claramente si los otros dos Vicepresidentes hubieran estado más vivos habrían hecho lo mismo, lo que hubiera resultado en un empate con 10 puntos cada alternativa.

El Presidente algo ha aprendido en este laberinto de tomar decisiones a partir de las opiniones de sus Vicepresidentes. Por una parte que pueden manipular sus preferencias de modo que consigan el mayor apoyo posible para su opción favorita a costa de desvalorizar artificialmente las demás. Por otra parte, que tomar decisiones con tres Vicepresidentes cuando hay más de dos alternativas puede dar lugar a irresolubles ciclos de mayorías. Dado que no parece fácil contrastar la honradez de las valoraciones en un concurso tipo Eurovisión, inspirador del anterior procedimiento, el Presidente hace un último intento de llegar a una decisión comparando dos alternativas cada vez y eliminando aquella que no resulte preferida. Toma ahora inspiración de los campeonatos internacionales de fútbol, donde cada dos equipos se enfrentan y el ganador pasa a la siguiente fase, hasta que solo queda un vencedor.

Así que nuestro Presidente retoma el cuadro inicial con las preferencias que le manifestaron sus Vicepresidentes y comienza el torneo entre las diversas alternativas.

Cuadro III.4

Opciones	Vicepresidente 1°	Vicepresidente 2°	Vicepresidente 3°
Mejor opción	I	II	III
Opción intermedia	II	III	I
Peor opción	III	I	II

Entre las alternativas I y II gana claramente la alternativa I, que es considerada mejor que la II por los Vicepresidentes 1° y 3° (digamos que el resultado del partido sería 2 a 1 a favor del equipo I). De modo que la alternativa II queda eliminada. Ahora juega la alternativa I frente a la III. Aquí gana la alternativa III, que es preferida por los Vicepresidentes 2° y 3° (gana también por 2 a 1), por lo que la alternativa I queda eliminada y la alternativa III resulta la opción vencedora ¡Por fin! exclama el Presidente, pronto a darle la buena nueva al Vicepresidente 3° que siempre había visto la III como la mejor opción.

Pero de nuevo el Jefe de Gabinete, arriesgando su puesto por cortar la alegría de su Presidente, le pregunta si está seguro de que ese es un resultado indiscutible. ¿Y por qué no? responde el Presidente sin disimular su disgusto. «Porque tal vez si en estos torneos empezamos con otra pareja de competidores, el resultado puede variar». Ante la parálisis que este comentario provoca en el Presidente, el Jefe del Gabinete saca una hoja de papel con los cálculos que él ha hecho. Si empezamos haciendo competir a las alternativas II y III, vemos que II gana a III, por lo que III queda eliminada. Y ahora entre I y II gana la alternativa I.

O sea, que la solución de estos «torneos» depende del orden con el que se ejecutan. Esto quizás rememore en algunos las votaciones en las asambleas estudiantiles, cuando había que elegir entre más de dos alternativas y se recurría a votar las alternativas por pares en forma de torneos. Rápidamente se observaba que el orden en que se votaban las alternativas podía afectar al resultado, lo que solía provocar interminables discusiones sobre «la forma adecuada» de hacerlo.

III.3 La Teoría de la Elección Social: El enfoque de Arrow y el Teorema de Imposibilidad

De manera general podemos decir que el objeto de la teoría de la elección social es el estudio de «las relaciones entre los objetivos de política social y las preferencias y aspiraciones de los miembros de la sociedad» [Sen (1970, p.1)]. Detrás de esta aparentemente sencilla definición existen tres aspectos esenciales que conviene subrayar. El primero, que los problemas que aborda la teoría de la elección social son de naturaleza eminentemente *normativa*. El segundo, que esta teoría se interesa por la obtención de reglas de evaluación que reflejen las preferencias de los individuos (a diferencia de otros criterios que podría tomar como referencia determinadas tradiciones, códigos preexistentes o «libros sagrados»). El tercero, que su ámbito de aplicación es enormemente general, abarcando problemas de muy diversa naturaleza (desde cómo elegir un alcalde a cómo repartir una herencia o cómo financiar una obra pública).

Dicho en otros términos, la teoría de la elección social trata de *encontrar criterios normativos de agregación de preferencias individuales en preferencias sociales que sean de aplicación a una gran variedad de contextos.*

Los elementos que tomaremos como referencia para la discusión serán pues los siguientes. En primer lugar, una ***sociedad***, término que designa simplemente una colección de agentes individuales que se enfrentan a un problema de decisión que les afecta colectivamente. En segundo lugar, las ***alternativas sociales***, que constituyen el conjunto de elección de nuestro problema. Estas alternativas varían de un problema a otro y pueden ser de naturaleza muy diversa. Si se trata de elegir un representante político el conjunto de alternativas será el elenco de candidatos; si se trata de decidir sobre proyectos de infraestructuras y formas de financiarlos el conjunto de elección estará dado por una colección de planes de obras públicas con sus costes respectivos y las posibles formas de cubrirlos mediante impuestos o tasas; y así sucesivamente. Por último, el ***criterio de valoración social***, es decir, la forma de valorar las alternativas sociales a partir de las preferencias individuales. Nos referiremos de manera genérica a este criterio como ***regla de elección social***.

Dada una sociedad y un conjunto de alternativas sociales, diseñar una regla de elección social requiere especificar tres aspectos diferentes, que pueden depender del tipo de sociedad de que se trate y de la naturaleza de las alternativas consideradas. Estos aspectos son:

(1) Qué tipo de variables resultan admisibles como argumentos de la regla de elección. O, dicho en otros términos, qué información usaremos como *input* de nuestra regla.

(2) Qué propiedades de coherencia posee la valoración social de alternativas. En otras palabras, cuál queremos que sea el tipo de *output* de la regla de elección. En particular si la regla debe ordenar completamente las alternativas sociales, o simplemente ser capaz de decir cuál es la alternativa mejor en cada ocasión.

(3) Las *propiedades de comportamiento*, es decir, cómo varía la valoración social con las valoraciones individuales.

Para que una regla de elección social tenga algún interés deberá satisfacer ciertos requisitos que puedan considerarse razonables, tanto desde un punto de vista operativo como desde un punto de vista ético. Estos requisitos son una expresión de los principios que gobiernan el proceso de agregación de preferencias. Nos indicarán en qué contextos resultará aplicable, cómo responderá la valoración social a cambios en las preferencias individuales, qué tipo de racionalidad social pedimos, etc. Idealmente buscamos un conjunto de propiedades *mínimas* que puedan ser universalmente aceptadas y nos permitan obtener una teoría significativa.

A partir de estos presupuestos el programa de trabajo de la teoría de la elección social podría enunciarse como sigue: Dada una sociedad y un conjunto de alternativas sociales, estudiar qué procedimientos de valoración social se derivan de agregar las preferencias de los individuos de acuerdo con ciertos requisitos. Siguiendo este esquema se obtendrían distintas reglas de elección social dependiendo del conjunto de principios éticos empleados en la agregación de las preferencias individuales.

Las perspectivas de este programa de trabajo se vieron drásticamente sacudidas por el notable resultado obtenido por Arrow en 1951, conocido como el *Teorema de Imposibilidad*. Este resultado indica que imponiendo unas condiciones relativamente débiles al proceso de agregación de preferencias individuales en preferencias sociales, la única regla posible resulta ser dictatorial (es decir, la regla de valoración de uno de los individuos de la sociedad). O dicho de otro modo: La conclusión central de Arrow es que no existe ninguna posible regla de agregación de preferencias que sea a la vez racional, eficiente, general y democrática. Esta contribución ha marcado buena parte del desarrollo de la teoría de la elección social.

El ejemplo que hemos presentado en la sección III.2 indica que el método de decisión mayoritaria, una de las reglas más sencillas y atractivas para la agregación de preferencias individuales en preferencias sociales, incumple un requisito básico: No es capaz de ordenar adecuadamente las alternativas. Ello implica que esta forma de valoración social puede generar ciclos de preferencia que nos impidan, como al Presidente del ejemplo, tomar una decisión. Este resultado, conocido de antiguo como la paradoja de las votaciones, no es en realidad más que la punta del iceberg. Kenneth Arrow demostró que el problema no afecta únicamente a la regla de mayoría simple. En realidad no hay *ninguna* regla capaz de satisfacer las buenas propiedades que presenta el método de votación mayoritaria y que evite la presencia de ciclos.

Arrow formula el problema de elección colectiva como una extensión de este método de decisión. Dada una sociedad y un problema de decisión colectiva, propone la construcción de una regla de elección social de acuerdo a los siguientes principios:

- Los argumentos de la regla de elección social deben ser simplemente las *ordenaciones individuales* de las alternativas[1]. En particular esto excluye la consideración de intensidades de preferencia, o de comparaciones interpersonales de utilidad.
- La regla de elección debe proporcionar una *ordenación* de las alternativas sociales como reflejo de las preferencias individuales. De esta forma estamos pidiendo que las preferencias sociales se comporten de forma análoga a las preferencias individuales. En particular, esta propiedad garantiza la ausencia de ciclos de preferencia.

Sean $M = \{1,2,...,m\}$ el conjunto de índices que identifica los agentes de una sociedad y X el conjunto de alternativas sociales. Una **Función de Bienestar Social** (FBS) es una regla de valoración social que determina una ordenación (transitiva y completa) de las alternativas como función de las ordenaciones de los individuos que constituyen la sociedad.

En cuanto a las propiedades de comportamiento que esta función de bienestar social debe cumplir, Arrow establece el siguiente conjunto de criterios mínimos:

- ***Dominio Universal***: *La regla de valoración social admite cualquier tipo de preferencias individuales.*

Esta propiedad nos dice que queremos que la regla de elección sea aplicable a cualquier tipo de sociedad (es decir, a cualquier configuración de las preferencias individuales). Una consecuencia relevante de la condición de Dominio Universal es que cambios en las preferencias individuales pueden alterar la valoración social de las alternativas, pero no cambian la regla de decisión colectiva. En este sentido la regla de valoración social constituye una especie de «Constitución», es decir un mecanismo de decisión colectiva que no cambia con los cambios de opinión de los individuos.

[1] Por «ordenación de las alternativas» nos referimos a la existencia de una relación de preferencia *completa* y *transitiva* (lo que formalmente corresponde a un «preorden», en el lenguaje de las relaciones binarias).

- ***Respeto de la Unanimidad***[2]: *Si todos los individuos de la sociedad consideran que una alternativa **x** es mejor que otra alternativa **y**, entonces la valoración social debe ordenar las alternativas del mismo modo. Es decir, **x** debe resultar socialmente preferida a **y**.*

Esta condición excluye la posibilidad de ordenar y antes que x si unánimemente todos los agentes estiman que x es mejor que y. Si no hay unanimidad esta condición no nos dice nada sobre cómo debe operar la regla de elección.

- ***Eficiencia Informacional***: *La valoración social de las alternativas **x** e **y** debe depender únicamente de cómo valoran los individuos **x** en relación a **y**.*

Hablamos de eficiencia informacional porque esta propiedad indica que la regla de elección social es capaz de operar con una cantidad mínima de información: Para comparar las alternativas x e y no necesitamos saber cómo valoran los individuos x o y con relación a una tercera alternativa z[3].

Consideremos ahora la siguiente definición, que introduce la idea del «dictador».

- ***Definición:*** *Se dice que una función de bienestar social es **dictatorial** si existe algún individuo j en M tal que, para cualquier configuración de preferencias y para todo par de alternativas sociales {**x, y**} de X, **x** es socialmente preferido a **y** si y sólo si el individuo j valora más **x** que **y**. Al individuo j se le denomina en tal caso un **dictador**.*

Una regla de elección social es dictatorial cuando responde a los deseos de un único individuo de la sociedad. Hay un aspecto de la definición que conviene no perder de vista: la valoración social de las alternativas cambia cuando cambian las preferencias del dictador (el individuo j): si hoy el dictador estima que x es mejor que y entonces la opción x es declarada socialmente mejor que la opción y. Pero si mañana el dictador decide que y es mejor que x, entonces la regla de elección social dirá que y es socialmente mejor que x.

Arrow obtiene el siguiente resultado (conocido como Teorema de Imposibilidad): No existe ninguna forma de obtener una ordenación social a partir de las ordenaciones individuales, que resulte de aplicabilidad universal, respete la unanimidad, no sea dictatorial y sea informacionalmente eficiente. Como señala Sen (1970, p.38), cada una de las condiciones propuestas resulta bastante inocua, pero «conjuntamente parecen producir un monstruo capaz de devorar todas las funciones de bienestar social del mundo».

Formalmente:

TEOREMA III.1: (Teorema de Imposibilidad de Arrow):

Sea un conjunto de elección con más de dos alternativas y una sociedad con más de dos individuos. Si una Función de Bienestar Social verifica las propieda-

[2] También llamado «Principio de Pareto».

[3] Esta propiedad se denomina en ocasiones «Independencia de Alternativas Irrelevantes». Hemos evitado el uso de este nombre con el fin de evitar la confusión con la propiedad del mismo nombre utilizada por Nash en el contexto de la teoría axiomática de la negociación.

des de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional, entonces es dictatorial.

Lo que el Teorema de Imposibilidad de Arrow establece es que no resulta posible construir un orden de preferencia social que resulte razonable, a partir únicamente de las preferencias individuales.

III.4 Las vías de escape al Teorema de Imposibilidad

Dada la naturaleza del resultado de imposibilidad de Arrow, no es sorprendente que haya generado una abundante literatura que trata de encontrar vías de escape al mismo modificando algunas de las condiciones propuestas. Repasaremos muy brevemente las posibilidades que aparecen al cambiar parte de las propiedades exigidas por Arrow, excepción hecha del Respeto a la Unanimidad que nos parece un requisito difícilmente renunciable.

La construcción de un orden social con propiedades de racionalidad más débiles

Una de las primeras ideas desarrolladas para escapar al resultado de imposibilidad fue la relajación del requisito de Ordenación (en particular debilitando el requisito de transitividad de la regla de elección). Se trata de sustituir la propiedad de transitividad por alguna otra más débil, compatible con la noción de racionalidad social. Nos referimos a las propiedades de *casi-transitividad* y de *aciclicidad.*

La ***casi-transitividad*** es una condición más débil que la transitividad ya que sólo pide que sea transitiva la relación de preferencia estricta (la relación «ser mejor que») pero no necesariamente la indiferencia (la relación «ser tan bueno como»)[4]. La ***aciclicidad*** es una propiedad todavía más débil que establece la ausencia de ciclos en la preferencia estricta. Lo que establece esta condición es que si existe una colección de k alternativas ($x_1, x_2, ..., x_k$) en X tales que x_1 sea mejor que x_2, x_2 sea mejor que x_3, ..., x_{k-1} sea mejor que x_k, no puede ocurrir que x_k lo consideremos mejor que x_1[5].

Amartya Sen observó que si cambiábamos el requisito de transitividad por el de casi-transitividad, existían reglas de elección social que verificaban las propiedades de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional. Sen propuso el siguiente ejemplo (una regla conocida como «Extensión del Principio de Pareto»). Dadas dos alternativas cualesquiera x e y,

[4] La idea es que largas cadenas de alternativas, cada una de las cuales resulta indiferente a la anterior, pueden generar intransitividades cuando existen pequeñas diferencias, imperceptibles una a una, que se acumulan conforme la cadena de comparaciones aumenta. Ya ilustramos este problema en el ejemplo del café con una variación infinitesimal en la cantidad de azúcar.

[5] Puede demostrarse que cuando la relación de preferencias es completa la aciclicidad es una condición necesaria y suficiente para que exista una elección, sobre cualquier subconjunto finito de alternativas en X.

declaramos que x es socialmente preferido a y si y sólo si todo el mundo prefiere x a y; caso contrario x e y son declaradas socialmente indiferentes. Es fácil comprobar que esta forma de establecer valoraciones sociales no es dictatorial y sin embargo verifica los requisitos establecidos. Esta regla de valoración no es muy atractiva porque todos los individuos tienen *poder de veto*; es decir, si alguien opina que x es mejor que y, no puede ocurrir que y resulte socialmente preferido a x, cualquiera que sea la opinión del resto de los miembros de la sociedad. Su interés deriva de que sirve para ilustrar que podemos escapar del Teorema de Imposibilidad.

Pero este camino no lleva muy lejos. Andreu Mas-Colell y Hugo Sonnenschein demostraron que la única regla de elección social casi-transitiva que cumple las propiedades de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional es una regla oligárquica. Una ***oligarquía*** se define como un grupo de personas que conjuntamente actúan como un dictador, es decir, x es considerado socialmente mejor que y si y sólo si todos los miembros de la oligarquía consideran que x es mejor que y. Cada miembro de la oligarquía tiene por tanto poder de veto. Esto nos indica que, en cierto sentido, la regla de Extensión del Principio de Pareto es la menos mala de las posibles, ya que la única oligarquía que admite es la compuesta por todos los miembros de la sociedad.

En realidad las cosas son todavía peores. Si, además de las propiedades de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional, pedimos a la regla de elección social que verifique un elemental principio adicional, conocido como «Monotonía», entonces el resultado de imposibilidad reaparece. Esta propiedad se define como sigue:

Definición: *Una relación de preferencia social es* ***monótona*** *si, cada vez que una opción x es declarada socialmente mejor o igual que una opción y, esta valoración no cambia si se modifican las preferencias de los miembros de la sociedad de modo que todos los individuos que consideraban x mejor o igual que y siguen pensando lo mismo.*

La condición de monotonía nos dice cómo debe responder la valoración social ante un cambio de preferencias individuales, en un caso particular. Establece que si en la sociedad una alternativa x es socialmente mejor que otra alternativa y, y algunos individuos que antes valoraban más y que x cambian de opinión, mientras que todos los que valoraban x más que y la mantienen, entonces la nueva valoración social debe determinar que x sigue siendo mejor que y (porque en la nueva situación habrá un número mayor de individuos que consideran x mejor que y).

Puede probarse el siguiente resultado:

Teorema III.2: [Mas-Colell y Sonnenschein (1972)]

Sea un conjunto de elección con más de dos alternativas y una sociedad con más de dos individuos. La única relación de preferencia social casi-transitiva que verifica las propiedades de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Monotonía, es dictatorial.

Este resultado establece que, si aceptamos como razonable el requisito de Monotonía, debemos concluir que no hemos conseguido nada al sustituir la transitividad del orden social por la casi-transitividad.

Sin embargo, todavía se puede relajar más el supuesto de transitividad conservando ciertas propiedades de racionalidad y operatividad. En lugar de la condición de casi-transitividad, podemos considerar la condición de *aciclicidad* que equivale a exigir que no haya ningún ciclo en la preferencia estricta.

Puede probarse que cuando la regla de elección es acíclica y verifica las propiedades de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Monotonía, existen siempre individuos con poder de veto sobre todas las alternativas (es decir, existen individuos tales que si prefieren x a y excluyen la posibilidad de que socialmente y sea declarado mejor que x).

A la vista de estos resultados es fácil convenir que no hemos adelantado mucho mediante la relajación del requerimiento de transitividad de la regla de elección social. No es ésta, pues, una vía satisfactoria de escapar al resultado de Arrow.

La existencia de un criterio razonable de elección social sin necesidad de recurrir a la ordenación de todas las alternativas

Otro camino alternativo es la renuncia a la ordenación de todas las alternativas. Ahora lo único que pedimos es que la regla de elección social sea capaz de seleccionar siempre alguna opción como la mejor, en cualquier subconjunto de las alternativas sociales. Formalmente, sea S un subconjunto cualquiera de X, y denominemos $C(S)$ al conjunto de elementos de S que resultan elegidos. Una ***función de elección social*** es una correspondencia que asocia a cada subconjunto S de X uno o más elementos $C(S)$ contenidos en S. La idea implícita es que los elementos de $C(S)$ son «los mejores».

Con esta aproximación al problema de elección social, que no exige que el criterio de elección genere una ordenación completa de todas las alternativas, adoptamos un enfoque más general. No obstante, para que las funciones de elección tengan algún interés deberán satisfacer algunos requisitos mínimos de consistencia (no vale todo, por así decir, si queremos elaborar una teoría significativa). Los dos requisitos más habituales de consistencia sobre funciones de elección son los siguientes:

Propiedad α.- *Para todo par de subconjuntos S, T de X, tales que S está contenido en T, si $x \in S$ es un elemento de $C(T)$, entonces también es un elemento de $C(S)$.*

Esta propiedad puede considerarse como un requisito de consistencia ante contracciones del conjunto de oportunidades: si x es un elemento elegido en un cierto conjunto y este conjunto se reduce, pero la alternativa x sigue siendo una opción alcanzable, entonces x también deberá ser un elemento elegido del nuevo conjunto de oportunidades. Para ilustrar lo que dice esta propiedad supongamos que X es el conjunto de los todos los jugadores profesionales de fútbol del mundo y sean S y T los conjuntos de todos los jugadores profesionales de fútbol de España y Europa, respectivamente. Si consideramos que Ronaldo está entre los mejores jugadores de Europa [$\{\text{Ronaldo}\} \in C(T)$], entonces, como Ronaldo juega en España [es decir, $\{\text{Ronaldo}\} \in S$] deberíamos concluir que Ronaldo está entre los mejores jugadores de España [es decir, $\{\text{Ronaldo}\} \in C(S)$].

Propiedad γ.- *Sean S, T dos subconjuntos de X, y supongamos que x está en $C(S)$ y, además, x está en $C(T)$. Entonces, $x \in C(S \cup T)$.*

Esta propiedad postula la consistencia frente a expansiones del conjunto de oportunidades que contienen alternativas dominadas. Si x es elegido en el conjunto S y también es elegido en un conjunto T, entonces x seguirá siendo elegido en el conjunto formado por la unión de S y T. El siguiente ejemplo ilustra el significado de esta condición. Si Fernando Alonso es mejor piloto de Fórmula 1 que todos los demás pilotos españoles, y también es mejor piloto que todos los pilotos portugueses de esa categoría, entonces Fernando Alonso es mejor que todos los pilotos de Fórmula 1 de la Península Ibérica.

Puede demostrase el siguiente resultado:

TEOREMA **III.3: [Blair, Bordes, Kelley y Suzumura (1976)]**

Una función de decisión social C(.) que cumpla los principios de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Monotonía, cumple las propiedades α y γ si y sólo si es dictatorial. Si únicamente cumple la propiedad α entonces existe algún individuo con poder de veto.

Tampoco parece que este camino constituya una vía de escape satisfactoria.

El Método de Decisión Mayoritaria y la ordenación social con tipos específicos de preferencias individuales

Comentaremos ahora algunos resultados obtenidos cuando sustituimos la propiedad de Dominio Universal por otra menos general. Se trata de analizar si hay alguna forma razonable de restringir las preferencias admisibles de los individuos que no sea demasiado restrictiva y al tiempo nos permita escapar al teorema de Arrow. Buena parte de los resultados de este enfoque están vinculados al análisis de las propiedades del Método de Decisión Mayoritaria. Ello se debe en gran medida a que, como veremos enseguida, las posibles formas de agregación de preferencias que se ajustan al enfoque de Arrow son esencialmente procedimientos de votación mayoritaria. Concentremos nuestra discusión en torno a este procedimiento.

La regla de decisión mayoritaria puede precisarse como sigue: La alternativa **x** se declara preferida a la alternativa y si más del 50 por 100 de los individuos consideran que x es mejor que y. Indudablemente esta regla de elección social resulta atractiva por su carácter eminentemente democrático. Puede comprobarse fácilmente que el Método de Decisión Mayoritaria satisface las condiciones de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional e incluso

Monotonía. El problema que plantea, como hemos visto en el ejemplo del Presidente atribulado, es que esta regla de elección genera valoraciones sociales intransitivas.

El propio Arrow sugirió un procedimiento de escapar a su resultado de imposibilidad con el método de decisión mayoritaria mediante una adecuada restricción de las preferencias individuales. Se trata de las ***preferencias unimodales***. Decimos que las preferencias de un conjunto de *m* individuos, definidas sobre un conjunto de alternativas sociales ***X***, son *unimodales* (*«single-peaked»*) si el conjunto de alternativas sociales puede ser ordenado sobre la recta real de forma que la función de utilidad de cada individuo sea monótona (creciente o decreciente), o bien creciente hasta un cierto punto y decreciente en adelante.

Las preferencias unimodales están asociadas a la existencia de una opción preferida por cada individuo de modo que cuanto más nos alejemos de esta mejor opción, menos valoradas son las alternativas. Este tipo de preferencias aparecen de forma natural en algunos contextos, como en el caso de las elecciones políticas. Los individuos suelen tener una opción política preferida y cuanto más lejos de esta opción están los candidatos, ya sea por la derecha o por la izquierda, peor valorados resultan. También encontramos este tipo de preferencias cuando se trata de instalar un servicio público (un colegio o un centro de salud, por ejemplo) y hay que decidir su ubicación. Los individuos pueden ordenar las diferentes ubicaciones según estén más o menos alejadas del lugar donde viven.

Para precisar más la idea de preferencias unimodales, consideremos el siguiente ejemplo. Una sociedad debe elegir el número de kilómetros de tren de alta velocidad que desea poner en funcionamiento, teniendo en cuenta que cada kilómetro tiene un coste promedio de q euros. Aquí las alternativas sociales aparecen descritas de modo natural mediante números reales, que representan los kilómetros de alta velocidad a poner en servicio[6]. Podemos pues disponer las alternativas sociales de forma que x_1 precede a x_2, que a su vez precede a x_3, ... y así sucesivamente. En este contexto las preferencias unimodales describen el siguiente tipo de valoración: Cada individuo tiene una alternativa $p(i)$ que es la mejor de todas, a la que llamamos su «pico» y que corresponde a un cierto número x de kilómetros de alta velocidad que el individuo considera como la cantidad óptima, dado el coste por kilómetro y el consiguiente impacto que supondrá en términos de impuestos. Para cualquier par de alternativas x_s, x_t se cumple que: Si $x_t \leq x_s \leq p(i)$ o bien $p(i) \geq x_s \geq x_t$, entonces x_t es peor que x_s (que desde luego es peor que $p(i)$). Es decir, conforme nos alejamos por un lado o por el otro de la opción preferida $p(i)$, encontramos opciones progresivamente peores.

Por supuesto cada individuo tendrá una opción preferida que puede ser diferente de las de los demás. Lo que requerimos cuando hablamos de preferencias unimodales es que todos ellos presenten este tipo de valoración: «cuanto más alejada esté una alternativa de la mejor opción, peor».

[6] La misma formulación es aplicable a muchos otros casos en los que las alternativas sociales se refieren a cantidades de un bien público, distancias al centro de la ciudad de un nuevo servicio, o pueden ser descritas mediante un índice que mide la magnitud de ciertas características.

Se dice que un individuo *i* es un ***agente mediano*** si el valor de su pico, $p(i)$, es tal que el número de individuos con valores de su pico menores o iguales a $p(i)$ es mayor o igual que $m/2$, y el número de individuos con valores de su pico mayores o iguales a $p(i)$ es también mayor o igual que $m/2$ (donde m es el total de individuos de la sociedad). El agente mediano es aquel que presenta un valor de su pico que divide a la población en dos mitades iguales, los que desean que se construya un número de kilómetros de alta velocidad mayor o igual que su pico y los que prefieren un número de kilómetros menor o igual. Es fácil comprobar que siempre existe algún agente mediano (y también que puede haber más de uno).

Puede comprobarse que la aplicación de la regla de mayoría en este caso genera una preferencia social acíclica y completa, aunque no necesariamente transitiva, que verifica los principios de Respeto a la Unanimidad y Eficiencia Informacional. En el caso especial en que el número de agentes es impar y no hay alternativas indiferentes para los agentes, entonces la preferencia mayoritaria es transitiva. La decisión social corresponde al pico del agente mediano.

Hay otras restricciones de dominio que también garantizan la «racionalidad» de los resultados obtenidos mediante el Método de Decisión Mayoritaria. Las más conocidas son las siguientes:

Restricción del Valor.- *Para todo trío de alternativas {x, y, z} de X, existe una alternativa, digamos x, tal que todos los individuos la consideran como que no es la mejor, o bien que no es la peor, o bien que no es la alternativa intermedia.*

Acuerdo Limitado.- *Para todo trío {x, y, z} de X, existe un par ordenado {x, y} de alternativas tal que todo el mundo considera que **x** es, al menos, tan buena como y.*

Puede probarse el siguiente resultado:

TEOREMA III.4: [Sen y Pattanaik (1969)]

La regla de decisión mayoritaria genera una relación social de preferencias acíclica, si y sólo si se verifica alguna de las dos siguientes condiciones: Restricción de Valor o Acuerdo Limitado.

¿En qué condiciones podemos esperar que se cumplan estas restricciones? La respuesta a esa cuestión depende esencialmente del número de individuos y de alternativas. Estas restricciones resultan plausibles en situaciones en las que un pequeño grupo de personas vota con respecto a un reducido número de alternativas. Sin embargo, para sociedades grandes y un número abundante de alternativas, la probabilidad de que tales condiciones sean satisfechas resulta minúscula. En particular «en los problemas económicos de asignación y distribución referidos a un espacio de mercancías rico, hay muy pocas posibilidades de que las condiciones requeridas sean satisfechas» [Sen (1982, pag. 12)].

Como resumen podemos concluir que, aunque en determinadas circunstancias las restricciones sobre las preferencias individuales permiten escapar al resultado de imposibilidad, no hemos encontrado todavía un procedimiento adecuado de establecer reglas de valoración social generales, eficientes, racionales y democráticas.

Observación.- *Un buen ejercicio de comprensión consiste en comprobar que las preferencias de los Vicepresidentes del ejemplo de la Sección III.2 no son unimodales, ni verifican la Restricción de Valor ni el Acuerdo Limitado.*

Como hemos señalado anteriormente, el enfoque dado por Arrow al problema de elección colectiva puede entenderse como un intento de extender las propiedades de los procedimientos de votación, evitando el problema de la intransitividad que presentan. Las propiedades que pide a su función de bienestar social son así generalizaciones de las propiedades que satisface el método de votación por mayoría. Sin embargo, es casi imposible encontrar algún otro método distinto del de votación mayoritaria en este contexto (y por consiguiente, si nos empeñamos en pedir además transitividad eliminamos también este procedimiento y nos quedamos sin ninguno).

La relación entre las condiciones de Arrow y el Método de Decisión Mayoritaria fue analizada por May en 1952. Este investigador se pregunta qué tipo de reglas de elección social, no necesariamente transitivas, son compatibles con las propiedades de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional, Monotonía, y otra propiedad muy débil, denominada Anonimato, que asegura el carácter democrático de la regla de elección social. El ***Anonimato*** establece que lo que es importante para la decisión colectiva son las preferencias de los individuos con independencia de qué individuo tiene qué preferencia (los nombres, posición social, la raza, el credo político o la confesión religiosa no cuentan). Así, la valoración social de las alternativas no debe cambiar si dos individuos cualesquiera intercambian sus valoraciones individuales.

May prueba el siguiente resultado:

Teorema III.5: [May (1952)]

La única forma de agregar las preferencias individuales que satisface Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional, Monotonía y Anonimato es el Método de Decisión Mayoritaria.

Como el Método de Decisión Mayoritaria no es transitivo, de aquí se sigue inmediatamente que no hay ninguna forma de agregar las preferencias individuales en un orden social que cumpla las propiedades de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional, Monotonía y Anonimato. El Teorema de Arrow es más fuerte dado que establece que lo mismo es cierto si sustituimos la condición de Anonimato por la de Ausencia de Dictador y eliminamos el requisito de Monotonía.

Ordenación social sin eficiencia informacional

¿Es posible encontrar reglas de decisión social no dictatoriales que ordenen consistentemente las alternativas y satisfagan los principios de Dominio Universal y Respeto de la Unanimidad? La respuesta es sí y además ya hemos visto alguna de ellas. Ejemplos relevantes de este tipo de reglas son la cuenta de Borda, los torneos, o los sistemas electorales «a doble vuelta».

Así pues si renunciamos al principio de eficiencia informacional podemos encontrar funciones de bienestar social que satisfacen todos los demás requisitos. Pero también hemos visto que algunas de estas reglas presentan otro tipo de inconvenientes. En el ejemplo del Presidente de la sección III.2 el sistema de elección que aplicaba la cuenta de Borda estaba sometido a la posibilidad de que los agentes manipularan sus preferencias para mejorar sus propios resultados. En el caso de los torneos veíamos que la alternativa elegida dependía del orden de ejecución de los torneos.

En realidad estos dos tipos de problemas no son peculiaridades del ejemplo considerado o de las dos reglas concretas aplicadas. Desafortunadamente afectan a todas las funciones de bienestar social que no cumplan la propiedad de eficiencia informacional.

En efecto, por un lado, sucede que todos estos métodos de agregación de preferencias resultan ***manipulables***. Es decir, los agentes tendrán en general incentivos a declarar unas preferencias que no se corresponden con las verdaderas, con el objetivo de conseguir resultados más próximos a sus intereses. Este resultado es conocido como el Teorema de Gibbard-Satterthwaite, cuyo mensaje esencial puede resumirse como sigue: Cualquier ordenación social de alternativas que satisfaga las propiedades de Dominio Universal y Respeto de la Unanimidad, y no sea dictatorial, es manipulable.

Formalmente:

Teorema III.6: [Gibbard (1973), Satterthwaite (1975)]

Toda regla de elección social que verifique las propiedades de Dominio Universal y Respeto de la Unanimidad o es manipulable o es dictatorial.

Así pues, combinando este resultado con el Teorema de Imposibilidad de Arrow, podemos concluir que de entre las reglas de decisión que verifican las propiedades de dominio universal y respeto de la unanimidad serán dictatoriales aquellas que cumplan el principio de eficiencia informacional y serán dictatoriales o manipulables aquellas que no lo cumplan.

Las reglas de decisión social que no verifican eficiencia informacional presentan otro serio inconveniente: La ordenación social de dos alternativas A y B puede verse afectada por la inclusión o la exclusión de una tercera alternativa C de la lista de alternativas sociales consideradas. Además de que esta propiedad puede considerarse indeseable por sí misma abre la puerta a otro tipo de manipulación, distinta de la manipulación de las preferencias. Se trata de la inclusión de alternativas más o menos ficticias que, sin embargo, pueden afectar a los resultados de la ordenación social[7].

[7] En muchas ocasiones la victoria en las elecciones donde compiten dos grandes partidos con apoyo popular similar se ve afectada por la aparición de una tercera opción, aunque ésta no tenga ninguna posibilidad de ganar (ya ocurrió con la elección de George Bush frente a Al Gore merced a la aparición de un candidato «ecologista»). La estrategia de introducir una alternativa «irrelevante» que no obstante puede afectar al resultado de una decisión, es usada en muchos contextos de forma sistemática.

Veamos ahora con algo más de detalle dos procedimientos de agregación de preferencias que no satisfacen la eficiencia informacional pero sí todas las demás propiedades de una función de bienestar social. El primero es la cuenta de Borda y el segundo el criterio lexicográfico que se aplica a la confección del «medallero» de los juegos olímpicos.

La cuenta de Borda

La regla de Borda (o *cuenta de Borda*) es uno de los métodos más usuales que no verifican eficiencia informacional (un método de decisión colectiva común en determinados concursos como premios literarios o festivales de música). Se trata de un sistema de valoración que consiste en asignar valores decrecientes a las distintas posiciones que ocupan las alternativas, ordenadas de mejor a peor. Así por ejemplo si existen n alterativas damos n puntos a la primera, $(n-1)$ a la segunda, ..., y un punto a la que ocupa la última posición. Sumando los puntos obtenidos por cada alternativa, según la valoración de los individuos de la sociedad, generamos un ranking completo que describe la valoración social. Este método de agregación de preferencias no es dictatorial y verifica las condiciones Dominio Universal, Respeto de la Unanimidad y Monotonía. Pero no la Eficiencia Informacional.

Para comprobar que no verifica la propiedad de eficiencia informacional consideremos el siguiente ejemplo en el que hay 4 alternativas y cuatro agentes. Cada agente ordena las alternativas de peor a mejor y le asigna un número de puntos que coincide con el número de orden (1 punto para la última, 2 para la antepenúltima, 3 para la siguiente y 4 para la mejor).

Cuadro III.5

	Agente 1	Agente 2	Agente 3	Agente 4	*Total*
Alternativa A	1	4	3	2	*10*
Alternativa B	2	1	4	3	*10*
Alternativa C	3	2	1	4	*10*
Alternativa D	4	3	2	1	*10*

Esta tabla nos dice que todas las alternativas son igualmente valoradas por la sociedad.

Supongamos ahora que eliminamos la alternativa B y aplicamos de nuevo el procedimiento de valoración anterior, que ahora otorgará 1, 2 y 3 puntos, respectivamente, a las alternativas peor, intermedia y mejor. Tendremos lo siguiente:

Cuadro III.6

	Agente 1	Agente 2	Agente 3	Agente 4	*Total*
Alternativa A	1	3	3	2	9
Alternativa C	2	1	1	3	7
Alternativa D	3	2	2	1	8

Observamos que al eliminar la alternativa B, ahora aparecen las alternativas perfectamente ordenadas de modo que A es la mejor, D es la segunda y C es la tercera.

Así la posición relativa de dos alternativas se ve influida por la presencia o no de una tercera, lo que prueba que esta regla no cumple la eficiencia informacional.

En cuanto a la posibilidad de manipulación de la regla de Borda como criterio de agregación de preferencias, es fácil identificar cómo opera. Los agentes tienen incentivos a minusvalorar estratégicamente las alternativas que compiten más directamente con la que uno quiere obtener.

Esta posibilidad de manipulación de las preferencias está presente en prácticamente todos los procedimientos subjetivos de valoración numérica y es un fenómeno bien conocido. Un ejemplo notable lo constituye la puntuación de las pruebas de gimnasia en los Juegos Olímpicos o en los Campeonatos del Mundo. Cada juez debe dar una puntuación subjetiva de 1 a 10 en cada prueba. En muchas ocasiones se tiene la impresión de que los jueces conceden sus puntos de forma estratégica para favorecer o perjudicar la clasificación de un determinado atleta o país. Por ello desde hace años se introdujo la práctica de eliminar del cómputo la puntuación más alta y más baja recibida por cada atleta. De este modo no se pueden manipular groseramente las puntuaciones.[8]

El Medallero olímpico

Después de cada nueva edición de los Juegos Olímpicos los medios de comunicación establecen una clasificación global de los países participantes en función del número de medallas obtenidas. Se trata del conocido «medallero».

Este medallero se construye ordenando los países según el número de medallas de oro obtenidas. Y de entre los que tienen las mismas medallas de oro poniendo delante a los que tienen más medallas de plata (o más de bronce, caso de que tuvieran igual número de oro y plata). De este modo se consigue una ordenación completa de los países sin recurrir a determinar «a cuántas medallas de plata equivale una medalla de oro».

[8] Sin embargo este procedimiento no elimina por completo la manipulación. El lector puede pensar en las posibilidades de manipulación que surgen entre dos jueces que actuaran de forma coordinada.

En realidad la clasificación del conjunto de países puede interpretarse como un problema de elección colectiva en el que las alternativas sociales a ordenar son los diferentes países participantes y las clasificaciones en las diferentes pruebas juegan el papel de las valoraciones de los agentes. Obsérvese que aquí no hay en realidad «preferencias» o valoraciones subjetivas sino que la regla de ordenación trabaja con *inputs* objetivos (las clasificaciones de los distintos países en las competiciones en que participan). No hay pues un problema de manipulación de preferencias por lo que el recurso a una regla que no verifique el principio de eficiencia informacional parece más razonable.

La regla de elección social con la que se confecciona el medallero corresponde a la aplicación de un criterio de tipo lexicográfico, que de forma general puede describirse del siguiente modo. Las alternativas se ordenan según el número de «primeras posiciones» que ocupen en las preferencias de los agentes. Dadas dos alternativas A y B colocamos delante aquella que tiene un mayor número de primeras posiciones. Caso en que tengan el mismo número de primeras posiciones, ordenamos primero a aquella que tenga mayor número de segundas posiciones en los ranking individuales. Caso de que las dos alternativas tengan igual número de primeras y segundas posiciones, colocamos en primer lugar a aquella que tenga mayor número de terceras posiciones. Y así sucesivamente.

Es fácil comprobar que este procedimiento ordena completamente las alternativas sociales y verifica las propiedades de Dominio Universal y Respeto de la Unanimidad. Pero no satisface el requisito de la eficiencia informacional, como ilustra el siguiente ejemplo correspondiente a 4 países (A, B, C, D) y 4 pruebas (competiciones) distintas. Los posibles tipos de medallas obtenidas en estas pruebas son: Oro, Plata, Bronce y Ninguna.

El siguiente cuadro ilustra los resultados obtenidos por estos países en las diferentes pruebas.

Cuadro III.7

	Prueba 1	Prueba 2	Prueba 3	Prueba 4
País A	Ninguna	Ninguna	Plata	Bronce
País B	Bronce	Plata	Ninguna	Oro
País C	Plata	Oro	Oro	Ninguna
País D	Oro	Bronce	Bronce	Plata

La clasificación resultante sería: C (dos oros), D (un oro, una plata y dos bronces), B (un oro, una plata y un bronce) y A (ningún oro).

Si ahora eliminamos al país B de la competición (imaginemos una descalificación global por consumo generalizado de sustancias prohibidas), y reasignamos consecuentemente las medallas, los nuevos resultados serían:

Cuadro III.8

	Prueba 1	Prueba 2	Prueba 3	Prueba 4
País A	Bronce	Bronce	Plata	Plata
País C	Plata	Oro	Oro	Bronce
País D	Oro	Plata	Plata	Oro

La nueva clasificación que se obtendría sería: D (dos oros y dos platas), C (dos oros y una plata) y A (ningún oro). Por tanto la posición relativa de D y C se ve afectada por la eliminación de una alternativa (la B), lo que prueba que este método clasificatorio no verifica la eficiencia informacional.

III.5 Más sobre votaciones

Consideremos ahora específicamente los problemas de decisión colectiva que consisten en la elección de una alternativa mediante algún procedimiento de votación. Para fijar ideas podemos identificar el conjunto de alternativas con una lista de *candidatos* (personas o proyectos a realizar) y la sociedad con el censo de *votantes*. Hemos visto que, en un contexto general, el método de decisión mayoritaria puede generar «ciclos» que nos impidan tomar una decisión sobre qué alternativa elegir. Y también que hay otros procedimientos que no presentan este problema, a costa de no verificar alguno de los otros requisitos considerados deseables (en particular la eficiencia informacional).

Nos ocuparemos en esta sección de estudiar las propiedades de diferentes procedimientos de votación, con especial atención a su capacidad resolutiva. En este sentido lo mejor que podemos esperar es que una regla de votación seleccione un único candidato. Y lo peor que no sea capaz de seleccionar a ninguno (como sucede en algunos casos con el método de decisión mayoritaria). Una situación intermedia es aquella en que hay varios candidatos que la regla propone como elegidos (un caso de «empate» que requiere algún procedimiento adicional para resolver). Todo ello sin olvidar que una regla decisión que respete la unanimidad o es dictatorial o es manipulable, de modo que aquellas reglas que sean democráticas resultarán siempre susceptibles de manipulación.

La discusión siguiente se refiere básicamente a reglas de votación que derivan de la valoración de las alternativas mediante un sistema de puntos (lo que se suele llamar *scoring rules*), o bien comparan las alternativas mediante alguna forma de *torneo*. Ambos enfoques se desarrollan originariamente en el siglo XVIII. El primero supone una generalización del criterio de valoración de Borda y el segundo está asociado al enfoque de Condorcet.

Votación por puntos y votación mediante torneos

La regla de Borda, que acabamos de discutir es, en realidad, un caso particular de una familia de sistemas de elección mediante la asignación de puntos a las diferentes posiciones en que se ordenan las alternativas y la elección de aquella que obtiene mayor puntuación.

Una ***regla de votación por puntos*** es un procedimiento de elección que consiste en que cada individuo (votante) asigna un cierto número de puntos a las diferentes alternativas (candidatos) según el lugar que ocupen. Suponiendo que hay q alternativas, cada individuo asigna un valor p_1 a la peor alternativa, un valor $p_2 \geq p_1$ a la siguiente, un valor $p_3 \geq p_2 \geq p_1$ a la que ocupa el tercer peor lugar, etc. hasta llegar a la alternativa preferida a la que se le asigna un valor p_q mayor o igual que todos los demás y estrictamente mayor que p_1. La regla de Borda es un caso particular de este procedimiento en el que tomamos $p_1 = 1, p_2 = 2, \ldots, p_q = q$.

Pero también el ejemplo del medallero olímpico podemos interpretarlo como un caso particular de votación por puntos, en el que tomamos $p_1 = p_2 = \ldots = p_{q-1} = 0$, $p_q = 1$. Ello equivale a decir que únicamente tomamos en consideración la alternativa más votada por cada candidato. En el ejemplo del Presidente ésta era precisamente una de las vías que había intentado para decidir la política a seguir, viendo cuál de las alternativas tenía más apoyo entre los militantes de su partido (cuadro III.2). Ya vimos allí los problemas de este tipo de regla, que se conoce en general como ***Regla Plural*** (*plurality voting*) y que consiste en que cada elector vota a su candidato preferido y la regla elige al candidato con más votos.

También podemos considerar un procedimiento de elección simétrico del anterior, conocido como Regla Antiplural, en la que cada votante propone descartar al candidato que menos le gusta y resulta elegido el candidato que menos objeciones recibe.

Adviértase que todas estas reglas siempre eligen a algún candidato pero también que pueden elegir a más de uno, en cuyo caso hace falta recurrir a algún expediente adicional para resolver. En la vida real esto se hace de formas variadas, como son eligiendo al de más edad (o al de menos edad), ordenando los candidatos pre-seleccionados por orden alfabético, o recurriendo a un «voto de calidad» de un votante distinguido (el Presidente de una comisión, por ejemplo). Pero en ocasiones suponen un problema fundamental, como ocurre en el ejemplo del Presidente de la Tabla III.1 si aplicamos el criterio de Borda, dando 3 puntos a la alternativa mejor, dos a la siguiente y uno a la última. Es inmediato comprobar que las tres alternativas obtienen 6 puntos cada una, de modo que todas son indiferentes y no hemos conseguido resolver nada (en este caso no por la presencia de ciclos de preferencias sino porque todas las alternativas son declaradas indiferentes).

Otro de los criterios de elección de políticas ensayado por el Presidente de la sección III.2 se refería a comparar las alternativas por pares, como si fueran torneos, e ir seleccionando en cada caso al ganador para enfrentarlo con los ganadores de otros torneos hasta llegar a un ganador total. También vimos que este procedimiento podía no ser resolutivo: puede ocurrir que el ganador dependa de la secuencia de emparejamientos (o, lo que es lo mismo, que no haya un candidato capaz de vencer a todos los demás en cualquier torneo que se plantee).

De forma general se dice que una alternativa es un ***ganador de Condorcet*** cuando es elegido frente a cualquier otro en comparaciones por pares. Es decir, dadas las alternativas $\boldsymbol{X} = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots, \boldsymbol{q}\}$ y los votantes $N = \{1, 2, \ldots, n\}$, la alternativa $\boldsymbol{a}$ es un ganador de Condorcet si para todo par de alternativas compuesto por $\boldsymbol{a}$ y otra cualquiera $\boldsymbol{b}$, hay más votantes que prefieren $\boldsymbol{a}$ sobre $\boldsymbol{b}$ que los que prefieren $\boldsymbol{b}$ sobre $\boldsymbol{a}$.

Es fácil deducir que cuando existe un ganador de Condorcet entonces es único, es decir, selecciona a un único candidato. Pero también sabemos, por el ejemplo descrito en el cuadro III.4, que un ganador de Condorcet no siempre existe. Puede comprobarse que, para un número dado de votantes, la probabilidad de que no exista un ganador de Condorcet crece rápidamente con el número de candidatos. Cuando tenemos tres alternativas (candidatos) y tres agentes (votantes), como en el ejemplo de referencia, la probabilidad de que no exista un ganador de Condorcet es aproximadamente del 6 %.

Se dice que una regla de votación es ***consistente con Condorcet*** si siempre elige el ganador de Condorcet, cuando existe. Las dos siguientes reglas de votación son reglas consistentes con Condorcet:

Regla de Copeland: Consiste en realizar comparaciones entre pares de candidatos, $\boldsymbol{a}$ y $\boldsymbol{b}$, dando una puntuación de +1 si una mayoría prefiere $\boldsymbol{a}$ sobre $\boldsymbol{b}$, de −1 en caso contrario, y de 0 si $\boldsymbol{a}$ obtiene tantos votos como $\boldsymbol{b}$. Sumando todas estas puntuaciones para cada alternativa $\boldsymbol{b}$ con la que $\boldsymbol{a}$ se compara obtenemos la puntuación de $\boldsymbol{a}$. La regla de Copeland selecciona al candidato con mayor puntuación, también llamado ganador de Copeland.

Regla de Simpson: Dado un candidato $\boldsymbol{a}$ llamamos $p(\boldsymbol{a}, \boldsymbol{b})$ al número de votantes que prefiere $\boldsymbol{a}$ sobre $\boldsymbol{b}$, para cada posible alternativa $\boldsymbol{b}$ que comparamos con $\boldsymbol{a}$. La puntuación de Simpson para la alternativa $\boldsymbol{a}$ es el mínimo de estos números $p(\boldsymbol{a}, \boldsymbol{b})$. La regla de Simpson escoge al candidato con una mayor puntuación calculada de este modo, también llamado ganador de Simpson.

Aunque estas dos reglas son consistentes con Condorcet aplican criterios diferentes. Para ser elegido según la regla de Copeland, hay que ganar en los torneos por pares al mayor número posible de candidatos rivales, aunque sea por un estrecho margen de votos. Para que $\boldsymbol{a}$ sea un ganador de Simpson debe ocurrir que, en las comparaciones por parejas, ningún otro candidato reciba un apoyo muy grande.

Recordemos que las preferencias de un colectivo se dicen unimodales cuando podemos disponer las alternativas sociales mediante un cierto orden fijo (por ejemplo sobre la recta real) y las preferencias de cada individuo se caracterizan por tener una alternativa preferida $\boldsymbol{p(i)}$ (su «pico») con la particularidad de que las alternativas sociales son peores cuanto más se alejan de esta opción preferida $\boldsymbol{p(i)}$. En este contexto se define el ***agente mediano*** como aquel cuyo pico $\boldsymbol{p(i)}$ divide a la población en dos mitades iguales (los que tienen picos mayores o iguales que el suyo y los que tienen picos menores o iguales).

El pico de un agente mediano tiene una propiedad interesante desde el punto de vista de la decisión mayoritaria: cuando las preferencias son unimodales el pico de cualquier agente mediano es un ganador de Condorcet.

Borda y Condorcet frente a frente

Las reglas de votación por puntos y las reglas consistentes con Condorcet aplican criterios de elección diferentes y tienen propiedades diversas. En particular la regla de Borda (o en general cualquier regla de votación por puntos) siempre determina algún candidato elegido, si bien en muchas ocasiones no es único. Por su parte hemos visto que la regla de Condorcet elige un único candidato, el ganador de Condorcet, pero que dicho ganador no siempre existe.

Consideremos el siguiente ejemplo en el que describimos las preferencias de una sociedad compuesta por 21 individuos que deben elegir entre cuatro candidatos ***a***, ***b***, ***c***, ***d***. La siguiente tabla muestra las preferencias completas de estos individuos, agrupados según ordenen a los candidatos (tres de ellos prefieren ***a*** sobre ***b***, ***b*** sobre ***c***, y ***c*** sobre ***d***; cinco de ellos prefieren ***a*** sobre ***c***, ***c*** sobre ***b***, y ***b*** sobre ***d***; etc.).

Cuadro III.9

3 votantes	5 votantes	7 votantes	6 votantes
a	a	b	c
b	c	d	b
c	b	c	d
d	d	a	a

Si en esta situación aplicamos la denominada «regla plural» (cada votante únicamente determina el candidato preferido), el candidato ***a*** resulta elegido con un total de 8 votos, frente a 7 del candidato ***b***, 6 del candidato ***c*** y ninguno del candidato ***d***. Sin embargo, como señala Condorcet, ***a*** es el peor candidato para la mayoría de los electores (13 sobre 21 son de esta opinión). Es fácil comprobar que una mayoría de 14 votantes estima que ***c*** es mejor candidato que ***d***, y otra mayoría de 11 votantes opina que ***c*** es mejor que ***b***. Por tanto, dado que ***c*** derrota en todos los torneos por pares a los demás candidatos, ***c*** debiera ser el candidato elegido.

Borda coincide con Condorcet en que ***a*** es una mala elección. Pero él propone que el candidato elegido debe ser ***b***, dado que ***b*** resulta la opción preferida por 7 votantes (mientras que ***c*** es la primera opción sólo para 6 votantes), y es la primera o segunda opción para 16 votantes (mientras que ***c*** lo es solo para 11). Y ***b*** está en los tres primeros puestos para todos los votantes, al igual que ***c***. Es fácil comprobar que en esta situación cualquier regla de votación por puntos dará siempre ganador al candidato ***b***.

Este ejemplo ilustra bien la contraposición entre los principios que inspiran los criterios de Condorcet y Borda, dado que nos encontramos en una situación en la que existe un ganador de Condorcet (la alternativa ***c***) que no resulta elegido por ninguna regla de votación por puntos.

Otro modo de confrontar estos sistemas de votación es tomando en cuenta su comportamiento en un contexto en el que la decisión se produce en una sociedad dividida en un conjunto de demarcaciones electorales. Consideremos el caso más sencillo en que hay únicamente dos demarcaciones, es decir, la sociedad N viene definida como la unión de dos subgrupos de población N_1 y N_2. Hay un conjunto X de candidatos de entre los que debemos elegir alguno a partir de los votos de las dos demarcaciones. Parece razonable suponer que si uno o varios candidatos son elegidos tanto en N_1 como en N_2, entonces la elección que se derivaría de considerar la sociedad N como una demarcación única debería elegir precisamente a este (o estos) candidato(s). Esta propiedad se conoce como ***refuerzo***.

Puede probarse el siguiente resultado:

Teorema III.7: [Young (1975)]

(i) Todas las reglas de votación por puntos verifican la propiedad de refuerzo.

(ii) No existe ninguna regla consistente con Condorcet que verifique la propiedad de refuerzo.

Así pues las reglas de votación inspiradas en los principios propuestos por Borda y Condorcet tienen propiedades diametralmente opuestas cuando las comparamos desde el puntos de vista de un proceso de votación en demarcaciones. Todas las reglas de votación por puntos eligen en la sociedad grande los candidatos seleccionados unánimemente por las sociedades pequeñas, y ninguna regla consistente con Condorcet es capaz de cumplir este requisito[9].

Nos queda por discutir las posibilidades que se derivan de utilizar más información que la contenida en las ordenaciones individuales. En particular, la utilización de comparaciones interpersonales de utilidad. Abordaremos este punto en el siguiente capítulo donde comprobaremos que la regla de elección social que obtenemos depende del tipo de comparación interpersonal de utilidad que admitamos.

[9] En realidad los sistemas de votación mediante puntos son prácticamente los únicos capaces de satisfacer la propiedad de refuerzo, como prueba Young (1975).

Apéndice al Capítulo III: Prueba del Teorema de Imposibilidad de Arrow

Sean $M = \{1,2,...,m\}$ el conjunto de índices que identifica los agentes de una sociedad, X el conjunto de alternativas sociales, $\succsim_i$ la relación de preferencias del agente i definidas sobre X, y sea R la ***relación de preferencia social*** sobre X. Una ***Función de Bienestar Social*** es una regla de elección colectiva F tal que $R = F(\succsim_1, \succsim_2, ..., \succsim_m)$ donde R es una relación binaria transitiva y completa. Denominaremos P, I a las relaciones sociales de preferencia estricta y de indiferencia, respectivamente, asociadas a R.

Consideremos las siguientes condiciones que la regla F debe cumplir:

Dominio Universal: F está definida para toda posible configuración de preferencias individuales.

Respeto de la Unanimidad: Para todo x, y de X, $x \succ_i y$ para todo i de M implica $x\ P\ y$.

Eficiencia Informacional: Sea Y un subconjunto de X, y sean $R_Y(X)$ el orden social de X restringido al conjunto Y, $R(Y)$ el orden social relativo al conjunto Y. Entonces, $R_Y(X) = R(Y)$.

Monotonía[10]: Sea $(\succsim_1, \succsim_2, ..., \succsim_m)$ una configuración de preferencias dada y sea R la relación social correspondiente. Sean x, y dos elementos de X tales que $x\ R\ y$, y sea A el subconjunto de M compuesto por todos aquellos individuos que consideran x mejor o igual que y. Sea $(\succsim^a_1, \succsim^a_2, ..., \succsim^a_m)$ una configuración alternativa de preferencias individuales tales que para todo agente i de A se cumple que $x \succsim^a_i y$. Entonces, $x\ R^a\ y$ (donde R^a es la relación social asociada a las nuevas preferencias).

Como ya señalamos, se dice que F es ***dictatorial*** si existe algún j de M tal que, para cualquier configuración de preferencias y para todo par de alternativas sociales x, y de X, se verifica que si j prefiere x a y entonces $x\ P\ y$. Al individuo j se le denomina en tal caso un ***dictador***.

Definición: *Dada una regla de elección social F, se dice que un grupo de individuos $G(x, y)$ contenido en M es un* ***grupo decisorio*** *con respecto al par $\{x, y\}$, si para toda configuración de preferencias se verifica que si x es preferido a y para todo i de $G(x, y)$, entonces $x\ P\ y$.*

Un grupo es decisorio con respecto a un par de alternativas $\{x, y\}$ si cuando todos prefieren x a y resulta que la valoración social es $x\ P\ y$. También aquí esta capacidad decisoria se cumple para cualquier configuración de preferencias.

[10] En realidad esta propiedad se denomina «Monotonía Débil», para diferenciarla de otra muy similar pero algo más exigente llamada Monotonía, que aquí no usaremos.

El Teorema de Imposibilidad de Arrow, también llamado *Teorema General de Posibilidad*, se prueba a partir de dos curiosos lemas. El primero nos dice que todo grupo decisorio sobre un cierto par de alternativas con dos o más miembros posee un subconjunto propio que también es decisorio (aunque no necesariamente sobre las mismas alternativas). El segundo establece que todo grupo decisorio sobre un par de alternativas resulta serlo sobre todas ellas.

Lema 1:

Sea F una Función de Bienestar Social que verifica las condiciones Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Monotonía, y sea G(x, y) un grupo decisorio de dos o más miembros. Entonces existe un subconjunto propio de G(x, y) que también es decisorio sobre algún par de alternativas.

Demostración

Sea $(\succsim_1, \succsim_2, ..., \succsim_m)$ un perfil de preferencias donde los individuos de $G(x, y)$ están separados en dos subconjuntos no vacíos G', G''. Sea z otro elemento de X y consideremos las ordenaciones individuales relativas a la terna $\{x, y, z\}$ (lo que es posible por la condición de Eficiencia Informacional). Sea G^C el conjunto complementario de $G(x, y)$ en M y, haciendo uso de la propiedad de Dominio Universal, supongamos que:

$$z \succ_i x \succ_i y \text{ para todo } i \text{ de } G'$$

$$x \succ_i y \succ_i z \text{ para todo } i \text{ de } G''$$

$$y \succ_i z \succ_i x \text{ para todo } i \text{ de } G^C$$

Necesariamente se sigue que $x\ P\ y$ dado que $G(x, y)$ es un grupo decisorio. Por su parte z puede ocupar las siguientes posiciones:

(a) $z\ R\ x\ P\ y$, en cuyo caso los individuos en G' resultan ser un grupo decisorio sobre el par $\{z, y\}$, puesto que son los únicos que ordenan z antes que y, y F verifica la propiedad de Monotonía[11].

(b) $x\ P\ y\ R\ z$, en cuyo caso los individuos en G'' constituyen un grupo decisorio sobre el par $\{x, z\}$, puesto que son los únicos que ordenan x antes que z y se verifica la condición de Monotonía.

(c) $x\ P\ z\ R\ y$, en cuyo caso G'' resulta ser un grupo decisorio sobre el par $\{x, z\}$, por las mismas razones de antes.

(d) $x\ R\ z\ P\ y$, en cuyo caso G' resulta ser un grupo decisorio sobre el par $\{z, y\}$, por las mismas razones de antes.

Q.e.d.

[11] Esta propiedad garantiza que G' es efectivamente un grupo decisorio, puesto que la ordenación social no cambiará si algunos agentes que no están en G' cambian sus preferencias sobre el par $\{y, z\}$ a favor de la alternativa z.

Lema 2:

Sea F una Función de Bienestar Social que verifica los supuestos de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Monotonía, y sea G(x, y) un grupo decisorio sobre el par {x, y}. Entonces este mismo grupo resulta decisorio sobre el par {w, z} para cualquier {w, z} de X.

Demostración

Dada la condición de Dominio Universal, podemos suponer que existe un perfil de preferencias $(\succ_1, \succ_2, \ldots, \succ_m)$ tal que:

$w \succ_i x \succ_i y \succ_i z$ para todo i de $G(x, y)$

$y \succ_i z \succ_i w \succ_i x$ para todo i de G^C, donde, como antes, G^C es el conjunto complementario de $G(x, y)$.

Haciendo uso de las propiedades de Eficiencia Informacional y de Respeto de la Unanimidad, podemos establecer: $w\,P\,x$, por aplicación del principio de Respeto a la Unanimidad; $x\,P\,y$, por ser $G(x, y)$ un grupo decisorio. Consecuentemente, por transitividad (ya que la Función de Bienestar Social ordena las alternativas) se cumplirá que $w\,P\,y$.

Por tanto, como los individuos de $G(x, y)$ son los únicos que ordenan w antes que y, resulta que también son decisorios sobre el par $\{w, y\}$ (usamos aquí nuevamente de forma implícita la propiedad de Monotonía). Podemos escribir pues: $w\,P\,y$, según acabamos de deducir. Además, $y\,P\,z$ por aplicación del principio de Respeto a la Unanimidad. Y, de nuevo por transitividad de la preferencia social, concluimos que $w\,P\,z$.

En consecuencia, como sólo los individuos en $G(x, y)$ ordenan w antes que z, este grupo también es decisorio sobre el par $\{w, z\}$.

Q.e.d.

Teorema: [Teorema de Imposibilidad de Arrow]

Sea X un conjunto de elección con más de dos alternativas. Si una Función de Bienestar Social F definida sobre X verifica las condiciones de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Monotonía Débil, entonces es dictatorial.

Demostración

El Respeto de la Unanimidad implica que M es un grupo decisorio sobre el par $\{x, y\}$. Aplicando recursivamente el Lema 1 nos quedaremos con un grupo decisorio de una sola persona, que tiene capacidad de decisión sobre un par de alternativas $\{z, w\}$.

El Lema 2 asegura entonces que esta persona tiene capacidad de decisión sobre *todas* las alternativas de X (es decir, es un *dictador*).

Q.e.d.

CAPÍTULO

IV

«Play it again, Sam»

(Frase nunca pronunciada por Bogart en *Casablanca*)

IV. Racionalidad colectiva (2)

IV.1 Más allá de Arrow: Comparabilidad interpersonal

En el capítulo anterior hemos visto que no parece haber vías de escape muy satisfactorias al resultado de imposibilidad de Arrow. La razón fundamental deriva del empeño de construir una regla de decisión muy potente (que genere siempre ordenaciones completas) a partir de muy escasos ingredientes (las ordenaciones individuales). Ninguna de las variantes consideradas altera sustancialmente esta situación.

Un camino alternativo y que abre posibilidades de definir criterios de ordenación social es aquel en que estamos dispuestos a establecer comparaciones interpersonales de utilidad. Es decir, juicios del tipo «el individuo i está mejor en la alternativa x que el individuo j», o del tipo «el individuo i está mucho mejor que el individuo j en la alternativa x que en la alternativa y». Con ello nos separamos del enfoque de Arrow en un aspecto sustancial ya que introducimos en la forma de valoración social algo más que las meras ordenaciones individuales de alternativas. Para ello se necesita tomar como variables de nuestra regla de elección social las funciones de utilidad individuales y no las relaciones de preferencia. El motivo es que con las funciones de utilidad podemos introducir condiciones que nos permitan establecer comparaciones interpersonales e intensidades de preferencia[1].

Así pues, a diferencia de la formulación de Arrow, ahora los argumentos de la regla de elección son las funciones de utilidad u_i para $i = 1, 2, \ldots, m$. La función u_i nos dice, en términos numéricos, cómo valora el individuo i las diferentes alternativas sociales. Así, para comparar socialmente dos alternativas $\{x, y\}$, los *inputs* de nuestra regla de elección social serán ahora los vectores de números $[u_1(x), u_2(x), \ldots, u_m(x)]$ y $[u_1(y), u_2(y), \ldots, u_m(y)]$.

El siguiente resultado garantiza que, bajo las condiciones habituales, el conocimiento de estos vectores de números es suficiente para establecer una ordenación social de las alternativas.

[1] El argumento formal es algo complicado y puede consultarse por ejemplo en Villar (1999, Cap. 7).

TEOREMA IV.1: [D'Aspremont & Gevers (1977)]

Una regla de elección social que ordena completamente las alternativas sociales satisface las propiedades de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional, si y sólo si cumple la siguiente propiedad: para todo par de alternativas {x, y} se verifica que la «alternativa x» es declarada socialmente mejor que la «alternativa y» si y sólo si el vector de utilidades de los individuos en la «alternativa x» es socialmente preferido al vector de utilidades de los individuos en la «alternativa y».

Este teorema establece que la valoración de las alternativas sociales depende *exclusivamente* de los valores que los individuos les asignan en sus funciones de utilidad (una propiedad conocida como *Neutralidad*). Por tanto, sólo la información contenida en las funciones de utilidad resulta relevante para la determinación del orden social. Esta implicación tiene un gran alcance desde el punto de vista de las propiedades normativas del proceso de agregación, como veremos más adelante, ya que impide tomar en cuenta aspectos como la naturaleza de las alternativas sociales, los derechos o necesidades de los individuos, la motivación de los agentes en su forma de ordenar alternativas sociales, etc.

Suele denominarse ***Orden de Bienestar Social*** al criterio de valoración social que ordena las diferentes alternativas sociales tomando como *input* las funciones de utilidad de los individuos.

Hay esencialmente dos criterios alternativos para establecer comparaciones interpersonales de utilidad. El primero es puramente ordinal, es decir, consiste en poder decir si un individuo está mejor que otro, sin necesidad de establecer «cuánto mejor». El segundo criterio de comparación supone que podemos medir la satisfacción de los individuos en términos de ciertas unidades comunes a todos ellos. En tal caso, podemos comparar no sólo los niveles de utilidad sino las ganancias de utilidad. Ambos tipos de comparabilidad son diferentes y tienen consecuencias diversas sobre la forma de valoración social. Bajo ciertas condiciones también puede resultar posible efectuar ambos tipos de comparaciones simultáneamente.

Comparaciones ordinales de utilidad: La Regla del Leximin

Consideremos el caso en que las utilidades de los individuos son ordinales pero interpersonalmente comparables. Ello nos permite comparar *niveles* de utilidad entre individuos, es decir, podemos establecer sin ambigüedad aserciones del tipo «con la alternativa x el individuo i está mejor o igual que el individuo j», lo que se describiría como $u_i(x) \geq u_j(x)$.

La regla de elección colectiva más usual en este contexto es la conocida como ***Regla del Leximin***, que es una extensión del principio del *Maximin* propuesto por John Rawls (1971) en su análisis de la Justicia. Esta idea puede resumirse diciendo que hay que elegir siempre aquella alternativa que maximiza el bienestar del individuo que está peor. Para precisar este concepto, sean $\{x, y\}$ dos alternativas sociales, y sean $u_i(x)$, $u_i(y)$, para $i = 1, 2, ..., m$, las utilidades de los diferentes individuos bajo la alternativa x y bajo la alternativa y, respectivamente. La regla del Maximin es el Orden de Bienestar Social definido como sigue: x es mejor que y si y sólo si:

$$\min_{i\in M}\{u_i(x)\} > \min_{i\in M}\{u_i(y)\}$$

Esta sencilla formulación esconde una dificultad importante que podemos ilustrar mediante el siguiente ejemplo: Sea $M = \{1, 2, 3\}$ y consideremos dos alternativas, x, y. Las utilidades de los tres individuos para cada una de estas alternativas vienen dadas por $u_1(x) = 1$, $u_2(x) = 5$, $u_3(x) = 7$, para la alternativa x, y por $u_1(y) = 1$, $u_2(y) = 50$, $u_3(y) = 100$. La regla del Maximin declara que las alternativas x e y resultan indiferentes (puesto que de la definición se sigue que x es mejor o igual que y, al tiempo que y es mejor o igual que x).

La Regla del *Leximin* es una extensión de esta idea de valoración social que afina el principio rawlsiano evitando el tipo de problema que acabamos de señalar. La idea es que cuando nos encontramos con alternativas frente a las cuales los que están peor resultan indiferentes, entonces recurrimos a ordenarlas según sean las preferencias de «el siguiente peor». En el ejemplo anterior la alternativa y resultaría preferida a la alternativa x porque quien está peor resulta indiferente entre x e y, pero el siguiente que está peor (el individuo 2 en este caso), prefiere y a x.

Al igual que el Maximin, la regla del Leximin pretende maximizar el bienestar del individuo que está peor. El cambio que introduce es que si, al comparar dos alternativas $\{x, y\}$ del conjunto X los que están peor tienen la misma utilidad, entonces la decisión de cuál resulta socialmente preferida recae en «el primer peor que no es indiferente».

Aun así uno podría pensar que la regla del Leximin es muy discutible, puesto que únicamente toma en cuenta la utilidad de quien está peor. En el contexto de Rawls este principio tiene sentido en la medida que estamos hablando de derechos básicos de los individuos («bienes primarios» en su terminología). Pero no es obvio que sea exportable a cualquier contexto de decisión colectiva puesto que ignora cuántos individuos (de los que no son los que están peor) prefieren una alternativa a otra.

El economista y Profesor de la Universidad de Stanford Peter Hammond propuso en 1976 un criterio de equidad que ha pasado a conocerse como el «Principio de Equidad de Hammond». Este principio aplica al caso especial de una sociedad compuesta por m individuos donde se comparan dos alternativas que afectan únicamente a dos personas, siendo todas las demás indiferentes entre ambas alternativas. En tal caso, cuando únicamente hay dos personas afectadas por las alternativas sujetas a elección, debemos hacer prevalecer la preferencia de aquel de estos dos individuos que está peor. Se trata sin duda de un principio mucho más débil que el Leximin, ya que únicamente aplica a situaciones en las que hay dos individuos afectados por el resultado de la decisión. En este caso optar por dar prioridad a quien está peor parece un argumento fácilmente admisible.

Aceptar este argumento tiene, sin embargo, consecuencias inesperadas. Cuando se combina el Principio de Equidad de Hammond con las otras propiedades que venimos pidiendo a una regla de agregación de preferencias, resulta que caemos inexorablemente en el Leximin como la única regla capaz de verificarlas. Formalmente:

TEOREMA IV.2: [Hammond (1976)]:

Una regla de elección social que toma como argumento las utilidades individuales que son comparables ordinalmente verifica Dominio Universal, Respeto a la Unanimidad, Eficiencia Informacional, Anonimato y el Principio de Equidad de Hammond, si y sólo si es la Regla del Leximin.

Una forma de interpretar este resultado es la siguiente: Si aceptamos que la decisión social venga dictada por el individuo que está peor en aquellas situaciones en las que hay únicamente dos individuos involucrados y dos alternativas, las propiedades de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Anonimato extienden este sistema de valoración a cualquier número de individuos y cualquier número de alternativas (una especie de «epidemia» en la que la forma de resolver un problema muy específico contagia a todos los problemas).

Comparaciones cardinales: El Utilitarismo Clásico

El utilitarismo clásico, desarrollado durante el siglo XVIII, parte de la idea de que es posible establecer comparaciones interpersonales de diferencias de utilidad. La forma de evaluación social usada en este contexto era la de la suma de utilidades individuales: dadas dos alternativas $\{x, y\}$ del conjunto X, x era declarada socialmente mejor o igual que y si la suma de las utilidades de todos los individuos con la alternativa x era mayor o igual que la suma de las utilidades de los individuos con la alternativa y.

El argumento heurístico es el siguiente. El término $u_i(x) - u_i(y)$ expresa cuánto gana en utilidad el individuo i al cambiar de la alternativa y a la alternativa x. Si estas ganancias en utilidad podemos medirlas en términos de las mismas unidades para todos los individuos, entonces la expresión:

$$\sum_{i=1}^{m} u_i(x) - \sum_{i=1}^{m} u_i(y)$$

nos dice cuál es la ganancia social de cambiar de la alternativa y a la alternativa x. Consecuentemente, una opción x es mejor que otra opción y cuando esta diferencia es positiva. El óptimo social consiste pues en buscar la opción que maximiza la suma de las utilidades.

Nótese que para que estas sumas tengan sentido todas la utilidades deben ser medidas en términos de una unidad común (si no estaríamos «sumando peras y manzanas», por decirlo de forma gráfica)[2].

[2] Por otra parte, el hecho de que los individuos midan las utilidades desde distintos puntos de referencia (usen diferentes valores «cero») no afecta a los resultados dado que al sumar las diferencias estos distintos orígenes se cancelan, como ya ilustramos en el Capítulo II.

Es fácil comprobar que, cuando las utilidades se miden en términos de una unidad común, esta regla de elección ordena completamente las alternativas sociales y cumple además todos los criterios de agregación formulados (Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Anonimato). Más aún, es la única regla que los satisface, como lo muestra el siguiente resultado:

TEOREMA IV.3: [D'Aspremont & Gevers (1977)]

Una regla de elección social que toma como argumento las utilidades individuales que son comparables en sus unidades, verifica las condiciones de Dominio Universal, Respeto de la Unanimidad, Eficiencia Informacional y Anonimato si y sólo si es el Utilitarismo Clásico.

Es decir, bajo las condiciones del teorema una alternativa x es declarada socialmente mejor que otra alternativa y si y sólo si la suma de las utilidades de todos los individuos en la alternativa x es mayor que la suma de las utilidades de todos los individuos en la alternativa y.

IV.2 Aplicaciones de la Teoría de la Elección Social a otros ámbitos

Ilustraremos ahora la teoría desarrollada en las secciones precedentes mediante algunas aplicaciones que se separan un poco en su contenido de lo que se supone es el objeto de la elección social, pero resultan muy ilustrativas al tiempo que formalmente equivalentes a los problemas de decisión colectiva. De este modo, queremos poner de manifiesto que también podemos aprender algunas cosas de este tipo de análisis para aplicarlas a otros ámbitos.

Un campeonato de Triatlón

Imaginemos que nuestro problema consiste en ordenar a tres concursantes que participan en un campeonato de Triatlón, que está compuesto por las pruebas de Natación, Carrera y Ciclismo.

El siguiente cuadro recoge los resultados obtenidos en la competición:

Cuadro IV.1: Clasificación de los atletas en las diferentes pruebas del Triatlón

Pruebas	Concursante 1	Concursante 2	Concursante 3
Natación	1º	2º	3º
Carrera	2º	3º	1º
Ciclismo	3º	1º	2º

Este problema tiene una estructura formal idéntica a la del problema de elección social del ejemplo del Presidente en la Sección III.2. Ahora lo que tenemos que ordenar son los distintos atletas (que jugarían el papel de «alternativas sociales» en el contexto anterior) a partir de las ordenaciones que han obtenido en las distintas pruebas (que juegan el papel de las «preferencias individuales»).

Veamos qué significan aquí las diferentes propiedades que hemos venido discutiendo. Generar una ordenación social equivale a decidir el orden de los atletas en el campeonato de Triatlón (es decir, quién obtiene la medalla de oro, quién la de plata y quién la de bronce). Usar como *input* informativo las preferencias individuales significa que para esa clasificación únicamente podemos tomar en cuenta la clasificación obtenida por cada uno de ellos en las tres pruebas consideradas.

La propiedad de Dominio Universal significa que consideramos que cada atleta podría obtener en cada prueba cualquiera de los resultados posibles (primero, segundo o tercero). La propiedad de Respeto de la Unanimidad establece que si un atleta queda delante de otro en las tres pruebas, entonces también queda delante en el campeonato. La propiedad de Eficiencia Informacional significa que para saber quién de dos atletas ha quedado delante en la competición de Triatlon es suficiente conocer los resultados obtenidos por estos dos atletas en las tres pruebas que lo constituyen.

La propiedad de Monotonía, por su parte, establece que si en una competición un atleta obtiene un cierto puesto y en otra competición mejora sus resultados en las tres pruebas, entonces no puede quedar peor situado que antes.

Por último, la propiedad de Anonimato significa que la ordenación de los concursantes no puede depender de sus nombres, su color de piel o sus nacionalidades.

El Teorema de Imposibilidad de Arrow nos previene sobre lo que puede ocurrir si queremos determinar la ordenación de los atletas a partir de las posiciones que ocuparon en las diferentes pruebas: podemos encontrarnos con que no es posible decidir cuál es el orden de los competidores de Triatlón. El ejemplo de la tabla anterior, que reproduce el de los tres Vicepresidentes, es una prueba de ello.

Los resultados obtenidos cuando se pueden establecer comparaciones interpersonales pueden ayudarnos también aquí a decidir el resultado. Supongamos ahora que, en cada prueba, damos una puntuación (utilidad) a los diferentes resultados. Por ejemplo, quien gana la prueba de Natación obtiene 10 puntos mientras el segundo y el tercero obtienen 5 y 0 puntos, respectivamente. El ganador de la prueba de Carrera obtiene 20 puntos y los que quedan segundo y tercero, 10 y cero puntos. Por último, la prueba de Ciclismo se valora como la prueba de Natación.

En este contexto la regla utilitarista es una fórmula natural de agregación, ya que ordena a los atletas en función del total de puntos alcanzados. Así, traduciendo a puntos los resultados anteriores, obtendríamos:

Cuadro IV.2: Puntuación de los atletas en las diferentes pruebas del Triatlon

Pruebas	Concursante 1	Concursante 2	Concursante 3
Natación	10	5	0
Carrera	10	0	20
Ciclismo	0	10	5
Puntuación total	20	15	25

En este caso proclamaríamos vencedor al concursante 3, siendo la medalla de plata para el concursante 1 y la de bronce para el concursante 2.

Quizás la ordenación de los concursantes de un campeonato de Triatlon nos parezca un ejemplo pintoresco muy alejado de nuestras preocupaciones más habituales. Sin embargo, este mismo tipo de dificultad se plantea en la selección de candidatos para muchos puestos de trabajo: Profesores para una plaza en la Universidad, opositores de una determinada convocatoria pública, aspirantes a un puesto de responsabilidad en una empresa, etc. En todos estos casos encontraremos diferentes candidatos acerca de los cuales podemos valorar distintas cualidades (actitud, experiencia, formación, etc.). Si no tenemos una idea clara de cómo comparar una cualidad con otra, por ejemplo mediante el uso de un baremo predeterminado, no podremos realizar una selección sistemática de los candidatos. Por supuesto si encontramos un candidato que sea el mejor en todos los aspectos el problema de decisión es trivial: por aplicación del Respeto de la Unanimidad éste será el candidato elegido. Pero no olvidemos que nuestro objetivo es construir una regla de elección que funcione bien en todas las circunstancias. Por eso hacemos especial hincapié en aquellos casos donde puede haber dificultades.

La compra de una vivienda: La utilización de criterios múltiples

Una interesante ilustración de la teoría de la elección social se refiere a las dificultades que surgen en los problemas de decisión individual cuando queremos usar varios criterios de valoración simultáneamente. Un caso particularmente sencillo y frecuente en nuestras vidas es el relativo a la adquisición de una vivienda. Cuando consideramos qué casa comprar solemos recurrir a la comparación de diferentes aspectos (criterios), como son su ubicación, el precio por metro cuadrado y su diseño, entre otros[3].

Imaginemos que, después de descartar las opciones menos interesantes, nos quedamos con tres viviendas alternativas: A, B y C. La siguiente tabla ilustra cómo ordenamos cada una de estas viviendas posibles con relación a los tres criterios mencionados:

[3] Ahora las «alternativas sociales» son las posibles viviendas que consideramos y los «agentes» son los criterios que tomamos en cuenta para valorarlas.

Cuadro IV.3: Ranking de viviendas según distintos criterios			
Criterios	**Vivienda A**	**Vivienda B**	**Vivienda C**
Ubicación	1ª	2ª	3ª
Precio por m²	2ª	3ª	1ª
Diseño	3ª	1ª	2ª

El Teorema de Imposibilidad de Arrow anticipa que podemos tener problemas para decidir qué casa comprar. La vivienda A es mejor que B y que C en términos de ubicación, pero peor que ambas en diseño y también peor que C en precio por m^2. La vivienda C es la que mejor relación precio por m^2 ofrece, pero es la peor en términos de ubicación. Y la vivienda B es la mejor opción en términos de diseño, la peor en términos de precio por metro cuadrado y ocupa una posición intermedia desde el punto de vista de la ubicación.

Recurramos, como en el caso anterior, a asignar puntos a las diferentes valoraciones. La siguiente tabla presenta un ejemplo de este tipo de valoración. En ella establecemos que la vivienda A es el doble de buena que B y seis veces mejor que C en términos de ubicación. Que la vivienda C es dos veces mejor que A y cuatro veces mejor que B en términos de precio. Y, por último, que la vivienda B es algo más de dos veces mejor que C y algo más de tres veces mejor que A en términos de diseño.

Cuadro IV.4: Valoración de las viviendas según distintos criterios			
Criterios	**Vivienda A**	**Vivienda B**	**Vivienda C**
Ubicación	12	6	2
Precio por m²	10	5	20
Diseño	3	10	4
Total	**25**	**21**	**26**

Si aplicamos aquí un criterio de valoración de tipo utilitarista concluimos que debemos comprar la vivienda C, que es la que mayor puntuación global alcanza, quedando A como segunda opción y B como la alternativa peor.

Pero en este caso también tendría sentido considerar el criterio de elección Leximin. La idea sería que queremos escoger aquella vivienda que alcance el mayor valor posible en aquel aspecto en que resulta menos favorecida. Se trata de una estrategia «conservadora» de minimización de riesgos que consiste en aplicar el principio de escoger aquella alternativa en la que *lo peor esté lo mejor posible*. En tal caso escogeríamos la vivienda B que es la que presenta una puntuación mayor en el aspecto menos valorado.

Medición de la desigualdad

Ilustraremos ahora el uso de las funciones de bienestar social en la valoración de la desigualdad en la distribución de la renta. Sea $\boldsymbol{y} = (y_1, y_2, ..., y_m)$ un vector que describe la distribución de la renta en una sociedad compuesta por m individuos. Llamemos μ a la renta media (o renta per capita), es decir,

$$\mu = \frac{\sum_{i=1}^{m} y_i}{m}$$

Sea $u_i(y)$ la utilidad del individuo i-ésimo en la distribución $\boldsymbol{y}$, y sea $F(y) = F[\,u_1(y), \ldots, u_m(y)]$ una función de bienestar social que valora las distribuciones de renta $\boldsymbol{y}$ a partir de las utilidades que proporcionan a los ciudadanos de esa sociedad. Sea pues $F(y)$ el número que describe la utilidad social de la distribución de renta $\boldsymbol{y}$. Imaginemos que ahora queremos alcanzar ese nivel de utilidad social mediante una distribución de renta absolutamente igualitaria, es decir, en la que todos los individuos tienen exactamente la misma renta. Definimos la ***renta igualitaria equivalente*** y^E como aquel nivel de renta que genera el mismo bienestar que la distribución de renta efectiva $\boldsymbol{y}$, cuando todos los individuos de la sociedad tienen exactamente la misma renta. Más formalmente, y^E es aquella renta que verifica:

$$F(y^E, y^E, ..., y^E) = F(y)$$

Donde $F(y^E, y^E, ..., y^E)$ describe la valoración social de una distribución en la que todos los individuos tienen exactamente la misma renta.

La idea de que la desigualdad en la distribución de la renta es un «mal social» se traduce en el supuesto de que la renta igualitaria equivalente es siempre menor que la renta per capita, salvo en el caso de perfecta igualdad. Es decir, $y^E < \mu$, con $y^E = \mu$ si y sólo si $y_i = \mu$ para todo i. Ello equivale a decir que la distribución igualitaria es la mejor posible y que la cantidad de renta necesaria para alcanzar el nivel de bienestar social $F(y)$ es:

$$my^E < \sum_{i=1}^{m} y_i = m\mu$$

Atkinson (1970), siguiendo una idea previa de Dalton (1920), sugiere medir la desigualdad asociada a una distribución como sigue:

$$A(y) = 1 - \frac{y^E}{\mu}$$

$A(y)$ es la desigualdad de la distribución $\boldsymbol{y}$. La fracción y^E/μ mide la relación entre la renta igualitaria equivalente y la media de la distribución. Observemos que:

$$0 \le \frac{y^E}{\mu} \le 1$$

de modo que cuando $y_i = \mu$ para todo i tendremos $A(y) = 0$. Esta fórmula nos dice que la desigualdad asociada a la distribución y podemos medirla en términos de la pérdida de bienestar debida a la dispersión de la renta. La fórmula concreta que adopta el índice de desigualdad depende del tipo de funciones de utilidad que supongamos y del criterio de agregación de las mismas, lo que se traduce en la expresión concreta que adopta y^E.

IV.3 Reconsideración de la racionalidad colectiva

«Preferencialismo»

El ámbito de la teoría de la elección social es muy amplio y abarca problemas muy diversos (cómo repartir un pastel, cómo elegir un alcalde, cómo decidir qué servicios públicos debe proporcionar el Estado, etc.). El proyecto de concebir un procedimiento de decisión que sea válido para problemas tan diversos es, por tanto, enormemente ambicioso. No debería sorprendernos demasiado averiguar que no hay muchas reglas de elección colectiva de aplicabilidad tan universal. El Teorema de Imposibilidad de Arrow muestra que no es posible obtener una valoración de las alternativas sociales mediante una relación de preferencias similar a las preferencias individuales. Pero no es ésta la única dificultad que presenta el enfoque de Arrow para el análisis de los problemas de decisión colectiva. Nos ocuparemos ahora de discutir otras dificultades asociadas a lo que podemos denominar «preferencialismo», que se refieren al uso exclusivo de las preferencias individuales como base de la valoración colectiva.

Tomar las preferencias individuales como variables de la función de bienestar social y asumir los principios de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional, tiene implicaciones relevantes sobre el tipo de información que podemos admitir en la valoración de las alternativas sociales. En particular, con este planteamiento hay dos tipos de información que no puede utilizarse, a saber:

(i) Intensidades de preferencia (cuánto más me gusta una alternativa que otra) o comparaciones interpersonales de satisfacción (cómo está un individuo en relación con otro según la alternativa social elegida).

(ii) La información no contenida en las preferencias individuales que puede referirse a aspectos como la naturaleza del problema, los derechos u obligaciones de los individuos, etc.

La imposibilidad de usar información relativa a las intensidades de preferencia o a la situación relativa de los individuos se deriva directamente de la utilización de las ordenaciones individuales como a argumento exclusivo de la valoración social. Si lo único que nos dicen los agentes es cómo ordenan las alternativas, no podemos inferir nada acerca de la intensidad de sus preferencias (juicios del tipo: «el individuo i prefiere la alternativa x a la alternativa y, pero prefiere *mucho más* la alternativa z»). Ni tampoco relaciones sobre la valoración relativa de los individuos (comparaciones del tipo «el individuo i prefiere la alternativa x a la alternativa y más que el individuo j»). Como hemos visto son precisamente este tipo de comparaciones las que nos permiten escapar al Teorema de Arrow.

La imposibilidad de usar información contextual en la valoración social, es decir, que la regla de decisión pueda depender del tipo de problema considerado y del tipo de sociedad involucrada, es una propiedad conocida como **Neutralidad**. Esta propiedad deriva de la conjunción de los principios de Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional, y resulta independiente de la posibilidad de efectuar comparaciones interpersonales de utilidad.

Para entender exactamente el alcance de esta propiedad, consideremos un problema de decisión colectiva caracterizado por un conjunto (finito) X de alternativas sociales, un conjunto M de m agentes y una regla F de elección colectiva que verifica un conjunto de propiedades entre las que se incluyen Dominio Universal, Respeto de la Unanimidad y Eficiencia Informacional. Sea ahora X' un conjunto diferente pero con el mismo número de alternativas y M' un conjunto de m agentes distintos del caso anterior. Sea F' una regla de elección colectiva asociada a este nuevo problema que verifica también las propiedades anteriores. La Neutralidad establece que la forma de ordenar las alternativas en ambos casos debe ser la misma, es decir, que F y F' deben coincidir.

Dos observaciones importantes con respecto a esta propiedad:

1. La Neutralidad *no* es una buena propiedad que «deseamos» que cumpla la regla de elección social. Se trata de una propiedad *derivada* de otros principios que sí consideramos «deseables».
2. La propiedad de Neutralidad es independiente de la consideración o no de comparaciones interpersonales de utilidad. Por tanto subsiste aun cuando consigamos agregar democráticamente las preferencias.

El Preferencialismo implica que el criterio de elección social resulta insensible ante aspectos como la naturaleza de las alternativas, las motivaciones de las preferencias o los posibles derechos o necesidades de los agentes involucrados. Cuando estos aspectos sean relevantes en un problema de elección colectiva podemos anticipar que encontraremos dificultades adicionales.

Implicaciones del Preferencialismo

Presentamos a continuación una serie de ejemplos que ilustran las implicaciones indeseables del Preferencialismo en distintos contextos.

La valoración de políticas redistributivas

Consideremos en primer lugar un problema de evaluación social de distribuciones de renta y supongamos que la valoración social se basa únicamente en las ordenaciones individuales, sin que puedan establecerse comparaciones interpersonales de utilidad ni intensidades de satisfacción[4].

El problema de valoración social con que nos enfrentamos consiste en la evaluación de medidas de política redistributiva en una sociedad compuestas por tres personas en dos situaciones diferentes. Suponemos que las preferencias de cada individuo son monótonas en la variable renta (es decir, cada individuo estima que cuanto más renta mejor). Observemos los datos de las dos tablas siguientes:

Cuadro IV.5: Distribución porcentual de la renta

Caso A	Situación inicial	Redistribución
Persona 1	90	80
Persona 2	5	10
Persona 3	5	10

Caso B	Situación inicial	Redistribución
Persona 1	3	1
Persona 2	48,5	49,5
Persona 3	48,5	49,5

En el caso A la persona 1 es muy rica (tiene el 90 % de la renta) mientras que las personas 2 y 3 son muy pobres. La política redistributiva consiste en transferir una parte de la riqueza del individuo 1 a los individuos 2 y 3. En el caso B las personas 2 y 3 son ricas mientras que la persona 1 es extremadamente pobre. La política redistributiva consiste en quitarle al pobre la mayor parte de lo que tiene y repartirlo entre los ricos.

Argumentaremos a continuación que el enfoque de Arrow obliga a valorar las políticas redistributivas en A y B de la misma forma. ¿Por qué? Supongamos que queremos decir que la redistribución está más justificada en A que en B. Esta conclusión podría derivarse del hecho de que la persona pobre está «mucho peor» en la situación B que en la situación A, mientras que las personas 2 y 3 están «mejor». Pero observemos que la idea de «mucho mejor» carece de sentido aquí ya que no podemos medir intensidades de preferencia. Por tanto sólo podemos decir que, en la situación B, con la política redistributiva la persona 1 está peor y las personas 2 y 3 mejor. Tampoco podemos interpretar que la persona 1 «está peor» con relación a las personas 2 y 3, ya que estaríamos introduciendo comparaciones interpersonales de satisfacción, que no son admisibles. Y si por «estar peor» aludimos al hecho obvio de que la persona 1 es más pobre en B que en A y la distribución de la riqueza más desigual, estamos incorporando información no contenida en las preferencias individuales, lo que tampoco es posible en el enfoque de Arrow. Por tanto, hemos de concluir que ambos casos resultan informativamente indistinguibles desde el punto de vista de la valoración social, de modo que si en el caso A estimamos que la redistribución es aconsejable lo mismo debemos concluir en el caso B. En particular si el método de decisión colectiva es el de votación por mayoría, que en este caso no presenta problemas puesto que las preferencias son unimodales, la conclusión es inmediata: Tanto en A como en B la alternativa socialmente preferida es la de la redistribución.

[4] Los problemas distributivos son problemas de decisión colectiva en los que se evidencia muy especialmente las limitaciones del Preferencialismo en la valoración social.

La naturaleza de las alternativas

El ejemplo del «ángel caído» que presentamos a continuación [Sen (1979, p. 477)] ilustra cómo al ignorar la naturaleza de las alternativas, como resultado inevitable de la Neutralidad, podemos llegar a resultados grotescos.

Consideremos una sociedad de dos individuos y dos alternativas sociales. La siguiente tabla nos da las correspondientes utilidades (que ahora supondremos comparables interpersonalmente tanto en niveles como en unidades, de modo que podemos medir el bienestar social ya sea mediante el criterio Leximin, ya sea como la suma de las utilidades):

Cuadro IV. 6

	Alternativa 1	Alternativa 2
Individuo *r*	10	8
Individuo *p*	4	7
Utilidad total	14	15

Supongamos primero que el individuo ***r*** es una persona rica y el individuo ***p*** es una persona pobre, y que la alternativa 1 representa la situación inicial y la alternativa 2 refleja el resultado de una política redistributiva. Como 15 > 14 concluimos que la alternativa 2 es socialmente preferida a la alternativa 1 (adviértase que si usamos el criterio Leximin la alternativa 2 sigue siendo socialmente preferida).

Supongamos ahora que la misma matriz de utilidad corresponde a una situación completamente distinta. Sea ahora ***r*** el conductor de una moto –dichoso, rico, ágil y rebosante de salud–, mientras que ***p*** es un peatón –malhumorado, pobre, frustrado y achacoso–. En la alternativa 1 el individuo ***r*** se pasea alegremente con su moto; en la alternativa 2 el motorista cae inadvertidamente en una zanja, destrozando la moto y magullándose. El conductor de la moto está peor en la situación 2 que en la situación 1, mientras que el peatón, que no ha tenido nada que ver con el accidente, disfruta del desconcierto y malestar del motorista

«¡Podría morirme de risa mirando a ese 'ángel caído'!». En este caso, también deberíamos concluir que la alternativa 2 es socialmente preferida.

Derechos: La paradoja liberal

Consideremos ahora que nuestro problema de elección se refiere a un conjunto X en el que se incorporan algunas alternativas que se consideran como *cuestiones privadas* de los individuos. Parece razonable suponer que en los aspectos puramente privados la regla de elección social debiera respetar la decisión de los individuos afectados. Podemos denominar «Liberalismo» a este principio que implica el reconocimiento de derechos individuales sobre determinados aspectos. La ***Paradoja Liberal*** es un resultado que establece que, bajo Dominio Universal, el Liberalismo y el Respeto a la Unanimidad son incompatibles.

Para ilustrar la naturaleza de este resultado consideremos el siguiente ejemplo [una versión libre del propuesto por Sen (1970, págs. 80-81)]. Hay una revista ilustrada (digamos el *Play Boy*) que puede ser «leída» por el señor A (el «devoto»), por el señor B (el «libertino») o por ninguno de los dos. Fijadas todas las demás variables relevantes en el problema de elección social, sean x, y, z estas tres alternativas. Supongamos ahora que el señor A prefiere en primer lugar la alternativa z (que nadie lea esa revista), en segundo lugar la alternativa x («si alguien tiene que leer esa revista debo ser yo, que poseo una mayor integridad moral»), y finalmente la alternativa y («hay que combatir el vicio»). El señor B, por su parte, prefiere en primer lugar la alternativa x («será un buen shock para ese meapilas»), en segundo lugar la alternativa y («es una pena que nadie disfrute de esa revista»), y finalmente la alternativa z. El Respeto de la Unanimidad establece que x es socialmente mejor que y, y sin embargo el individuo A no quiere leer la revista y el individuo B sí.

Este resultado viene a cuestionar la validez de la unanimidad en determinadas circunstancias (en particular en presencia de «externalidades»)[5]. Más en general, lo que la Paradoja Liberal señala es que el uso exclusivo de las preferencias individuales para la valoración social, independientemente de su motivación, puede resultar inadecuado cuando existen derechos individuales que deben ser respetados.

La comparación interpersonal de utilidades

Al abandonar el enfoque de Arrow, para buscar vías democráticas y generales de obtener reglas de elección colectiva, hemos pasado a tomar las utilidades de los individuos como los *inputs* de nuestra valoración social. La razón de dicho cambio es que de este modo podemos establecer comparaciones interpersonales de satisfacción. Los resultados de posibilidad obtenidos en este contexto sugieren que las dificultades de la formulación desarrollada por Arrow no derivan tanto de

[5] Las externalidades se refieren aquí a la influencia en la valoración social de la opinión de un individuo sobre un aspecto que pertenece a la esfera privada de otro.

que exija *demasiadas* propiedades a la regla de elección colectiva, sino de que exige *demasiado pocas:* se quiere que la regla de elección sea muy operativa usando muy poca información sobre las utilidades.

Si bien la formulación desarrollada a este respecto es perfectamente coherente desde el punto de vista formal, la interpretación de las comparaciones interpersonales de utilidad no está exenta de dificultades. ¿Qué debemos entender cuando decimos que el individuo i está mejor con la alternativa x que el individuo j con la alternativa y? ¿Qué significa que todos los individuos midan su utilidad en las mismas unidades?

Aunque hay diversas propuestas sobre la forma de interpretar estas comparaciones interpersonales de utilidad, sugerimos aquí una forma particular: las comparaciones interpersonales son efectuadas por un agente externo que tiene la consideración de un «juez imparcial». La idea básica que hay detrás de esta interpretación es que el bienestar social no sólo debe tomar en cuenta las preferencias individuales, cualesquiera que sean las características personales y la posición social de estos individuos, sino que debe evaluar conjuntamente estas circunstancias. Así, haciendo de la necesidad virtud, podemos contrarrestar los efectos indeseables del preferencialismo.

Para precisar el tipo de interpretación que proponemos, sea X un conjunto de alternativas sociales y sea $\Omega(X)$ el conjunto de características personales y sociales que resultan relevantes para el problema de valoración asociado a X. Este conjunto cambia obviamente de un problema a otro. Sea θ_i el vector de características del individuo i-ésimo. Entonces, para todo x de X, y para todo individuo $i = 1, 2, \ldots, m$, podemos tomar

$$v_i(x) = V(x, \theta_i)$$

donde v_i es la función de utilidad que asignamos al individuo i, y V es una función que asocia valores a cada par de elementos de la forma (x, θ_i) del conjunto $X \times \Omega(X)$. Esta función V refleja la valoración de este «juez imparcial» sobre las circunstancias del problema. La imparcialidad implica, entre otras cosas, el respeto a las preferencias individuales, es decir, para todo par de alternativas $\{x, y\}$ de X, y para todo individuo i, debe cumplirse que

$$V(x, \theta_i) > V(y, \theta_i) \Leftrightarrow u_i(x) > u_i(y)$$

Donde u_i es la función de utilidad del individuo i que describe sus preferencias convencionales.

Nótese que esta interpretación de la comparabilidad interpersonal puede ser operativa sin ser arbitraria, en la medida que se expliciten los juicios de valor empleados y se deduzcan de ellos los criterios de comparación[6]. Una ventaja de esta formulación es que puede anular los inconvenientes derivados de la

[6] Ésta es precisamente la forma de proceder empleada en el diseño de reglas de elección colectiva, donde el criterio de valoración social dependía de los criterios de agregación asumidos.

Neutralidad: puesto que cada conjunto de características podemos escogerlo en función de X, la Neutralidad puede convertirse en una propiedad operativa sin mayores inconvenientes.

A modo de ilustración consideremos los siguientes ejemplos relativos a un subconjunto finito X de alternativas sociales. Para cada agente i llamamos u_i^{min}, u_i^{max} al menor y mayor valor, respectivamente, de la función de utilidad de i sobre el conjunto de alternativas X.

Ejemplo 1:

Para cada individuo i definimos $\theta = (u_i^{min}, u_i^{max})$ y tomamos:

$$V(x, \theta_i) = \frac{u_i(x) - u_i^{min}}{u_i^{max} - u_i^{min}}$$

La función V nos da la ganancia relativa de utilidad del individuo i-ésimo, en términos de sus propias preferencias. Adviértase que siempre que u_i sea una función de utilidad cardinal, este ratio es independiente de la representación de las preferencias.

Ejemplo 2:

Para cada individuo i definimos $\theta_i = u_i^{min}$ y tomamos:

$$V(x, \theta_i) = u_i(x) - u_i^{min}$$

La función V nos dice ahora cuál es la ganancia en de utilidad de la alternativa x con respecto a la peor opción.

El significado de la función de bienestar social en ambos ejemplos es claro y su interpretación no presenta ambigüedad. La comparabilidad interpersonal se ha introducido en el primer caso al tomar como valores relevantes las *diferencias relativas* de bienestar, y en el segundo al tomar las *diferencias absolutas*. En ambos casos estas diferencias se miden de acuerdo con las propias valoraciones de los individuos (es decir, en términos de sus propias funciones de utilidad).

Bajo ciertas condiciones razonables puede comprobarse que las funciones V del ejemplo 1 permiten comparar tanto niveles de utilidad como unidades. Por tanto podemos elegir entre funciones de bienestar social de tipo leximin o de tipo utilitarista.

Por su parte, las funciones V del ejemplo 2 no permiten la comparación de niveles de utilidad, pero sí la comparación de unidades. Por tanto, únicamente podemos aplicar aquí la valoración social de tipo utilitarista.

Consecuentemente:

(i) La función de bienestar leximin aplicada a las funciones V del ejemplo 1 nos indica que el bienestar social se maximiza buscando la igualación de las ganancias relativas para todos los individuos[7]. La función de tipo utilitarista establece que el máximo bienestar social lo obtendremos al maximizar la suma de las ganancias relativas.

(ii) La función de bienestar utilitarista aplicada al ejemplo 2 nos dice que x es socialmente preferida a y si la suma de las ganancias de utilidad es mayor en x que en y[8].

¿Qué hemos aprendido?

El Teorema de Imposibilidad de Arrow nos dice que si utilizamos únicamente las ordenaciones individuales como argumentos de valoración social, no existe ninguna regla de elección social satisfactoria. Más precisamente, que sólo en contextos muy específicos conseguimos obtener un procedimiento democrático de agregación de preferencias. Dicho procedimiento está muy ligado al método de decisión mayoritaria.

Acabamos de ver, mediante unos sencillos ejemplos, que no es éste el único problema asociado con este enfoque. Aunque consigamos un procedimiento aceptable de agregación de preferencias individuales en preferencias sociales, puede resultar insatisfactorio (un problema al que hemos denominado «preferencialismo»).

Otra conclusión central del análisis precedente es que para conseguir reglas de valoración social democráticas en contextos generales debemos *estar dispuestos a hacer comparaciones interpersonales de utilidad*. Ello supone usar como argumento de la regla de elección social más información de la que proporcionan las meras ordenaciones individuales de alternativas.

Las reglas de elección social no dictatoriales que hemos obtenido cuando las utilidades son comparables interpersonalmente dependían del tipo de comparabilidad admitida. Cuando comparamos simplemente niveles de utilidad entre los individuos, obtenemos el criterio del Leximin como principio de ordenación social. Cuando medimos las utilidades individuales en términos de una unidad común para todos los individuos, obtenemos el utilitarismo clásico, como procedimiento de valoración colectiva.

Ambos procedimiento son democráticos y generan valoraciones sociales transitivas. Pero también es cierto que ambas reglas de elección social verifican la propiedad de Neutralidad, que implica descartar toda la información no contenida en las funciones de utilidad. Por tanto estas reglas de elección social pueden resultar insatisfactorias para aquellos problemas en los que hay información relevante no contenida en las utilidades de los individuos.

[7] La maximización del bienestar social resulta entonces equivalente a la solución de Kalai-Smorodinsky asociada al problema de negociación asociado.

[8] Cuando medimos las ganancias en términos de logaritmos, la maximización del bienestar social resulta equivalente a la solución de Nash del problema de negociación asociado.

Sin embargo podemos usar la comparabilidad interpersonal como una herramienta que nos permite introducir elementos contextuales en los criterios de agregación de las valoraciones individuales en valoraciones sociales. Los ejemplos presentados ilustran que es posible contrapesar el efecto de la Neutralidad mediante una adecuada formulación de la comparabilidad interpersonal y obtener al tiempo un modelo de decisión operativo y no arbitrario.

Señalemos para terminar que hay formas alternativas de abordar problemas de decisión colectiva que son de aplicación en muchos problemas económicos y que no se basan en la construcción de una función de bienestar social. La teoría axiomática de la negociación y de la justicia, la teoría de los juegos cooperativos o los esquemas de reparto de costes, son buenos ejemplos de ello. Estos enfoques se refieren en general a problemas mucho más concretos, pero a cambio proporcionan resultados más precisos.

Nos ocuparemos en los siguientes capítulos de presentar algunos de estos resultados.

CAPÍTULO

V

«All together now...»

LENNON & MCCARTNEY

V. Cooperación

V.1 Introducción

Cooperación simple y cooperación multilateral

Dedicamos este capítulo al estudio de una familia específica de problemas de decisión colectiva cuyo tema común puede describirse como sigue: Un grupo de agentes tiene que decidir cómo se reparten las ganancias derivadas de su cooperación en algún proyecto conjunto. Las ganancias pueden expresarse en términos de utilidades, de cantidades de un cierto bien o de indices que miden la consecución de un determinado objetivo. En ocasiones la decisión colectiva también puede referirse al reparto de pérdidas o a la asignación de costes conjuntos. El punto de partida es que la cooperación puede hacer que el grupo consiga mejores resultados que los que se derivarían de la acción aislada de sus miembros. Lo que hace interesante el problema de reparto es que las ganancias de la cooperación son limitadas, de modo que el hecho de que a partir de cierto punto que uno reciba más significa que algún otro recibirá menos.

Estos problemas presentan rasgos específicos que inducen un tratamiento diferenciado con respecto a los problemas abstractos de elección social analizados en los capítulos anteriores. En particular:

(i) La noción de «distribuir las ganancias de la cooperación» requiere la consideración un punto de referencia desde el que se miden estas ganancias. Ello supone el incumplimiento de la propiedad de *eficiencia informacional* que establecía que la comparación de dos alternativas cualesquiera depende únicamente de las dos alternativas consideradas[1].

(ii) En los problemas distributivos se manifiestan las mayores limitaciones del *preferencialismo* característico de las funciones de bienestar social. Veremos que, en ocasiones, este tratamiento diferenciado posibilita la

[1] Ahora la comparación de dos alternativas depende de su posición relativa con respecto a una tercera (aquella desde la que medimos las ganancias o pérdidas y que juega el papel de *statu quo*).