EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS XX JORNADAS

VOLUMEN 16 (2010)

Pío García
Alba Massolo

Editores

ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES
UNIVERSIDAD NACIONAL DE CÓRDOBA

Decisiones sociales e imposibilidades

*Marcelo Auday**

Introducción

La vida en comunidad supone la interacción social y la acción colectiva y ambas involucran procesos de decisión tanto individuales como grupales. Todos estos fenómenos, objetos básicos de estudio de las ciencias sociales, están atravesados por una mezcla de motivaciones e implican usualmente problemas de coordinación a la vez que posibles conflictos de intereses.

La teoría de la elección racional, en particular la teoría de la elección social y la teoría de juegos se ocupan específicamente de este tipo de problemas. Dos resultados bien conocidos son el teorema de imposibilidad de Kenneth Arrow (Arrow,1951) y la tragedia de los bienes comunes de Garrett Hardin (Hardin, 1968). Mientras que el primero establece una limitación a la agregación de opiniones individuales para establecer un resultado u opinión social, el segundo muestra cómo las decisiones individuales sobre el uso de un bien común pueden terminar destruyendo dicho bien.

La politóloga estadounidense Elinor Ostrom es la primera mujer en ganar el premio Nobel en economía; dicho premio se le ha otorgado fundamentalmente por sus investigaciones orientadas al esclarecimiento de la tragedia de los bienes comunes, las condiciones bajo las cuáles es probable que ésta surja y el análisis de las distintas formas de resolver dicho problema. En particular, Ostrom ha mostrado cómo la dicotomía usual, establecida entre soluciones privadas (redefinición de derechos de propiedad) y la solución estatal, ignora un tercer tipo de solución, a saber, el control comunitario. Una de las características salientes del trabajo de Ostrom es la de combinar en sus investigaciones la construcción de modelos con el estudio de casos y el diseño y realización de experimentos.

En su artículo en homenaje a Amartya Sen, Ostrom (2009) establece un paralelismo entre los resultados de imposibilidad de Arrow y de Hardin, a la vez que un paralelismo entre el trabajo de Sen y su propio trabajo. Sen, en su conferencia del Nobel (Sen,1999) expuso el teorema de Arrow para mostrar que no puede soslayarse pero tampoco debe asumirse sin más. Parte del análisis de Sen ha sido estudiar detalladamente los supuestos del teorema y relacionarlos con investigaciones empíricas sobre procesos de decisión colectiva para establecer con claridad cuándo es posible que se produzcan las incongruencias y cuándo la decisión social funciona adecuadamente.

* Centro de lógica y filosofía de la ciencia Universidad Nacional del Sur, ccauday@criba.edu.ar

Ostrom (2009) hace lo mismo en relación con la tragedia de los bienes comunes resumiendo los principales aportes realizados y señalando los problemas que quedan por resolver.

En lo que sigue, luego de presentar el teorema de Arrow y la tragedia de los bienes comunes de Hardin, desarrollamos con mayor detalle el paralelismo articulado por Ostrom, estableciendo una conexión formal entre ambos resultados bajo el supuesto, bastante usual, de que la tragedia de los bienes comunes puede modelarse como un dilema del prisionero. Por cuestiones de simplicidad lo hacemos solamente para el caso de dos jugadores. Finalmente, hacemos unas aclaraciones sobre las motivaciones subyacentes a las preferencias individuales supuestas en el dilema del prisionero y en el teorema de Arrow.

El teorema de Arrow

La paradoja de Condorcet muestra que la regla de mayoría puede generar preferencias no transitivas a partir de preferencias individuales transitivas. Supongamos que tres individuos tienen que establecer un ranking entre tres opciones (a, b y c) y para ello usarán la regla de mayoría, supongamos que las preferencias individuales son las siguientes: aP_1bP_1c, bP_2cP_2a, cP_3aP_3b.[1] Hay entonces una mayoría a favor de aPb (individuos 1 y 3), otra a favor de bPc (1 y 2) y otra a favor de cPa (2 y 3).

Visto este problema de la regla de mayoría es natural preguntarse si otras reglas de votación también presentarán este problema o, más generalmente, qué otras limitaciones pueden tener dichas reglas. Kenneth Arrow, en su tesis doctoral (Arrow, 1951) estudia de manera sistemática los mecanismos de agregación de preferencias, en particular, de la agregación de preferencias individuales completas y transitivas para producir preferencias sociales que también sean completas y transitivas[2]. Arrow postula un conjunto de condiciones que considera toda función de bienestar social debe cumplir y demostra que, de hecho, ninguna puede satisfacerlas simultáneamente. Específicamente, el teorema de imposibilidad de Arrow establece que ninguna función de bienestar social puede cumplir a la vez las condiciones de dominio irrestricto, principio débil de Pareto, independencia de alternativas irrelevantes y no dictadura, si hay al menos dos votantes y tres opciones.

La condición de dominio irrestricto establece que cualquier preferencia individual (completa y transitiva) sobre las opciones es admisible. El principio débil de Pareto establece que si todos los votantes consideran que la alternativa x es estrictamente mejor que la alternativa z, entonces la sociedad prefiere estrictamente x a z, es decir, la regla de decisión debe respetar la unanimidad. La condición de independencia establece que la preferencia social sobre cualquier par de alternativas $\{x,z\}$ depende sólo de las preferencias individuales sobre ese par. En particular, si dos n-tuplas de preferencias individuales $(R_1, \dots R_n)$, $(Q_1, \dots, Q_n)$ son iguales respecto del par de opciones $\{x,z\}$, entonces ambas deben producir la misma preferencia social respecto de dicho par. La condición de no-dictadura establece que ningún individuo puede imponer siempre sus preferencias, esto es, no hay un individuo tal que en toda n-tupla de preferencias individuales y sobre todo par de alternativas $\{x; z\}$ cada vez que él prefiere estrictamente x a z la sociedad prefiere estrictamente x a z.

La tragedia de los bienes comunes

La tragedia de los bienes comunes representa una situación de interacción social en la cual un grupo de individuos, que comparten un recurso común limitado, actuando cada uno de ellos de manera independiente y guiados por sus preferencias individuales, terminan por sobreexplotar el recurso y destruirlo. Hardin (1968) no da una representación formal de tal situación sino que construye un ejemplo: hay una pradera accesible a cualquiera, es decir cualquier pastor puede llevar a la misma la cantidad de ganado que desee para que pasten allí, se asume, además, que los pastores maximizan beneficios. La causa de la tragedia (es decir, de la sobreexplotación y eventual destrucción de la pradera) yace en que cada pastor tiene incentivos a enviar el mayor número posible de animales a la pradera, y esto es porque mientras los beneficios obtenidos de la venta de un animal son capturados únicamente por el dueño de dicho animal, los costos generados por el animal (en el ejemplo, su alimentación) se distribuyen entre todos los usuarios de la pradera.

Si bien, como dijimos, Hardin no da una representación formal, usualmente se representa la tragedia como un dilema del prisionero. Seguimos aquí el análisis dado por Arce & Sandler (2005) para el caso de dos pastores: cada uno de ellos puede llevar un animal a pastar (acción) o ninguno (inacción). Si un pastor lleva un animal a pastar obtiene un beneficio privado b y genera un costo público C para ambos pastores. La matriz del juego es la siguiente.

		Pastor 2			
		Acción		Inacción	
Pastor 1	Acción	b-2C,	b-2C	b-C,	-C
	Inacción	-C,	b-C	0,	0

Para que dicha matriz represente un dilema del prisionero se requiere que $2C > b > C$. Así, $b-C>0$ y $b-2C>-C$, por lo cual para cada pastor llevar un animal a pastar (acción) es una estrategia dominante y terminan produciendo un resultado ineficiente: que ambos lleven un animal a pastar (acción, acción) (único equilibrio de Nash del juego) produce un resultado inferior, para cada uno de ellos, al que obtendrían si ninguno de los dos lleva animales a pastar (inacción, inacción).[3]

El dilema del prisionero y el teorema de Arrow

Si aceptamos que la tragedia de los bienes comunes queda apropiadamente representada mediante el dilema del prisionero con 2 o más jugadores, podemos tratar de establecer una conexión más estrecha entre la tragedia y el teorema de Arrow vinculando a éste último con el dilema del prisionero. Rusciano (1990) ofrece una forma de establecer dicha conexión. En particular, sostiene que el dilema del prisionero es un caso particular del teorema de Arrow. Ahora bien, dado que son dos estructuras formales distintas (una un juego en forma normal, la otra una función de agregación), es necesario establecer una traducción de una a la otra. Para esto, lo que Rusciano hace es transformar el dilema del prisionero en una situación de decisión colectiva o votación: hay un único perfil de preferencias que agregar, constituido por las preferencias de cada uno de los jugadores sobre los resultados del dilema; el objetivo es construir un orden de preferencias social y la regla de agregación es la unanimidad en la siguiente versión: dadas dos opciones cualesquiera *a* y *b*, *a* es estrictamente preferida a *b* si y sólo si ambos jugadores prefieren estrictamente *a* a *b*; de lo contrario, el orden social establece que *a* y *b* son indiferentes. Según Rusciano, dado que no hay medios para obligar a alguien a cooperar, la unanimidad es la única regla justificable en el contexto del dilema. Simbolizando "acción" con "a" e "inacción" con "i" el perfil de preferencias extraído del dilema es el siguiente: jugador 1: (a,i)P(a,a)P(i,i)P(i,a), mientras que las preferencias del jugador 2 son: (i,a)P(a,a)P(i,i)P(a,i). Por unanimidad obtenemos (i,a)I(i,i), (i,i)P(a,a), (a,a)I(i,a) lo cual viola la transitividad[4] Por otra parte, la regla de unanimidad cumple las demás condiciones establecidas en el teorema de Arrow.

Si bien la idea de Rusciano en términos generales es aceptable, no lo es desde un punto de vista más riguroso. Por una parte, lo que Rusciano prueba es que unanimidad puede violar transitividad, no que cualquier regla aplicada a ese perfil específico de preferencias violará alguna de las condiciones de Arrow. Desde luego, puede que Rusciano tenga razón en que la unanimidad sea la única regla aceptable para el contexto bajo consideración, pero el tema central aquí es establecer una conexión directa entre el dilema del prisionero y el teorema de Arrow y no con alguna alguna regla de agregación de preferencias en particular. Por otra parte, el teorema de Arrow original considera más de un perfil de preferencias, mientras que Rusciano sólo considera un único perfil. Ahora bien la literatura ad usum distingue ambos tipos de situaciones y una versión aceptable del teorema de Arrow para el caso de un único perfil de preferencias requiere de modificaciones a las condiciones originales del teorema.

Feldman & Serrano (2008) ofrecen una demostración simple para el teorema de Arrow con un único perfil, tanto para el caso de dos votantes como para el caso de más de dos votantes. Consideramos aquí sólo el primer caso. A las condiciones usuales (preferencias sociales completas y transitivas, principio débil de Pareto[5] y no dictadura) hay que anexarles las condiciones de neutralidad y diversidad simple. Neutralidad es el reemplazo natural de la condición de independencia en un contexto de único perfil.

La condición de *neutralidad* implica que todas las alternativas reciben el mismo trato o, en otras palabras, que el mecanismo de decisión no hace distinciones entre las alternativas. De forma un poco más precisa. sean $\{x,z\}$ y $\{w, p\}$ dos pares cualesquiera de alternativas, y sean dos perfiles cualesquiera de preferencias individuales $(R_1, \ldots R_n)$ y $(Q_1, \ldots, Q_n)$. Entonces, si la información individual sobre $\{x,z\}$ en $(R_1, \ldots R_n)$ es la misma que la información individual sobre $\{w,p\}$ en $(Q_1, \ldots, Q_n)$, la preferencia social sobre $\{x,z\}$ en $(R_1, \ldots R_n)$ tiene que ser la misma que la preferencia social sobre $\{w, p\}$ en $(Q_1, \ldots, Q_n)$ En un contexto de múltiples perfiles neutralidad implica independencia.

En un contexto de único perfil, independencia es irrelevante (se cumple vacuamente); sin embargo, neutralidad sí se aplica: sean $\{x,z\}$ y $\{w, p\}$ dos pares cualesquiera de alternativas, entonces si para cada individuo *i* se cumple que xR_iz sii wR_ip y zR_ix sii pR_iw, entonces las preferencias sociales serán xRz sii wRp y zRx sii pRw.[6] Finalmente, la condición de dominio irrestricto no tiene sentido en el contexto de perfil único, su reemplazo aquí es la condición de diversidad simple: hay al menos tres alternativas *x,z,w*, sobre las cuales las preferencias individuales son las siguientes: todos prefieren estrictamente *x* a *z*, mientras que un votante prefiere estrictamente *x* a *w* y el otro prefiere estrictamente *w* a *x*, y lo mismo sucede sobre el par {z,w}.

El punto importante es que el perfil de preferencias generado por el dilema del prisionero cumple la condición de diversidad simple: sea x =(i,i), z=(a,a) y w=(a,i) o w = (i,a). Consideremos entonces cómo agregar estas preferencias indiduales mediante una regla que cumpla las condiciones arriba mencionadas. Por el principio débil de Pareto, obtenemos (i,i)P(a,a) Las preferencias individuales sobre todos los demás pares son opuestas. Si (a,i)P(i,i), por neutralidad (a,a)P(i,a) y el jugador 1 es un dictador Si (i,i)P(a,i), por neutralidad (i,a)P(i,i) y (a,a)P(a,i) y el jugador 2 es un dictador. Finalmente, si (i,i)I(a,i), por neutralidad (a,i)I(a,a) lo cual viola la transitividad.

Motivaciones subyacentes

Bajo la interpretación usual del dilema del prisionero, lo que causa semejante resultado socialmente negativo (el único equilibrio de Nash es pareto ineficiente) es que los individuos involucrados, al tomar sus decisiones, sólo tienen en cuenta sus costos y beneficios personales y no los costos y beneficios sociales. En términos más generales, se dice que tal resultado es producidos por una estructura de preferencias egoístas.

Por su parte, Ostrom sostiene que el teorema de Arrow también presupone preferencias de este tipo: "Arrow uses the definition of rationality accepted by most economists in the midtwentieth century He assumed that individuals making a social choice had well-defined, complete and transitive preferences regarding the likely net personal returns given the set of alternatives under consideration." (Ostrom, 2009, p. 522)

Ahora bien, la teoría de la elección social en general y el teorema de Arrow en particular sí presuponen racionalidad pero no presuponen que dicha racionalidad implique una motivación egoísta, "racionalidad" en este contexto significa simplemente que las preferencias son completas y transitivas. Nada se estipula acerca del criterio o los criterios mediante los cuales se genera el orden de preferencias de un individuo. A modo de ejemplo, consideremos la situación en que un jurado integrado por tres personas tiene que establecer un ranking entre tres candidatos (a,b,c) para un puesto en una institución. El jurado 1, en base a su evaluación de la inteligencia de los candidatos establece el ranking aPbPc, por su parte, el jurado 2 establece bPcPa (en base al criterio de responsabilidad), finalmente, el jurado 3 establece cPaPc (en base al criterio de liderazgo). Es obvio que si a este perfil de preferencias le aplicamos la regla de mayoría obtenemos la paradoja de Condorcet (que podemos verla como una instanciación del teorema de imposibilidad de Arrow). Ahora bien, ninguno de los rankings mencionados implica, en algún sentido natural, preferencias egoístas por parte de los jurados. Lo que hay es una evaluación diferente en virtud de la utilización de criterios diferentes, y tales criterios no implican en principio ningún costo y/o beneficio para los jurados. El ejemplo puede reforzarse aún más suponiendo que, de hecho, los tres jurados están de acuerdo en cuál es el ranking correcto para cada uno de los criterios mencionados (es decir, para los tres el ranking de los candidatos es aPbPc según la inteligencia, bPcPa según la responsabilidad, y cPaPc según el liderazgo) pero difieren en términos de cuál de los criterios es el adecuado para realizar la evaluación y seleccionar a uno de los candidatos.

Ahora bien, algo similar puede decirse del dilema de los prisioneros. Formalmente, lo que define a una situación como un dilema del prisionero es que la estructura de las preferencias involucradas es tal que en el juego cada jugador tiene una estrategia dominante, hay un único equilibrio de Nash y el mismo es Pareto ineficiente. Que la motivación egoísta pueda generar tal estructura de preferencias no implica que sea el único tipo de motivación que pueda hacerlo. Para finalizar, el siguiente ejemplo, tomado de Osborne (2004), muestra cómo preferencias egoístas pueden generar un resultado socialmente óptimo, mientras que preferencias altruistas pueden generar un dilema del prisionero. Consideremos un ómnibus donde sólo quedan dos asientos adyacentes y sin ocupar Dos individuos tienen que decidir si sentarse o no, para ambos, en términos de confort personal, lo mejor es sentarse sólo, en segundo lugar sentarse acompañado y, finalmente, quedarse parado. En este último caso, donde el jugador no se sienta hay tres opciones: que sea indiferente respecto de que el otro esté sentado o no, que prefiera que el otro se siente, o que prefiera que el otro tampoco se siente En verdad, la estructura del juego es invariante respecto de estas opciones. Consideramos sólo el primer caso (la indiferencia). Supongamos entonces que cada jugador toma decisión sólo considerando su confort personal.

		Jugador 2	
		Sentarse	No sentarse
Jugador 1	Sentarse	3 , 3	4 , 2
	No Sentarse	2 , 4	2 , 2

En este juego, cada jugador tiene una estrategia dominante (sentarse) y hay un único equilibrio de Nash (sentarse, sentarse), que es Pareto eficiente, por lo cual este juego no es un dilema del prisionero.

Consideremos ahora la situación en que cada jugador toma su decisión en base al confort personal del otro jugador y además por cortesía prefiere estar parado si el otro jugador también lo está.

		Jugador 2	
		Sentarse	No sentarse
Jugador 1	Sentarse	3 ,3	1, 4
	No Sentarse	4, 1	2 , 2

En este juego, cada jugador tiene una estrategia dominante (no sentarse) y hay un único equilibrio de Nash (no sentarse, no sentarse) el cual es Pareto ineficiente. Tenemos aquí sí un dilema del prisionero.

Aunque el ejemplo es simple, alcanza para mostrar que la existencia de situaciones con la estructura del dilema del prisionero no depende de un determinado tipo de motivación.

Conclusiones

Ostrom (2009) establece un paralelismo entre la tragedia de los bienes communes y el teorema de imposibilidad de Arrow. En este trabajo, profundizamos dicho paralelismo estableciendo una conexión formal entre ambos resultados bajo el supuesto de que la tragedia de los bienes comunes puede representarse razonablemente como un dilema del prisionero. Secundariamente, corregimos una apreciación de Ostrom respecto de qué tipo de preferencias están supuestas en el teorema de Arrow. Mostramos brevemente que ni este teorema ni el dilema del prisionero requieren que la motivación subyacente sea el egoísmo.

Dado que la conexión formal mencionada arriba entre el dilema del prisionero y el teorema de Arrow sólo la hemos establecido para el caso de dos individuos, queda como tarea extender esto para el caso de más de dos individuos. Por otra parte, nuestro análisis ha tomado como punto de partida el supuesto bastante extendido de que la tragedia de los bienes comunes puede modelarse como un dilema del prisionero. Más allá de su razonabilidad esto no está exento de críticas7. Parte del trabajo de Elinor Ostrom8 ha sido tratar de mostrar bajo qué condiciones la gestión de un bien común puede presentar las características de un dilema del prisionero y bajo cuáles no.

Notas

1 Los subíndices distinguen a los individuos. "aP_1bP_1c" abrevia aP_1b, bP_1c y, como son transitivas, aP_1c. "aP_1b" significa "el individuo 1 prefiere estrictamente la opción *a* a la opción *b*".

2 Este tipo de funciones se denominan usualmente *funciones de bienestar social*.

3 Además de sólo considerar dos pastores, la otra simplificación es que cada pastor sólo puede llevar un único animal. Una forma de interpretar esto es que se está considerando la situación límite en la cual agregar dos animales más a la pradera produce la sobreexplotación, es decir, podemos suponer que ya hay un número determinado de animales en la misma.

4 "P" representa la relación de preferencia estricta (o la parte asimétricade las preferencias) e "I" la relación de indiferencia (o la parte simétrica de las preferencias). Que las preferencias de un individuo sean transitivas implica que tanto P como I son transitivas y, además, que si xPz y zIw, entonces xPw, y si xIz y zPw, entonces xPw.

5 Feldman & Serrano (2008) usan, en verdad, el principio fuerte de Pareto, pero dado que aquí consideramos un perfil de preferencias estrictas, puede reemplazarse este principio por su versión débil.

6 Un ejemplo puede ser aclaratorio: supongamos que hay sólo dos individuos, que las preferencias sobre el par de opciones {x,z} son xP_1z, zP_2x y que la regla de agregación produjo el resultado social xPz. Entonces, si dicha regla es *neutral*, para cualquier otro par de opciones {w,p} tal que las preferencias individuales sean wP_1p, pP_2w, el resultado social tiene que ser wPp.

7 Para algunas críticas a este supuesto ver Kuhn (2009)

8 Ver, por ejemplo, Ostrom (2007)

Referencias

Arce, D. & Sandler, T. (2005) "The Dilemma of the Prisoners' Dilemma", Kyklos, vol. 58, nro 1, 3-24

Arrow, K. (1951) Social Choice and Individual Values, Wiley, New York.

Feldman & Serrano (2008) "Arrow's Impossibility theorem: two simple single-profile versions" [en línea] Brown University Department of Economics Working Paper No. 2006-11. Disponible en SSRN. http://ssrn.com/abstract=896742 (consulta. 21-06-10)

Hardin, G (1968) "The Tragedy of the Commons", *Science*, vol 162: 1243-1248.

Kuhn, S. (2009) "Prisoner's Dilemma", *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, Edward N Zalta (ed.),

http://plato.stanford.edu/archives/spr2009/entries/prisoner-dilemma (consulta: 21-06-10)

Osborne, M. (2004) "An Introduction to game theory", New York. Oxford University Press.

Ostrom, E. (2007) "A diagnostic approach for going beyond panaceas", *PNAS*, vol. 104, nro 39, 15181-15187

Ostrom, E. (2009) "Engaging with Impossibilities and Possibilities", en Basu, K. & Kanbur, R. (eds) "Arguments for a better World" Essays in honor of Amartya Sen, vol II Society, Institutions and Development, New York: Oxford University Press.

Sen, A. (1999) "The Possibility of Social Choice", *American Economic Review* vol. 89, nro 3, 349-378.