1206

IESA

Instituto de Estudios Sociales Avanzados de Andalucía
Consejo Superior de Investigaciones Científicas / JUNTA DE ANDALUCÍA

JUNTA DE ANDALUCIA

Documentos de Trabajo

LIBERTAD, JUSTICIA Y JUEGOS

FERNANDO AGUIAR

CSIC

IESA WORKING PAPER SERIES

www.iesaa.csic.es

1. Introducción

Desde muy pronto, apenas una década después de que John von Neuman y Oskar Morgenstern publicaran su *Theory of Games and Economic Behaviour*, se apreció la utilidad de esta nueva rama de la matemática aplicada para la ética. Si la teoría de juegos se ocupaba del estudio formal de las interacciones estratégicas de individuos racionales, esas interacciones podían tener un claro componente ético (Braithwaite, 1955). Desde muy pronto también, los pioneros de la teoría de juegos dieron un nuevo impulso a viejos problemas filosóficos, éticos y políticos al tiempo que desarrollaban su disciplina. Hoy en día resulta imposible reflexionar sobre racionalidad, utilidad, normas sociales, normas de justicia o cooperación, por poner sólo algunos ejemplos, sin tener en cuenta las aportaciones de la teoría de juegos.

Ahora bien, alabar el uso de la teoría de juegos en ámbitos como la ética y la filosofía política puede hacernos creer que hay un claro acuerdo sobre cómo debe usarse esta herramienta formal y cuáles son sus límites. Sin embargo, esto no es así. Hay al menos tres respuestas distintas a esa cuestión y cada una de ellas establece límites distintos al uso de la teoría de juegos en el terreno de la reflexión ético-política: una concepción puramente instrumental, otra contractualista y una tercera evolutiva.[1] El uso y la concepción puramente instrumentales de la teoría de juegos para la ética es el más restringido. Desde este punto de vista la teoría de juegos no es más que una herramienta formal que, como la lógica, aporta claridad a las complejas situaciones de interacción que implica un problema de decisión moral. Los modelos de la teoría de juegos evidencian cuál es la naturaleza concreta de un problema moral y las teorías éticas proporcionan soluciones a esos problemas. Así, por ejemplo, si la teoría de juegos nos ha llevado mucho más lejos que cualquier otra teoría en la demostración formal de los fallos de la racionalidad, la misión de las teorías morales consistiría, entre otras cosas, en tratar de resolver esos problemas de racionalidad. En el caso paradigmático del dilema del prisionero, en el que la racionalidad individual conduce a un resultado colectivo irracional al impedir que los jugadores cooperen, las teorías morales proporcionan las normas que permiten superar tales dilemas.

[1] Esta división tripartita es una versión modificada de la que se puede encontrar en Verbeek y Morris (2004).

Las aproximaciones a la teoría de juegos en términos que no son meramente instrumentales consideran que las normas morales y los principios de justicia son el producto de la interacción estratégica de individuos racionales. La teoría de juegos no sería ya una herramienta útil para la ética, sino que la ética misma sería teoría de juegos, una rama más de esta disciplina (Harsanyi, 1992: 671). En este terreno nos encontramos sobre todo con teorías contractualistas que "interpretan la moralidad como el resultado de un proceso de negociación" (Verbeek y Morris, 2004: 4). Los jugadores negocian estratégicamente los principios morales y políticos que regularán la distribución de los recursos sociales. Ahora bien, esto no deja de ser un artificio, una construcción hipotética a partir de la cual deducir los normas morales y de justicia. La concepción evolutiva de la relación entre la ética y la teoría de juegos comparte con los contractualistas la idea de que la moralidad emerge de la interacción estratégica de los individuos. Pero ese surgimiento no se establece en un marco hipotético-deductivo, sino en un marco evolutivo. La normas morales son producto de la evolución humana; a lo largo de su historia evolutiva los seres humanos han tenido que resolver complejos problemas de coordinación para sobrevivir. La solución de esos problemas es lo que se denomina normas morales y de justicia. La teoría de juegos evolutiva nos permite reconstruir esas soluciones, nos proporciona los equilibrios evolutivos a los que denominamos moralidad, justicia y libertad.

En este capítulo presentaré algunos ejemplos de los diversos usos de la teoría de juegos en ética y filosofía política. Para ilustrar la aproximación instrumental me centraré en el análisis de la libertad –derechos y libertades y libertad interior- desde la teoría de juegos (secciones 2 y 3). En la cuarta sección veremos el vínculo que existe entre la teoría de la negociación y la teoría del contrato. En la quinta sección me detendré en la teoría de juegos evolutiva y en su explicación del surgimiento de normas de justicia.

2. Elección social, liberalismo y libertad

Cuando Kenneth Arrow publicó en 1951 su obra *Social Choice and Individual Values* (Arrow, 1963) inauguraba una línea de investigación en el seno de la economía del bienestar que rápidamente se extendería a otras disciplinas. Su propósito original residía en hallar un mecanismo (una constitución, una regla de arbitraje, un sistema de votación, etc.) que permitiera tomar decisiones colectivas a partir de las preferencias de cada individuo sobre distintas opciones sociales. Arrow demostró en su trabajo que es imposible encontrar un mecanismo semejante –una función de decisión social- sin que viole una serie de condiciones tan razonables como las siguientes: que nadie imponga a los demás su preferencia, que en el proceso de decisión colectiva tengan cabida todas las preferencias individuales, que las preferencias sean transitivas para que el individuo (y la sociedad) no se contradigan a sí mismos, etc. El análisis del Teorema de Imposibilidad de Arrow (sus consecuencias, los intentos de sortearlo y la aparición de nuevas imposibilidades) ha generado lo que hoy se conoce como teoría de la elección social.[2]

Una de las condiciones que deben cumplir las reglas de elección social tiene que ver con el derecho de las personas a que las decisiones ajenas no violen su esfera privada: si se concede que en cierto ámbito privado la elección individual ha de ser libre, se debe imponer a las decisiones colectivas una condición que asegure tal derecho a la libre elección en asuntos privados. Si yo quiero que las paredes de mi casa sean verdes no puede establecerse una regla de elección social –una sistema de voto, por ejemplo- que contemple la posibilidad de que las pinte de rojo. En el lenguaje de la teoría de la elección social se dice que la persona que determina la elección sobre el par de posibles estados o situaciones sociales (**x**, **y**) en los que se abordan asuntos de su incumbencia es *decisiva* sobre ese par. Para asegurar la decisividad de los individuos Amartya Sen propuso una *condición débil de liberalismo* (condición L) que sirve para proteger la esfera privada de cualquier persona cuando se ve inmersa en un proceso de elección social, esto es, un proceso en el que a partir de las preferencias individuales se debe llegar a una preferencia social o colectiva:

> "La condición L exige que para toda persona haya *al menos* un par de estados sociales, digamos x e y, tales que su preferencia en ese par sea decisiva para el juicio social; esto es, si prefiere x a y, se tiene que reconocer entonces que x es socialmente mejor que y, y de igual modo si prefiere y a x. La aceptabilidad de la condición L dependerá de la naturaleza de las alternativas que se ofrecen para elegir, y si ninguna elección fuera personal...esta condición no tendría mucho atractivo" (Sen, 1982: 292).

[2] Una excelente introducción a los problemas de la elección social se puede encontrar en Sen (1976).

No se trata, según lo entiende Sen, de que tan sólo los liberales en sentido estricto acepten esta condición. El término liberalismo puede mover aquí a confusión. Lo que se pretende destacar es que cualquiera que acepte un mínimo de libertad individual -sea cual fuere su concepción general de la misma- debe estar conforme con que la condición L se integre en el proceso de elección social.

En "La imposibilidad de un liberal paretiano" (Sen, 1970), artículo extensamente comentado desde su aparición, Amartya Sen demostró que la condición L entra en conflicto con el *criterio de optimidad o eficiencia de Pareto*. En su forma débil, este criterio afirma que si todos los miembros de una sociedad prefieren un estado social **x** a otro **y**, el primero se debe considerar socialmente preferible al segundo. El Teorema de Imposibilidad de un liberal paretiano demuestra que no existe ninguna función de decisión social que satisfaga al mismo tiempo la condición liberal (condición L) y el criterio de Pareto. En otras palabras, la condición liberal no es compatible con la eficiencia paretiana, resultado que tiene una enorme trascendencia para le economía del bienestar, que ha hecho del criterio de Pareto su piedra angular.

Merece la pena recordar el ejemplo que propuso Sen para ilustrar su teorema, pues aunque se trata de un ejemplo un tanto anticuado se ha convertido en todo un clásico de la literatura sobre elección social. Puritano y Lascivo tienen un ejemplar de *El amante de Lady Chatterley*, y Puritano quiere evitar que Lascivo lea un libro con fama libertina. Las situaciones sociales posibles en el ejemplo de Sen son las siguientes:

a: Puritano lee el libro
b: Lascivo lee el libro
c: Ninguno de ellos lo lee.

Puritano preferiría ante todo que nadie lo leyera (situación social o resultado final **c**), optando por leerlo él mismo (**a**) si con ello impide que lo lea Lascivo (**b**). Lascivo desearía que lo leyera Puritano (**a**) para ver si así se suaviza su puritanismo. Mas si Puritano se niega a leerlo, Lascivo prefiere leerlo él (situación **b**) antes de que el libro se quede sin leer (**c**). Tenemos, pues, la siguiente ordenación de las preferencias de ambos individuos[3]:

[3] Las opciones que se encuentran más arriba se prefieren a las que se encuentran más abajo.

Puritano	Lascivo
c	**a**
a	**b**
b	**c**

¿Pueden obtener Lascivo y Puritano una sola ordenación social a partir de sus preferencias individuales que cumpla con la Condición L y con el criterio de Pareto? La condición liberal dicta que Lascivo lea el libro antes de que se quede sin leer, ya que así lo desea. Es decir, la sociedad debe respetar ese deseo, por lo que **b** se ha de considerar socialmente mejor que **c** gracias a la condición L, pues Lascivo es decisivo sobre ese par (b, c). Por la misma razón diremos que si Puritano no desea leer el libro de Lawrence, las alternativa **c** resultará socialmente mejor que **a**. Entonces, si **b** es socialmente preferida a **c** (lo que expresamos como bPc) y cPa obtendremos, por transitividad, **bPa**. Pero ocurre que el criterio de Pareto impone **aPb** (pues **a** es preferida a **b** de forma unánime por Lascivo y por Puritano), por lo que surge una contradicción: el criterio de Pareto lleva a la preferencia social **aPb** y la condición L a la preferencia social **bPa**. La condición liberal conduce, pues, a un resultado social supóptimo.

Al calor de la Paradoja de Sen surgieron otras muchas paradojas y otros tantos intentos de resolverlas. Una de las más interesantes y discutidas, relacionada también con la libertad de elección, es la Paradoja de Gibbard. En esta paradoja se demuestra que no sólo la Condición L puede entrar en conflicto con el criterio de Pareto, como demostró Sen, sino que también los derechos y libertades individuales -tal y como se plasman en dicha condición- pueden dar lugar a resultados incoherentes entre sí. En el teorema de imposibilidad de Sen se establecía la decisividad de una persona sobre un par de alternativas sociales que sólo le afectaban a ella. No se planteaba la posibilidad de que dos personas fueran decisivas sobre el mismo par. Fue Allan Gibbard quien presentó una paradoja ligada a este problema. Según Gibbard (1974), a todo el mundo debería permitírsele fijar ciertas "características" de las alternativas o estados sociales (no los estados sociales mismos) que fueran sólo asunto suyo, de tal forma que su preferencia por tales características fuera socialmente decisiva. De esta forma, si una opción social sólo se diferencia de otra en una característica que es de mi sola incumbencia, Gibbard considera que soy moralmente decisivo sobre esa característica. Pues bien, Gibbard demostró que la decisividad moral sobre asuntos privados puede conducir por sí sola a una contradicción. El siguiente ejemplo, del propio Gibbard, nos ayudará a comprender la dinámica de su demostración. Supongamos que la persona **A** desea ponerse una chaqueta del mismo color que la

persona **B** que, a su vez, quisiera diferenciarse de **A**. Esto daría lugar al surgimiento de cuatro alternativas (**R** es rojo y **V** verde. La primera letra representa el color elegido por **A** y la segunda el elegido por B): **RR**, **VV**, **RV**, **VR**. Puesto que la elección del color de la chaqueta es sin duda un asunto privado, **A** y **B** son moralmente decisivos sobre ese asunto, es decir, que lo que ellos prefieran deberá ser aceptado por la sociedad y elegido por ella, en este caso una sociedad de dos miembros. Mas esto conduce a una contradicción, pues tendríamos que el deseo de **A** de imitar a **B** conduce a que **RR** (esto es, que ambos vistan chaqueta roja) sea socialmente preferible a **VR**, y que **VV** sea preferible a **RV**; y la voluntad de **B** por diferenciarse de **A** nos lleva a que **RV** sea preferible a **RR**, y que **VR** sea socialmente mejor que **VV**. De esta forma, la combinación de las preferencias de ambos nos aboca al siguiente resultado paradójico: **RR p VR p VV p RV p RR**.[4] De esta forma, el derecho de los individuos a vestir la chaqueta del color que ellos quieran conduce a un resultado social paradójico, intransitivo, en el que uno de los dos miembros de la sociedad (el que quiere llevar una chaqueta distinta) sale peor parado que el otro, pues al final los dos llevan el mismo color.

La sencillez o, aún más, el carácter superficial de los ejemplos no debe hacer pensar que no nos hallamos ante verdaderos problemas de agregación de preferencias, ante conflictos sociales que surgen como producto de la interacción de preferencias individuales. Preferencias individuales que, por otra parte, se hallan protegidas por principios sociales que tratan de asegurar que los mecanismos de decisión social o colectiva beneficien a todo el mundo (criterio de Pareto) y que nadie vea en peligro su libertad de elección. Tales principios, bien arraigados en la cultura moral y política de nuestras sociedades, pueden conducir a complejos conflictos y contradicciones sociales que los modelos formales de la teoría de la elección social tratan de desentrañar. Ahora bien, aun siendo esto cierto, se ha criticado que la Condición L de Sen no resulta adecuada para captar lo que es un derecho, pues, "por lo común, los individuos no tienen derechos sobre estados sociales o resultados, sino sobre acciones (suyas o de otros)" (Dowding, 2004: 151). Si se acepta esto, la vía para estudiar formalmente los problemas de la libertad de elección no es la teoría de la elección social, sino la teoría de juegos: habría que transformar las paradojas de Sen y de Gibbard en juegos de estrategia pues, de esa forma, los derechos se convierten en derechos sobre estrategias, que son acciones o conjuntos de acciones individuales. A la hora de analizar la compatibilidad entre diversos principios sociales –eficiencia paretiana y libertad, conflictos entre derechos, etc.- un enfoque centrado en la teoría de juegos no considera

[4] De nuevo la letra p significa "es preferido a". La uso en minúscula esta vez para que se distinga mejor.

que los sujetos de derecho puedan fijar los resultados finales de la interacción social, sino que se les garantiza el derecho a elegir entre diversas estrategias que conduzcan al resultado social que respeta su libertad. De esta manera, las reglas o funciones de decisión social se transforman en juegos de estrategia abstractos (*game forms*, formas de juego como el de más abajo), esto es, juegos en forma normal o estratégica en los que no se muestran los pagos (Gärdenfors, 1981; Peleg, 1998; Van Hees, 2000). Los derechos son ahora estrategias con las que cuentan los jugadores, estrategias que, combinadas entre sí, dan lugar diferentes situaciones sociales. En algunas de esas situaciones los derechos entrarán en conflicto, en otras, chocarán con el criterio de Pareto. Esto se verá más claro en los ejemplos que siguen.

		Jugador 2	
		B	A
Jugardor 1	A	(a, b)	(a, a)
	B	(b, b)	(b, a)

El jugador 1 y el jugador 2 tienen dos estrategias cada uno, A y B (el derecho de hacer A y el derecho de hacer B), que producen cuatro combinaciones de estrategias por jugador. Cada combinación genera un situación social distinta –(b, b), por ejemplo, o (b, a)- que tendrá lugar o no según las preferencias de los jugadores por una u otra acción o estrategia a la que tienen derecho. Así, si el jugador 1 opta por realizar su derecho a A y el jugador B opta por realizar el suyo a B, la situación social que se genera es (a,b), que es producto de la combinación del derecho a la estrategia A y del derecho a la estrategia B.

Si interpretamos el ejemplo de Sen en los términos de la matriz anterior, una de esas combinaciones se daría cuando ambos leen el libro, otra distinta es aquella en la que ninguno lo lee, y otras dos surgen cuando lo lee uno o lo lee otro. La Condición L de Sen se cumple, pues ambos jugadores pueden leer el libro o no leerlo, y eso queda reflejado en las distintas situaciones sociales. Ahora bien, ¿cuáles son las preferencias de los jugadores que en el ejemplo de Sen conducen a la paradoja del liberal paretiano? Si A es “no leer el libro” y B “leerlo”, tendríamos los siguientes órdenes de preferencias y las siguiente matriz:

		Puritano	
		(L) Leer	(NL) No leer
Lascivo	NL (No Leer)	(3, 3)	(1, 4)
	L (Leer)	(4, 1)	(2, 2)

Orden de preferencias de Lascivo:[5] L-L=4 > NL-L=3 > L-NL=2 > NL-NL=1

Orden de preferencias de Puritano: NL-NL=4 > L-NL=3 > NL-L=2 > L-L= 1

En efecto, esta matriz recoge bien la idea de Sen, pues lo que quiere Lascivo ante todo es que el libro sea lea, y si lo leen ambos mejor, mientras que Puritano quiere lo contrario, es decir, que nadie lea el libro o, como poco, leerlo sólo él para que Lascivo no se condene. Se trata, como vemos, del conocido Dilema del Prisionero, en el que ambos jugadores tienen una estrategia dominante: Lascivo leer el libro y Puritano no leerlo. Lo que en definitiva nos dice la Paradoja de Sen interpretada así es que para todo juego tipo que satisfaga la Condición L habrá un perfil de preferencias tal que todo equilibrio de Nash en estrategias puras será ineficiente en el sentido de Pareto. O dicho de otra forma, que la condición liberal y el criterio de Pareto son incompatibles si los jugadores tienen estrategias dominantes, como habíamos visto ya, sólo que ahora se definen los derechos como estrategias. Por otra parte, en el caso de la paradoja de Gibbard las preferencias y los derechos de los jugadores transforman en el juego tipo en este juego normal.

Jugador B

[5] El símbolo “>” se suele usar en estos casos, en los órdenes de preferencias de los juegos, con el sentido de “preferido estrictamente a”. El orden de preferencias no es cardinal: sólo indica que se prefiere más que otra, pero no en qué medida se prefiere más o menos. De esta forma, 4 es lo que más se prefiere, y 1 lo que menos, pero esto no significa que la opción más preferida se prefiera cuatro veces más.

		Verde	Roja
Jugador A	Roja	(2, 4)	(4, 1)
	Verde	(3, 2)	(1, 3)

Orden de preferencias de A: RR=4 > VV=3 > RV=2 > VR=1

Orden de preferencias de B: VR=4 > RV=3 > VV= 2> RR= 1

Como vimos más arriba, ambos jugadores tienen derecho a ponerse la chaqueta del color que quieran. Dado que el Jugador A quiere coincidir con el otro y el Jugador B lo que quiere sobre todo es evitar coincidir, el juego carece de equilibrio en estrategias puras. De esta forma lo que nos dice la paradoja de Gibbard es que para toda forma de juego que satisfaga la Condición L habrá un perfil de preferencias para ese juego que no contenga ningún equilibrio de Nash en estrategias puras.

A menudo se ha señalado que la Condición L de Sen pone un énfasis excesivo, como se vio más arriba, en la efectividad o poder de los individuos para que se realice o lleve a cabo de hecho el estado de cosas que asegura sus derechos. De esta forma, parece que la condición de Sen captaría mejor la idea de la libertad positiva que la de libertad negativa, más propia del liberalismo. La libertad negativa sólo nos asegura que nadie impedirá que se produzca tal o cual resultado, no que el resultado se producirá de hecho. Yo soy negativamente libre de leer *El amante de Lady Chatterly* si nadie me impide leerlo, pero eso no asegura que lo *pueda* leer: quizá no pueda si no tengo un ejemplar a mano. La Condición L recoge, pues, el poder hacer algo, no la ausencia de constricciones exteriores que me impiden hacer ese algo. Si la Condición L se reformula de forma que refleje sólo la idea de libertad negativa la Paradoja de Sen desaparece, pero la de Gibbard se mantiene. Sin embargo, al interpretar los derechos como estrategias admisibles para los individuos hemos visto que la paradoja de Gibbard pierde fuerza, pues ya no nos encontramos con un resultado incoherente (la violación de la transitividad), sino con un resultado inestable (la ausencia de equilibrio de Nash). Como veremos más adelante, la inestabilidad se puede resolver mediante algún convenio o acuerdo y no supone mayor problema para el liberalismo y la libertad negativa como principios sociales que han de regir la agregación de preferencias individuales. Las paradojas que hacían incompatibles la libertad y la eficiencia paretiana, por una lado, y las libertades entre sí, por otro, desaparecen al reformularlas como

juegos de estrategia en los que la libertad se reduce al derecho de que los demás no interfieran en nuestras decisiones.[6]

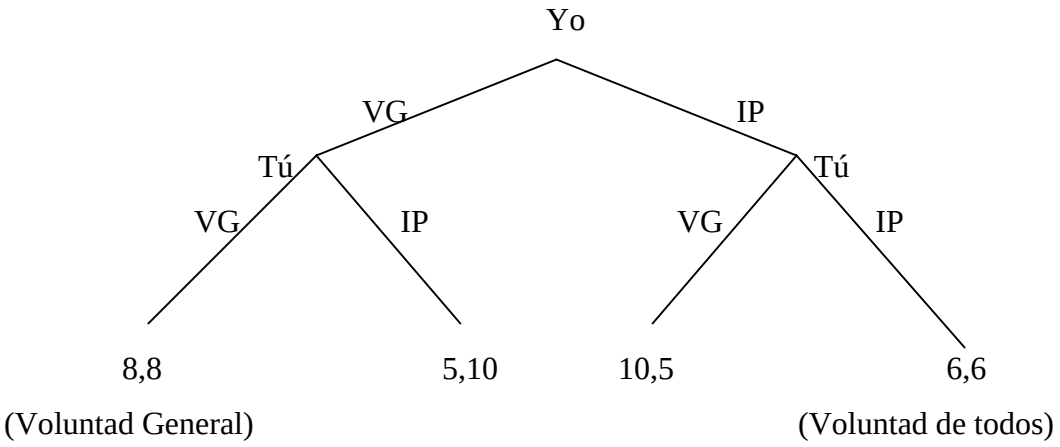
3. Voluntad general y libertad interior

En uno de los pasaje más olvidados de *Elección Social y Valores Individuales* Kenneth Arrow traza, casi a vuelapluma, una posible conexión entre la voluntad general de Rousseau, el imperativo categórico de Kant y el concepto de ordenación social. En cierto modo, asegura Arrow, la ordenación social podría considerarse una suerte de voluntad general "que se supone...que es la misma para todos", o un imperativo moral que es tanto una ordenación social como "una ordenación moral para todo individuo: representa la voluntad que tendría todo individuo si fuera plenamente racional". La posible relación entre la voluntad general y la elección social, que nos aleja de los tradicionales fundamentos utilitaristas de la teoría de la elección social, nos lleva a una cuestión que ha sido tratada también desde la teoría de juegos: me refiero a la liberad de los sujetos para elegirse a sí mismo, es decir, para ser interiormente libres. En la propuesta de Arrow el resultado de la elección social, el orden social de preferencias, la preferencia colectiva, se convierte en la voluntad general que los individuos asumen –hacen suya- si son plenamente racionales. Pero esto puede plantear al individuo un conflicto entre su voluntad privada y la voluntad general (que, como señala Arrow, es también su propia voluntad), en la medida en que ambas se contradigan. En otras palabras, ¿cómo se puede resolver, si se da, el conflicto entre el interés particular y el interés público? Históricamente se han dado dos respuestas a esta pregunta: el conflicto se puede resolver mediante la ley o mediante la virtud. En un mundo de egoístas recalcitrantes, en un mundo de maximizadores de su bienestar privado, en un mundo, en fin, hobbesiano, los individuos serán incapaces de seguir los dictados del bien público.[7] Si acaso se llegara en semejante mundo a un resultado colectivo mediante un mecanismo de elección social, los individuos preferirían que los demás se atuvieran a ese resultado común, mientras que ellos defienden su interés egoísta. La consecuencia bien conocida es que en el estado de naturaleza hobbesiano reina el

[6] Sin embargo, en opinión de Sen "las limitaciones de la aproximación [a los problemas de la libertad] como formas de juego reside en parte en el hecho de que se concentra exclusivamente en el aspecto de elección de la libertad" (1991: 211). Esto impide, en su opinión, analizar los ataques a la libertad procedentes de las "acciones invasoras" o de la "inhibición a la hora de elegir", que padecen las mujeres en las sociedades sexistas. La teoría de la elección social puede tratar esos casos al poder usarse para analizar situaciones en las que se va más allá de la decisión real de los jugadores.

[7] En las páginas que siguen hago un uso ilustrativo de conceptos clásicos como el de estado de naturaleza o el de voluntad general. No se trata, por supuesto, de un análisis de la teoría de Hobbes o de Rousseau.

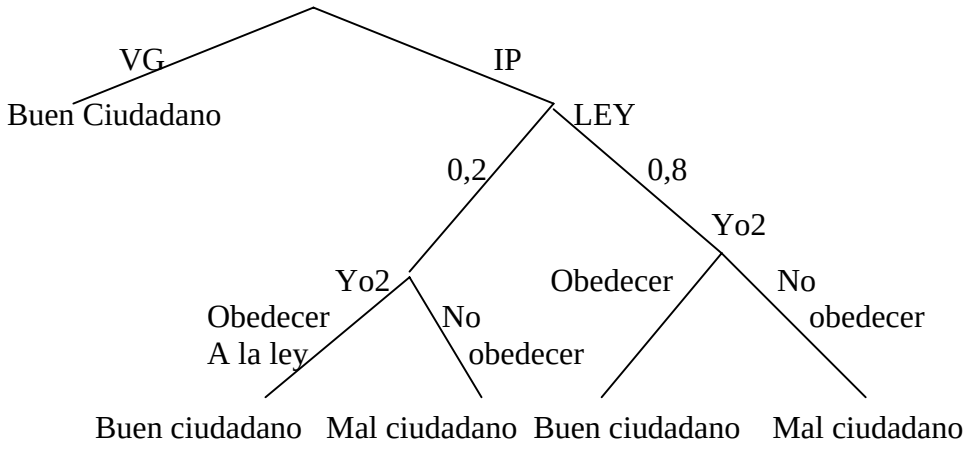
casos, la guerra de todos contra todos, y la única forma de salir de tal situación consiste en alienar la libertad en un soberano que, representante de Dios en la tierra, imponga la ley y el orden mediante el temor. El estado de naturaleza hobbesiano sería, pues, un dilema del prisionero en el que sólo el soberano puede imponer la cooperación a los jugadores.[8] Que en el estado de naturaleza pueda haber individuos dispuestos a cooperar por propia voluntad es indiferente, pues serán arrollados por los egoístas: no cooperar es la estrategia dominante.



Supongamos que en el juego de más arriba (un dilema del prisionero en forma extendida en el que el segundo jugador, Tú, toma su decisión sabiendo ya lo que ha hecho el otro) Tú y Yo queremos poner en práctica el acuerdo colectivo al que se ha llegado en nuestra sociedad para salir del Estado de naturaleza hobbesiano. Si Yo pretendo seguir los dictados del acuerdo que es producto de la voluntad general, VG, Tú tienes un incentivo enorme para seguir tus intereses privados, IP (pues así obtienes la máxima ganancia, 10, aunque sea a mi costa). Al conocer Yo el juego que jugamos y ver en el árbol que si juego VG tú vas jugar IP, esto es, me dejarás que cargue con los costes de poner en práctica el bien público mientras te dedicas a tus intereses privados, Yo no jugaré VG al comenzar el juego, sino que también elegiré IP y el resultado será colectivamente malo (pues el resultado agregado de VG más VG es mejor que el resultado de IP más IP). El papel de la ley para salir de esta situación sería el siguiente:

Yo1

[8] David Gauthier (1969) es el primero que analiza la teoría hobbesiana del Estado como si se tratara de la superación de un dilema del prisionero.



Al introducir la ley el juego se transforma. Ya no se trata sólo de que Tú y Yo intentemos cooperar para salir del estado de naturaleza, sino que ahora hay un juego en el que Yo juego contra mí mismo (Yo1 y Yo2) y debo decidir primero si atiendo a la voluntad general o a mis intereses privados asociales.[9] Si decido satisfacer mis intereses privados incluso a costa de ejercer la fuerza contra los demás como en el estado de naturaleza, ahora deberé tener en cuenta un nuevo elemento: las leyes que nos hemos dado libremente y que me van a forzar a ser un buen ciudadano. Pero si las leyes son blandas (esto es, si la probabilidad de que me castiguen es baja, 0,2) quizás me merezca la pena no obedecer a la ley y seguir defendiendo mis intereses privados a la fuerza si es preciso. Si las leyes son duras (si la probabilidad de ser castigado es muy alta), se me forzará con mayor seguridad a ser un buen ciudadano. Los miembros de una sociedad hobbesiana que ha salido del estado de naturaleza no son ya libres del todo, pues han de trocar parte de la libertad irrestricta de que gozaban en el estado de naturaleza por seguridad. El soberano, la ley, protege a los súbditos, pero lo hace obligándolos a asumir la coacción que la propia ley implica.

Nos falta por analizar, sin embargo, una rama del juego anterior, la rama izquierda, la de quien elige de primeras honrar la voluntad general. ¿Qué hay de ese ciudadano que elige de buen grado seguir la voluntad general, ese ciudadano que no promueve el bien público por temor a la ley sino por convicción?¿Qué hay de ese ciudadano que hace suya la voluntad general? Desde el punto de vista de los resultados, parece que no existe diferencia alguna entre el ciudadano que elige por convicción seguir los dictados de una voluntad general que

[9] No se trata de que los individuos no puedan tener intereses privados, sino de que esos intereses privados son los del estado de naturaleza, son intereses privados que no dudan en aprovecharse de la voluntad general.

hace suya y el ciudadano que sigue la voluntad general por temor a la ley: ambos se presentan a ojos de los demás como buenos ciudadanos. Hay una diferencia obvia, por supuesto, que el juego resalta con claridad. Si los beneficios de ser un mal ciudadano son muy altos y la probabilidad de que la ley castigue al mal ciudadano es muy baja, éste no tendrá incentivo alguno para comportarse como un buen ciudadano. En cambio, el ciudadano virtuoso, el ciudadano que opta por la voluntad general, que la hace suya, no necesitará esa coacción externa para ser un buen ciudadano. Para el ciudadano virtuoso la voluntad general ya no será impuesta desde fuera, sino que será libremente –virtuosamente- elegida desde dentro por la razón:

> "¿Qué es, pues, el hombre virtuoso? Es el que sabe vencer sus afectos. Porque entonces sigue su razón, su conciencia, cumple su deber, se mantiene en el orden y nada puede apartarlo de ahí. Hasta ahora tú sólo eras libre en apariencia; no tenías sino la libertad precaria de un esclavo al que no se ha mandado nada. Sé libre ahora en efecto; aprende a volverte tu propio dueño; manda en tu corazón, oh Emilio, y serás virtuoso" (Rousseau, 1998: 666).

El hombre y la mujer virtuosos no tendrán que ser coaccionados desde fuera por la ley para ser buenos ciudadanos: la voluntad general es su voluntad y por eso son libres. Ahora bien, la ley debe existir porque siempre habrá quien se deje arrastrar por las pasiones, por los intereses, por la vanidad, por el *amour propre*. El ciudadano virtuoso, que es interiormente libre, juega un juego contra sí mismo en el que doblega al egoísmo, pues su estrategia dominante es la cooperación. El mal ciudadano juega contra sí mismo un dilema del prisionero en el que el egoísmo siempre gana, pues el mal ciudadano no es libre interiormente y sólo la coacción externa puede forzarlo a cooperar.[10]

[10] Queda abierta aquí la pregunta de por qué el virtuoso lo es en primera instancia. Se trata de un problema de psicología moral, no de teoría de juegos. Ahora bien, una vez que el ciudadano ha decidido hacer suya la voluntad general se podría representar la lucha interna para doblegar las pasiones o los intereses antisociales como un juego de reputación, pero un juego de reputación interna (Kim, 2006), en el que el virtuoso juega contra sí mismo y recibe un beneficio psicológico si es capaz de no dejarse vencer por las pasiones. En ese juego el ciudadano, si es virtuoso, ya no se ve forzado por una ley exterior, sino por una ley interna, una ley que se da a sí mismo, una ley de su voluntad. Para una análisis de la voluntad general rousseauniana y el imperativo categórico kantiano en los términos de la teoría de juegos véase Doménech (1989: 265-292)

4. Teoría de la negociación y teoría del contrato

Ahora bien, ¿por qué una sociedad de virtuosos sería mejor que una sociedad de ciudadanos que obedecen la ley por temor a la sanción? Hemos visto que sin temor a la sanción el ciudadano virtuoso promueve también la voluntad general. Eso no implica, sin embargo, que no puedan surgir también entre virtuosos dilemas de la cooperación que les obliguen a arbitrar mecanismos de cooperación, acuerdos, para superar esos dilemas. La teoría de juegos es una teoría matemática, una teoría formal, y por lo tanto los supuestos que establece sobre los jugadores son también formales: exige de ellos racionalidad, entendida como coherencia lógica (que sus preferencias sean transitivas y completas) y como maximización de utilidades. Si las preferencias del jugador son egoístas, el jugador tratará de maximizar su propio beneficio; sin son altruistas tratará de maximizar el beneficio ajeno. La teoría de juegos no supone, pues, que los individuos son egoístas. Lo que sí nos permite evidenciar el uso instrumental de la teoría de juegos es que ya sean egoístas o altruistas, la interacción de los jugadores puede ocasionar dilemas cuya solución es compleja si no se llega a algún tipo de acuerdo. En otras palabras, aunque parece que en la sección anterior insinuamos que en una sociedad de individuos virtuosos interiormente libres la voluntad general de cooperar surgiría por sí sola, lo cierto es que la teoría de juegos demuestra que también en una sociedad de individuos virtuosos se precisa a veces algo más que preferencias virtuosas. Esto queda patente en el Dilema del Altruista y del Egoísta, que se refleja en la siguiente matriz.

	Pasa tú primero	Paso yo primero
Pasa tú primero	0, 0	1,1
Paso yo primero	1,1	0, 0

Cuando dos personas se encuentran ante un puerta por la que quieren pasar y no pueden hacerlo a la vez es necesario que una de ellas ceda el paso. Si cada una de esas dos personas se empeña en pasar primero, si ninguna cede el paso, se atascarán en la puerta y no podrán

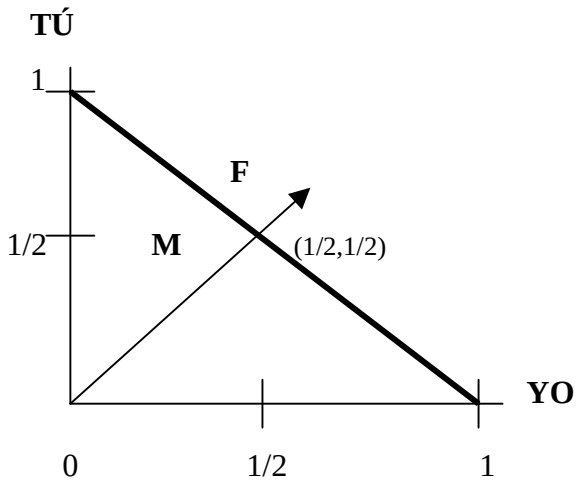
pasar más que a empujones, hasta que uno entre primero. El par de estrategias (Yo primero; yo primero) conduce a un mal resultado, esto es, (0,0): ninguno pasa. Pero tampoco resuelve el dilema el hecho de que ambos jugadores venzan sus afectos egoístas y, de forma virtuosa, cedan el paso al otro, porque si los dos ceden a la vez y no salen del círculo vicioso en que han caído por virtuosos tampoco pasarán. El par de estrategias (Pasa tú primero; pasa tú primero) conduce al mismo resultado que la estrategia egoísta, esto es, al peor resultado, (0, 0), de forma que los jugadores tampoco pasarán por la puerta, se quedarán ante ella cediéndose amablemente el paso una eternidad. La diferencia con los jugadores egoístas que tratan de pasar a la vez es que al menos aquí no hay empujones. Como vemos, la teoría de juegos no presupone nada respecto del contenido de las preferencias de los jugadores. Ya sean egoístas o virtuosos, el trabajo de la teoría consiste en analizar el resultado emergente de su interacción. Y en este caso lo que nos dice la matriz es que para alcanzar uno de los dos equilibrios de Nash –(1, 1) y (1, 1)- algún jugador tiene que ceder el paso y el otro tiene que pasar.

Si el dilema del altruista se juega una sola vez y no se incluye ninguna variable que nos permita resolverlo (uso de la fuerza, normas, etc.), los egoístas se enfrentarán y los virtuosos altruistas se quedarán ante la puerta. Sin embargo, si el juego se repite, si los jugadores se encuentran muchas veces ante la misma situación, pueden salir del atolladero recurriendo a lo que se conoce como estrategias mixtas. En el Dilema los jugadores tienen dos estrategias puras, esto es, pasar primero o ceder el paso. Se denominan así porque el jugador escoge una o escoge la otra, pero no las mezcla, es decir, no elige una estrategia unas veces y la otra otras veces. En el Dilema de la matriz anterior, si los jugadores deciden poner en practica una estrategia mixta –jugar unas veces “paso yo primero” y otras “pasa tú primero”- el equilibrio de Nash se produce cuando los jugadores eligen la mitad de las veces pasar ellos primero y la otra mitad que pase primero el otro. Si el juego se repite esta es sin duda la solución.

Como se puede apreciar, en este caso nos hemos alejado de la concepción puramente instrumental de la teoría de juegos, pues ya sean virtuosos o egoístas los jugadores, la solución a su dilema se la proporciona el propio juego, y no su idea de lo que está bien y de lo que está mal, no su particular teoría moral. O dicho en otras palabras, cuando se hace un uso instrumental de la teoría de juegos, ésta presenta los problemas en su aspecto formal para que resulten más claros, pero las teorías políticas o morales los resuelven. Así, por ejemplo, en el caso del dilema del prisionero la teoría de juegos demuestra que la racionalidad individual

puede conducir a un resultado colectivo irracional. Cuando el dilema se juega una sola vez, la teoría de juegos considera que los más racional es no cooperar. Sin embargo, una teoría moral podría recomendar que se cooperara incluso cuando se juega una sola vez: para un utilitarista la solución cooperativa es la correcta moralmente porque cuando se coopera la suma de utilidades es mayor; para un kantiano cooperar en un dilema del prisionero es la única forma de que no se considere al otro como un medio para maximizar el beneficio privado. El uso instrumental de la teoría de juegos sirve, pues, como herramienta descriptiva de los problemas de interacción moral o política, pero ni ofrece soluciones ni las explica.

En el dilema del altruista que vimos más arriba la situación es distinta, pues no sólo modela la teoría de juegos el problema, sino que ofrece la solución, en principio, más justa. Desde esta perspectiva, la ética se puede considerar una rama más de la teoría de decisión y no una simple herramienta heurística. Como afirma David Gauthier, uno de los principales defensores de la concepción que relaciona íntimamente ética y teoría de juegos, "elegir racionalmente es elegir moralmente...La moralidad...se puede generar como una constricción racional a partir de las premisas no morales de la elección racional" (Gauthier, 1986: 4). En efecto, si el dilema del altruista se repite, los jugadores llegarán al equilibrio de Nash en estrategias mixtas que, además, es la solución más justa, por imparcial, al problema que se les plantea. Y del mismo modo, si el dilema del prisionero se juega de forma repetida los jugadores aprenderán a cooperar, que es la solución correcta desde un punto de vista moral. La teoría de juegos no es una teoría moral, es una teoría formal, pero las soluciones que propone cuando los jugadores se encuentran en repetidas ocasiones pueden adoptarse como soluciones morales. Un resultado éste –la elección racional que es a la vez elección moral- que se puede forzar, es decir, no es necesario esperar a que la interacción repetida lo produzca. Los jugadores pueden regatear y llegar a acuerdos justos –imparciales- de forma que se obtenga un resultado que beneficie a todos. El resultado de la negociación, la moral por acuerdo, llevaría a un contrato social aceptable para todos. Así, por ejemplo, en el caso del dilema del altruista, la negociación conduce al mismo resultado que la repetición indefinida del juego, como se ve en el siguiente gráfico:



Si los jugadores negocian llegarán a la conclusión de que el dilema que se les plantea se resuelve pasando cada cual primero por la puerta la mitad de las veces en que se encuentren. Los puntos que caen por debajo de la frontera del conjunto de negociación –determinada por la línea que une el punto 1 de cada coordenada-, así como los que están en la misma frontera, se corresponden con el conjunto de negociación: sólo esos puntos (como el M, por ejemplo) son posibles. Lo que caen más allá de la frontera (como el F) no son posibles, no pueden ser objeto de la negociación. Por otro lado, los puntos que están en la frontera son todos óptimos de Pareto.[11] El punto cero es el punto de desacuerdo en este juego, una suerte de "estado de naturaleza" del que los jugadores quieren salir pero al que pueden amenazar con volver. Pues bien, si los jugadores tienen el mismo poder de negociación el equilibrio de Nash coincide con el mayor resultado de multiplicar entre sí los posibles resultados de la negociación –en nuestro sencillo los resultados son pasar por la puerta (1) o no pasar (0)- descontándoles el punto de desacuerdo. (El equilibrio se obtiene también en el punto en que la diagonal que sale del punto desacuerdo corta la frontera del conjunto de negociación).

(1/2 – 0) x (1/2 – 0) = 1/4

(1 – 0) x (0 – 0) = 0

Una negociación de esta naturaleza resolvería, por ejemplo, la paradoja del liberal paretiano, pues los jugadores podrían llegar a un acuerdo sobre quién lee el libro y quién no, y un acuerdo similar resolvería también la paradoja de Gibbard, al determinarse mediante negociación quién ha de llevar un jersey de un color y quién de otro y en qué (o en cuántas)

[11] Incluido (0, 1).

ocasiones. De esta forma, la negociación da lugar a la interacción cooperativa de los jugadores:

> "La idea de la negociación nos permite incorporar a nuestra descripción de la elección racional cooperativa lo que se pierde desde la perspectiva de la elección social y del utilitarismo: la implicación activa de quienes cooperan. También nos permite captar el requisito de que el acuerdo sobre una estrategia común sea voluntario" (Gauthier, 128).

Sin embargo, al filósofo o la filósofa curtidos en los complejos problemas de la justicia distributiva les costará aceptar que la teoría de la negociación sirva para resolver tales cuestiones, y menos aún para explicar la posible adopción por parte de agentes morales reales de los posibles contratos sociales que emerjan de la negociación. ¿Cómo olvidar que los agentes tienen poder?¿Cómo no tener en cuenta que en la sociedad hay ricos y pobres, fuertes y débiles?¿Cómo dejar a un lado el hecho de que el resultado de esas negociaciones hipotéticas no es vinculante? En primer lugar, las teorías contractualistas de la negociación pueden suponer, siguiendo la estela de Rawls, que en el "estado de naturaleza" los individuos negocian tras un velo de ignorancia que iguala su poder. La negociación se desarrolla así de forma imparcial. Si se supone, en cambio, que los individuos tienen información completa, como en la teoría del contrato de Gauthier (1989), la negociación se somete a restricciones que eviten que los jugadores puedan negociar con ventaja: Gauthier impone a los negociadores una condición lockeana (*Lockean proviso*) según la cual nadie puede mejorar su situación empeorando la de otro.[12] En segundo lugar, se considera preciso adoptar algún tipo de índice social para poder comparar el interés que puedan tener los jugadores por los distintos bienes. En nuestro ejemplo de quien pasa primero por la puerta, un índice social que compare el intereses de los jugadores podría ser el siguiente: el altruista puede tener el doble de interés que el egoísta en ceder el paso, de forma que se llegue a un acuerdo en el que el egoísta pasa primero tres cuartos de las veces y el altruista un cuarto. Este resultado sería también justo dados los intereses de los jugadores. Si lo que se distribuyen son bienes básicos, por ejemplo, la negociación tiene que partir de un mínimo y todo lo que se salga de ese mínimo es innegociable. Dicho en otros términos, el punto de desacuerdo se puede situar donde se quiera y tendrá, como los índices sociales, un marcado carácter cultural y social. Por último, el hecho de que se suponga que los individuos son maximizadores de utilidades hace

[12] Obsérvese que esta es una forma de definir el criterio de Pareto. La teoría de la justicia de Gauthier se analiza a fondo en Vallentyne (1991).

que las decisiones que adoptan sean vinculantes, pues son las que más les benefician. No cumplir los acuerdos perjudica a la larga a quien incumple.

5. Evolución de la moral y teoría de juegos

Pese a todas las restricciones que se puedan aplicar a la negociación, su uso como base de la teoría de la justicia no carece de detractores, especialmente por la naturaleza hipotética del contrato. El interés de los jugadores a la hora de cumplir con los términos del contrato aparece como algo difuso que no tiene fuerza categórica. Además, no está claro que toda negociación tenga una, y sola una, solución racional dadas las muchas soluciones posibles que hay en un proceso de negociación, ni está claro tampoco que esa solución racional, si existe, coincida siempre con la solución moral de un problema de interacción. Los factores irracionales que pueden influir en la negociación son tantos, puede haber tantos sesgos, que el resultado puede quedar indeterminado.

Ahora bien, ya hemos visto que en algunos casos el resultado de la negociación es el mismo que se daría sí los jugadores se encontraran en repetidas ocasiones y aprendieran a cooperar. Si en un dilema del prisionero los jugadores se enfrentan en múltiples ocasiones terminan cooperando porque es lo que más les beneficia (Axelrod, 1986). La cooperación puede surgir, pues, como resultado de la evolución del propio juego en un tiempo indefinido y no tanto de la negociación expresa de jugadores hipotéticos perfectamente informados. Pero en otros casos ocurre lo contrario, es decir, que la evolución del juego lleva a un resultado muy distinto al de la negociación hipotética. Teoría de juegos evolutiva y teoría de la negociación suponen, pues, aproximaciones distintas al estudios de los problemas morales y de justicia distributiva.

La teoría de juegos evolutiva surge al aplicar la teoría de juegos a cuestiones de evolución biológica. En el trabajo que inaugura esta línea de investigación se analiza la adaptación biológica –el éxito o el fracaso a la hora de transmitir los genes a las siguientes generaciones- como una estrategia. Si la estrategia tiene éxito la población que la adopte será genéticamente eficaz, esto es, estará bien adaptada al medio y bien preparada para resistir el asalto de otras estrategias evolutivas. A las estrategias que tienen éxito se la denomina evolutivamente estables. Según la definición de Maynard Smith, “una estrategia es estable en sentido evolutivo cuando no existe una estrategia mutante que dé una eficacia darwiniana superior a

los individuos que la adoptan" (1978: 122). Supongamos, para simplificar las cosas, que dos individuos pueden adoptar una estrategia adaptativa agresiva (a la que llamaremos, siguiendo la costumbre, Halcón) o una estrategia que no sea agresiva (Paloma). En el juego Halcón-Paloma hacer siempre de halcón o hacer siempre de paloma (estrategias puras) no es una estrategia evolutivamente estable. Seguir una estrategia agresiva es mortal si uno se encuentra con otros Halcones o una Paloma se encuentra con un Halcón: los Halcones se exterminan pero son más eficaces en la adaptación genética que las palomas. Asumir siempre la estrategia menos agresiva funciona sólo cuando uno se encuentra con otras Palomas. En este juego la estrategia evolutivamente estable es la estrategia mixta (que es el equilibrio de Nash), que consiste en hacer de Halcón 8/13 de las veces y de Paloma 5/13. En otras palabras, "la estrategia del Halcón no puede invadir a una población que emplee la estrategia mixta M".

	HALCÓN (H)	PALOMA (P)
HALCÓN (H)	-5, -5	10, 0
PALOMA (P)	0, 10	2,2

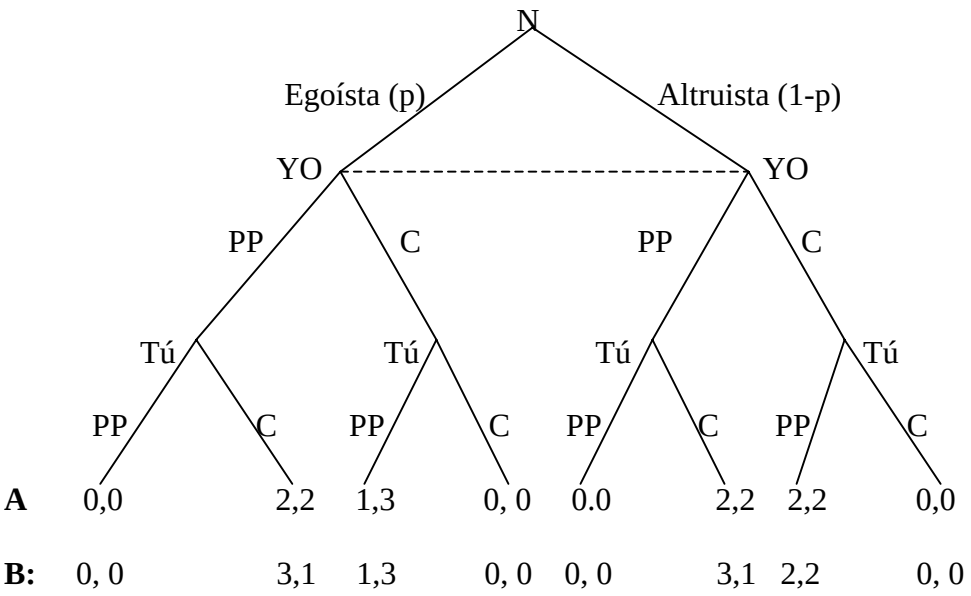
Figura 3: Juego Halcón-Paloma

La teoría de juegos evolutiva ha encontrado numerosas aplicaciones fuera de la biología, sobre todo en el terreno de la teoría moral y de la sociología. Así, por ejemplo, el juego Halcón-Paloma se ha aplicado con notable éxito al estudio de normas sociales. Si suponemos que H es una estrategia que promueve la desconfianza en los demás y P una que promueve la confianza, este juego nos dice que la confianza y la desconfianza incondicionales no son evolutivamente estables: no puede sobrevivir una sociedad en la que todos desconfían entre sí, y una sociedad en las que todos confían es, a su vez, "invadida" con facilidad por desconfiados que traten de aprovecharse de los confiados. Habría que seguir una estrategia mixta en la que se mezclaran en distinta medida la confianza y la desconfianza. Eso es, precisamente, lo que suele hacer la gente en sociedad.

Por otro lado, desde la perspectiva de la teoría moral, el uso de la teoría de juegos evolutiva permite suponer que las normas morales son soluciones evolutivamente estables a los problemas de coordinación que se han encontrado los seres humanos en su evolución cultural (Binmore, 2005). La teoría de juegos evolutiva tiene un marcado carácter explicativo del que carecen los usos instrumental y contractualista de la teoría de juegos, pues trata de

analizar formalmente por qué tenemos las normas morales que tenemos, cómo surgen, por qué perviven. Ahora bien, una apuesta decidida en esta línea implica suponer, además, que tenemos las normas morales que tenemos porque resulta que son las mejores evolutivamente hablando: las mejores para la convivencia humana en términos de su evolución cultural. Veamos esto con un ejemplo para comprender mejor lo que implica para la teoría moral el enfoque de la teoría de juegos evolutiva.

Supongamos que, al igual que en el juego de más arriba, dos jugadores (Tú y Yo) tienen que decidir quién pasa primero por una puerta, pero en este caso supongamos que el primero que pase puede comerse hasta tres porciones de un pastel dividido en cuatro partes que hay al otro lado. Supongamos también que los dos jugadores tienen hambre y son golosos. Si a esto añadimos que un jugador, Tú, tiene preferencias egoístas o preferencias altruistas, pero el otro (Yo) desconoce si se encuentra ante un egoísta o ante un altruista, ¿cómo debe comportarse? Antes de abordar esta pregunta podemos ver cómo se plantea la situación en términos de la teoría de juegos. En primer lugar, habrá que representar un contexto (por convención se denomina a ese contexto N, del inglés *nature)* en el que Yo se pueda encontrar con un egoísta o con un altruista. Luego se representa la decisión de Yo, que no sabe si se halla ante un egoísta o ante un altruista: la línea de puntos del juego en forma extendida de más abajo representa la incertidumbre de Yo, que no sabe dónde está, si ante un tipo de jugador u otro. Una vez que Yo decida pasar primero por la puerta o ceder, le toca el turno a Tú, que toma su decisión sabiendo ya lo que ha decidido Yo. Pues bien, ¿qué es lo que hay que hacer en este juego?¿Hay que pasar o hay que ceder?



Una teoría moral deontológica que incluya el altruismo y la generosidad entre sus primeros principios hará un uso instrumental de la teoría de juegos para analizar la situación y prescribirá la conducta altruista con independencia de sus consecuencias. Una teoría del contrato recomendará una negociación entre los jugadores que les lleve al resultado (2, 2), que es el más justo porque ambos jugadores se encuentran en las mismas circunstancias (tienen hambre, son golosos, tienen el mismo poder). En caso de que los jugadores no cumplan el trato la teoría apelará a una institución (el Estado, por ejemplo) que lo haga cumplir. El análisis evolutivo del juego llega a la solución equitativa por otro camino. ¿Qué hará Yo, que decisión tomará? En la teoría de juegos evolutiva Yo ya no es un jugador individual maximizador de utilidades, sino un miembro de una población. Yo vive en un mundo en el que hay muchos Yos y muchos Tus, y en esa población hay egoístas y altruistas. A Yo le gustaría saber ante quien se halla, pero no siempre lo sabe. Si supiera que está ante un egoísta quizás adoptaría la estrategia de un altruista. Como se ve en el árbol de más arriba, cuando Yo actúa como un egoísta (línea B de resultados o pagos del juego) sale muy mal parado si se encuentra con otro egoísta: son dos Halcones que se pelean en la puerta. En cambio, si cede ante el egoísta puede llevarse una porción de pastel. Por otro lado, si se encuentra con un altruista y adopta una estrategia egoísta (Línea A de pagos) saldrá muy beneficiado, pues pasará primero por la puerta y se llevará tres trozos. Mas el caso es que Yo no sabe si se encuentra ante un egoísta o ante un altruista. Si cree que es muy alta la probabilidad (p) de que en la población haya más egoístas que altruistas, puede adoptar una estrategia de altruista de supervivencia. Si cree que la probabilidad (1-p) de que en la población haya muchos altruistas es muy elevada puede adoptar una estrategia egoísta. Ahora bien, si en la población hay muchos altruistas, su estrategia será invadida con facilidad por egoístas que siempre pasan primero por la puerta y se llevan más ración. Eso les hará comportarse a veces, si no son santos, como egoístas, para defenderse. De esta forma, y ante la incertidumbre, los individuos terminarán pasando los primeros por la puerta la mitad de las veces y cediendo el paso la otra mitad. En promedio el pastel se habrá repartido equitativamente y la norma de ceder el paso la mitad de las veces ante la puerta habrá tenido éxito adaptativo en la prehistoria evolutiva de la sociedad de Yos y Tus que hemos imaginado aquí.[13] Desde esta perspectiva cabe suponer, pues, que las normas de justicia y la normas morales son estrategias

[13] "En una población en la que todo el mundo exige la mitad del pastel, cualquier mutante que exija algo diferente obtendrá menos que la media de la población. Exigir la mitad del pastel es la estrategia evolutivamente estable en el sentido de Maynard Smith y Price" (Skyrms, 1996: 11). En nuestro ejemplo lo que se termina exigiendo es pasar por la puerta el primero la mitad de las veces. Así todos terminan comiendo a la larga la misma cantidad de pastel.

evolutivamente estables. Cabe suponer también que la moral y la justicia surgen como subproducto de la interacción de los individuos. Y cabe suponer también, en consecuencia, que la justicia y la moral no precisa de individuos racionales plenamente informados, sino de individuos que siguen normas que son producto de la evolución cultural humana.

6. Conclusión

Cabe suponer todo esto empleando la teoría de juegos evolutiva, ¿pero se puede demostrar? ¿Son tan similares la evolución biológica y la evolución cultural? Y si una norma de justicia y o una norma moral no son evolutivamente estables, si son evolutivamente inestables, ¿qué me puede importar como agente moral en el momento en que he de tomar una decisión moral inaplazable? Hay que resolver todavía muchas preguntas en este terreno tan novedoso. Cuando se habla de evolución puede hacerse referencia a un proceso simulado en el laboratorio, como el ingeniado por Axelrod, para analizar el surgimiento de normas, y no a la evolución moral real de la especie humana. Ahora bien, aunque se pueda simular en el laboratorio el surgimiento evolutivo de normas morales, que las normas de justicia sean equilibrios de Nash seleccionados por la evolución implica una idea muy distinta, pues aquí los lapsos temporales se miden en miles de años. ¿Son las normas de justicia producto de la evolución humana, de nuestra convivencia en pequeñas sociedades de cazadores y recolectores? ¿Hay normas inmorales que también resultan evolutivamente estables? ¿Se le puede dar fuerza categórica a lo evolutivamente estable por el mero hecho de serlo?

En este capítulo no podemos hacer frente a semejantes preguntas, pero merece la pena plantearlas porque evidencian la intensa relación que puede darse entre la ética y la teoría de juegos. De los tres usos de esta teoría que hemos presentado aquí, quizá sea el evolutivo el que tiene ante sí mejores perspectivas. Sin embargo, aún no existe demasiada relación entre expertos en teoría de juegos y expertos en filosofía moral y política. Para los primeros, para algunos de ellos al menos, los filósofos académicos se contentan con levantar cortinas de humo carentes de sentido (Binmore, 2005: 37). Para los segundos, para la gran mayoría, la teoría de juegos resulta inútil si su estructura formal no se interpreta de forma especial, pero al hacerlo nos las veremos inevitablemente con “tradiciones muchos más viejas” (Rawls, 1999: 58). Los filósofos temen encontrarse con una nueva ética matematizada, una nueva ética demostrada según el orden geométrico, que aporte poco a los verdaderos problemas morales. La teoría de juegos, sin embargo, ha demostrado hasta ahora que puede ser de gran utilidad

para desentrañar complejos problemas de interacción humana y proponer soluciones. La ética, la teoría de la justicia, se enfrentan también a uno de los grandes problemas de la interacción humana, a saber, el de construir una sociedad mejor, más racional y más justa. La colaboración entre la teoría de juegos y la filosofía moral y política puede ser, pues, muy fructífera y no tiene por qué estar lastrada por otros intentos –ya sea el más antiguo de Spinoza, ya sea el más reciente de la filosofía analítica- de emplear disciplinas formales para abordar cuestiones morales.

BIBLIOGRAFÍA

Arrow, K. (1963), *Social Choice and Individual Values* [*Elección social y valores individuales*, Madrid: Instituto de Estudios Fiscales, 1974]

Axelrod, R. (1986), *La evolución de la cooperación*. Madrid: Alianza Editorial.

Binmore, K. (2005), *Natural Justice*. Oxford: Oxford University Press.

Braithwaite, Richard Bevan. 1955. *Theory of Games as a Tool for the Moral Philosopher*. Cambridge: Cambridge University Press.

Domènech, A. (1989), *De la ética a la política*. Barcelona: Crítica.

Dowing, K. (2004), "Social Choice and the Grammar of Rights and Liberties", *Political Studies*, 52: 144-161.

Gärdenfors, P. (1981). "Rights, games and social choice", *Noûs*, 15: 341-356.

Gauthier, D. (1969), *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes.* Oxford; Clarendon Press.

Gauthier, D. (1986), *Morals by Agreement*. Oxford: Clarendon Press.

Gibbard, A. 1974. "A Pareto consistent libertarian claim", *Journal of Economic Theory*, 7: 388-410.

Harsanyi, J. (1992), "Game and decision theoretic models in ethics", en R. J. Aumann y S. Hart, *Handbook of Game Theory. Vol. 1*. Londres: Elsevier, pp. 671-707.

Kim J-Y (2006), "Hyperbolic discounting and the repeated self-control problem", *Journal of Economic Psychology*, 27: 321-462.

Maynard-Smith, J. (1978), "La evolución del comportamiento", *Investigación y Ciencia* 26: 116-127.

Peleg, B. (1998), "Effectivity functions, game forms, games, and rights", *Social Choice and Welfare*, 15: 67-80.

Rawls, J. (1999) [1958], "Justice as fairness", en S. Freeman (ed.), *John Rawls. Collected Papers*. Cambridge, Mass.: Harvard University Press, pp. 47-72.

Rousseau, J.-J- (1998) [1762], *Emilio*. Madrid: Alianza Editorial

Sen, A. (1970), "The impossibility of a Paretian liberal", *Journal of Political Economy*, 78 (1970), pp. 152-7

Sen, A. (1976) *Elección colectiva y bienestar social*. Madrid: Alianza Editorial.

Sen, A. (1982), "Liberty, unanimity and rights" en A. Sen, *Choice, Welfare and Measurement*, Oxford: Basil Blackwell, 1982, pp. 291-326.

Sen, A. (1995), "Minimal liberty", en A. Sen, *Nueva economía del bienestar*. Valencia: Universidad de Valencia, pp. 193-215.

Skyrms, B. (1996), *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Vallentyne, Peter (ed., 1991), *Contractarianism and Rational Choice*. Cambridge: Cambridge

University Press.
Van Hees, M. (2000), “Negative freedom and the liberal paradox”, *Rationality and Society*, 12: 335-352.
Verbeek, B. y Morris, Ch., (2004), "Game Theory and Ethics", *The Stanford Encyclopedia of Philosophy (Winter 2004 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2004/entries/game-ethics/>.