

# COOPERAR POR PRINCIPIO

## COOPERATION ON PRINCIPLE

IGNACIO SÁNCHEZ-CUENCA

*Instituto Juan March y Universidad Complutense. Madrid*

*pacho@ceacs.march.es*

### RESUMEN

En este trabajo se analiza el papel potencial de las motivaciones morales en la acción colectiva. En concreto, se sugiere que la motivación moral puede funcionar de manera distinta a las normas sociales de reciprocidad que se han estudiado en la economía del comportamiento. Se presenta una nueva función de utilidad que integra los componentes clásicos de la literatura de acción colectiva y en la que se añade un componente moral. Según este componente moral, el agente aumenta o disminuye su utilidad en función de cuánto se desvíe con respecto a la media social de cooperación. Se muestra cómo la utilidad moral puede modificar un Dilema del Prisionero en otros juegos más favorables para la cooperación. Finalmente, se presentan algunas implicaciones del modelo con respecto al efecto *crowding out* y al concepto de sacrificio personal.

### PALABRAS CLAVE ADICIONALES

Acción colectiva, Utilidad moral, Efecto *crowding out*, Sacrificio.

### SUMMARY

This article analyzes the potential role of moral motivation in collective action. It is suggested that moral motivation works very differently from the social norms of reciprocity that have been studied in behavioral economics. A new utility function, combining the traditional arguments of the literature on collective action and a moral factor, is presented. In moral terms, the agent increases or decreases its utility depending on how much it deviates from the social average level of cooperation. Moral utility may transform the Prisoner's Dilemma into a game more favourable to cooperation. Finally, some implications from the model are drawn with regard to the crowding out effect and the concept of personal sacrifice.

### ADDITIONAL KEYWORDS

Collective Action, Moral Utility, Crowding Out Effect, Sacrifice.

---

*Agradecimientos:* Numerosas personas han realizado comentarios sobre versiones anteriores de este trabajo. Estoy especialmente en deuda con Víctor Lapuente, Pablo Lledó, Adam Przeworski y John Roemer. Agradezco además los comentarios de los participantes en la sesión sobre "Modelos formales de acción colectiva" organizada por José Manuel Robles y Luis Miguel Miller, así como las críticas y sugerencias de dos evaluadores anónimos.

## INTRODUCCIÓN

Desde los años sesenta, economistas y psicólogos han realizado decenas de experimentos de laboratorio en los que se pone de manifiesto de forma rotunda y clara que los seres humanos cometen fallos de razonamiento, desviándose de lo que establecen los modelos más exigentes de racionalidad; igualmente, las personas no siempre actúan de acuerdo con los estrictos supuestos de auto-interés que se manejan en la ciencia económica. La consolidación de lo que se ha convenido en llamar “economía del comportamiento” (*behavioral economics*) ha generado un debate fascinante sobre cómo conseguir que la teoría económica se ajuste mejor a lo que hemos aprendido sobre la conducta real de los seres humanos. En general, hay cierto consenso en que es preciso incorporar en los modelos formales de la acción un abanico de motivaciones que va más allá del supuesto tradicional de auto-interés (Bowles, 2004: Cap.3; Camerer, 2003; Camerer, Loewenstein y Rabin, 2004; Sobel, 2005).

Aunque economistas y psicólogos se han centrado fundamentalmente en experimentos de laboratorio, el mundo real ofrece también innumerables ejemplos de conductas que difícilmente pueden explicarse considerando que las personas actúan en función de su propio interés, entendido éste en un sentido materialista (y muchas veces incluso monetario). Por ejemplo, sería sin duda muy forzado intentar dar cuenta de las decisiones de un terrorista suicida en términos de mero auto-interés material. Sacrificar la vida propia matando a otras personas con el fin de contribuir a una causa colectiva, ya sea la liberación de un territorio ocupado o la defensa de ciertos valores religiosos, no parece un tipo de conducta que se ajuste al comportamiento auto-interesado.

Curiosamente, aunque la inmensa mayoría de las acciones que cuestionan el principio del auto-interés ocurren en el ámbito político, la teoría política positiva, que no es sino teoría económica adaptada a los problemas que se plantean en la ciencia política, apenas si se ha hecho eco de los debates que tienen lugar en la economía del comportamiento. En cierto sentido, la ciencia política basada en la elección racional es más purista que la propia economía. Quizá por eso la introducción en la ciencia política del enfoque económico o de elección racional ha dado lugar a un enconado debate metodológico (Cox, 1999; Green y Shapiro, 1994; Friedman, 1996; Lewin, 1991; MacDonald, 2003; Mackie, 2003). El contraste entre la realidad y la teoría resulta demasiado chocante en ocasiones: sin ir más lejos, así sucede por ejemplo con algo tan sencillo como el hecho de que mucha gente vote en las elecciones de los países democráticos (Blais, 2000). Si los ciudadanos fueran auto-interesados en el sentido más estrecho del término, no acudirían a votar, puesto que se trata de una actividad costosa en la que una contribución individual tiene un impacto infinitesimal sobre el resultado agregado de las urnas. Un ciudadano racional entiende que el resultado no va a cambiar por el hecho de que él vote y por tanto no encuentra razones para incurrir en el coste de ir a votar.

Si dentro de la ciencia política hay algún campo en el que los modelos formales de auto-interés no encajan adecuadamente con lo que ocurre en los laboratorios o en el mundo real, es el de la acción colectiva, en el que lo que es bueno para todos no

coincide necesariamente con lo que es bueno para cada uno desde el punto de vista del auto-interés egoísta. El problema de la acción colectiva se formula de distintas maneras en función del instrumental analítico que se emplee: a veces se expone como un asunto de provisión subóptima de bienes públicos (Olson, 1965), a veces como un Dilema del Prisionero para  $n$  personas (Hardin, 1982)<sup>1</sup>.

Si la teoría de la elección racional resulta más cuestionable en la ciencia política que en la teoría económica se debe, entre otras razones, al hecho de que los problemas de acción colectiva son mucho más frecuentes en el ámbito político que en el económico. Aquí van algunos ejemplos: la participación en una protesta política (una manifestación, una huelga, una revolución, etc.), la afiliación a organizaciones políticas (sindicatos, partidos, grupos de interés), el uso de recursos comunales, el reciclaje de residuos, la donación de sangre o de órganos, la donación caritativa, etcétera. En casi todos los casos, se observa mayor grado de cooperación del que cabe esperar según la teoría.

Como apunté antes, lo anómalo es que a pesar de todos estos desmentidos empíricos, la teoría de la acción colectiva haya mostrado tan poco interés por los avances realizados en la economía del comportamiento. Mi propósito en este trabajo consiste en analizar una función de utilidad original en la que se contempla la posibilidad de que las personas cooperen en virtud de ciertas motivaciones morales. Por un lado, la función de utilidad que propongo integra buena parte de la literatura estándar sobre acción colectiva (función de producción del bien colectivo, incentivos selectivos, “la seguridad de los números” —*safety in numbers*). Por otro, incorpora un nuevo argumento en el seno de la función que recoge la idea de que algunas personas actúan movidas por principios morales, tratando de ser fieles a los valores que esos principios encarnan.

Este componente moral de la función de utilidad es bastante distinto a las normas de justicia (*fairness*) y reciprocidad que se han propuesto en el campo de la economía del comportamiento para dar cuenta de los resultados sorprendentes que se observan en experimentos con el juego del Ultimátum y sus variantes (un buen ejemplo de este tipo de trabajos es Bicchieri, 2006). Mientras que las funciones de utilidad basadas en normas sociales tratan de explicar ciertas regularidades que se producen en las interacciones sociales, en el sentido de que hay unas normas que inducen a las personas a actuar de una determinada manera, en coherencia con lo que socialmente se espera de ellas, en mi caso me centro en la posibilidad de comportamientos extremos e inusuales, como sacrificios muy costosos, que están en mayor medida motivados por convicciones personales que por lo que se considera correcto desde el punto de vista social. Con ello no pretendo dar a entender que los modelos de economía del comportamiento basados en normas sociales son inapropiados: tan sólo quiero subrayar que son parciales, puesto

---

<sup>1</sup> Ambas formulaciones son casi equivalentes, aunque hay una diferencia importante. Mientras que en la teoría de los bienes públicos el tamaño del grupo es una variable crucial, en teoría de juegos el Dilema del Prisionero estipula la no cooperación incluso en el grupo más pequeño posible, formado por dos personas. Véase Sandler (1992: 38-49).

que no cubren, como es lógico, todas las posibilidades que se abren a la persona cuando se enfrenta a un dilema de acción colectiva.

Por un lado, la función de utilidad que presento puede reproducir los mismos resultados que la literatura sobre la función de producción del bien colectivo, sólo que a través de una vía alternativa. En lugar de explicar la existencia de equilibrios cooperativos según las características del bien colectivo que se persigue, muestro que esos mismos equilibrios se pueden alcanzar mediante variaciones en los motivos que mueven a la gente a actuar.

Por otro lado, aunque no intento comprobar si la función de utilidad que aquí sugiero es empíricamente superior a otras, sí creo poder mostrar, mediante un análisis de posibles extensiones y desarrollos, que esta función recoge e integra diversos aspectos de la acción colectiva a los que hasta el momento se les ha prestado más bien poca atención en los análisis formales (no así en los aplicados o empíricos). En concreto, en la última parte de este artículo desarrollo la función de utilidad para poder explicar la idea de sacrificio personal, cuya importancia en la acción colectiva es obvia.

La justificación que realizo de la función de utilidad se basa por tanto en consideraciones filosóficas y metodológicas, y en lo que Whewell (1984) llamó *consilience of inductions*, es decir, la capacidad de una teoría para unificar fenómenos que hasta el momento habían recibido explicaciones inductivas e independientes unas de otras.

## **EL PAPEL DE LAS MOTIVACIONES EN LA TEORÍA DE LA ACCIÓN COLECTIVA**

Las teorías de elección racional sobre acción colectiva comparten una cierta desconfianza de partida hacia cualquier solución basada en las preferencias o motivaciones del agente. Se considera que no resuelve nada suponer que el agente pueda tener “preferencias cooperativas”, es decir, preferencias que induzcan la cooperación y que eliminen la tensión entre lo que es bueno para todos y lo que es bueno para cada uno. Una solución así parece trivial o tautológica.

La razón última de esta desconfianza hacia las soluciones basadas en preferencias o motivos es metodológica. Aunque las preferencias sean el motor de la acción dentro de la teoría de la elección racional, son también entidades inobservables y por tanto sospechosas en una explicación. Lo que observamos es nada más la acción, no sus motivaciones. De ahí que a pesar de que la teoría no estipule el contenido de las preferencias, en la práctica, en los modelos formales que pretenden tener aplicación empírica, siempre haya un supuesto de auto-interés o egoísmo. Sólo desde ese supuesto la teoría tiene “parsimonia” suficiente para producir consecuencias que puedan someterse al tribunal de la experiencia.

No es que la teoría de la elección racional descarte la existencia de preferencias favorables al comportamiento cooperativo, sino que lo que se admite como explicación satisfactoria requiere un supuesto previo de egoísmo. Olson contempla la posibilidad de que pueda haber preferencias morales que favorecen la cooperación (evitando un sentimiento de culpa o de pérdida de auto-estima en caso de que el agente se comporte como un *freerider*), pero rechaza esta vía expeditivamente:

Primero, no es posible obtener una prueba empírica de la motivación que hay tras la acción de la persona: no es posible decir con seguridad si una persona actuó en un caso concreto por razones morales o por cualesquiera otras. Confiar en explicaciones morales podría hacer que la teoría no pudiera ponerse a prueba. Segundo, no hace falta una explicación así, puesto que hay otro tipo de explicaciones que resultan suficientes para toda la acción colectiva que se analiza [en este libro]. (Olson, 1965: 61)

Nótese que las razones de Olson son puramente metodológicas, pues no descarta la posibilidad de que existan las preferencias cooperativas. Su argumento es que incluso aunque esas preferencias existieran, no tendrían cabida en la explicación. Las explicaciones basadas en preferencias egoístas son “suficientes” para Olson.

La literatura posterior sobre acción colectiva no se ha desviado demasiado de esta postura. Las motivaciones no son una variable fundamental para explicar la cooperación. No obstante, cuando se da por supuesto en diversos estudios que el juego de referencia para analizar la acción colectiva no es el Dilema del Prisionero, sino el Juego de la Seguridad (*Assurance game*) o algún otro similar (Chong, 1991; Lichbach, 1996), se está asumiendo, de forma más o menos explícita, que las preferencias de las personas son tales que ante la cooperación de los demás, prefieren ellos mismos cooperar que defraudar. Por qué sea ese el orden de preferencias no se discute abiertamente.

Lo mismo sucede con la abundante literatura de los modelos de umbral (Granovetter, 1978; Heckathorn, 1996; Lohman, 1994; Macy, 1991; Marwell y Oliver, 1993; Schelling, 1978). En casi todos estos modelos hay siempre un grupo inicial de cooperadores incondicionales que pone en marcha la cooperación de los cooperadores condicionales. La existencia de esa masa crítica inicial de cooperadores incondicionales se da por buena sin mayores averiguaciones, como si no planteara un problema el supuesto de que hay personas libres de los dilemas de acción colectiva que atenazan a la mayoría de la gente.

En parte se ha intentado superar este tipo de problemas haciendo recaer todo el peso de la explicación en las características del bien colectivo que se persigue. Esto es lo que propone Heckathorn (1996) en su modelo: el tipo de interacción estratégica en el que participan las personas afectadas por un dilema de acción colectiva depende enteramente no tanto de las preferencias o motivaciones cuanto del tipo de bien colectivo. Cada bien colectivo tiene su propia función de producción, es decir, una forma específica de agregación de actos individuales, que da lugar a distintos tipos de juegos. Heckathorn, sin embargo, no proporciona apenas ejemplos de bienes colectivos concretos que por tener distintas funciones de producción generen distintos tipos de dilemas de acción colectiva. Más bien supone que si la gente coopera, ha de ser porque el bien colectivo es tal que la estructura de interacción resultante corresponde, por ejemplo, a un juego de la Seguridad y no a un Dilema del Prisionero. En este sentido, puede decirse incluso que la forma específica de la función de producción resulta tan inobservable como las preferencias mismas de las personas.

No quiero dar a entender con estos comentarios que la función de producción del bien colectivo sea irrelevante para explicar la dinámica de la acción colectiva: de hecho,

dicha función desempeña un papel importante en el modelo que presento en este trabajo. No obstante, la función de producción, por sí misma, no es suficiente. Dada la misma tecnología de producción de dos bienes colectivos, es posible que en un caso observemos cooperación y en otro no. Pensemos en dos bienes colectivos que tienen una producción semejante: la (ausencia de) inflación y la (ausencia de) contaminación medioambiental. En ambos casos el grupo es muy grande y en ambos casos parece que la función de producción es acelerativa (las primeras contribuciones producen una cantidad muy pequeña del bien colectivo, las contribuciones posteriores van produciendo cada vez mayor cantidad del bien). A pesar de estas similitudes, la solución del problema de acción colectiva será en cada caso muy distinta dados los distintos tipos de motivaciones que cada problema implica. En el caso de la inflación, lo lógico es considerar que la gente no va a cooperar en absoluto (no va a reducir su consumo individual para que disminuya la inflación), por lo que hay que activar soluciones centralizadas (gravar el consumo, aumentar los tipos de interés, etc.). En cambio, en el caso de la contaminación es posible que la concienciación de la gente y la intervención de grupos ecologistas sean capaces de inducir a muchas personas a cooperar. Estas personas no verán el problema de la contaminación como algo ajeno a sus vidas, hasta el punto de que muchos pueden llegar a sentir una contradicción o un conflicto entre desear un medioambiente limpio y no cooperar en la consecución del mismo. Si esta contradicción es suficientemente fuerte, no cabe descartar que la persona acabe decidiendo que le compensa cooperar. Este tipo de motivación moral, que quizá no sea dominante en la sociedad, pero que sin duda existe entre amplios grupos, no es probable que fructifique en el caso de la inflación, pues la contención de los precios como bien colectivo tiene una conexión muy lejana con los asuntos moralmente relevantes para la gente.

Lo que esta sencilla ilustración pretende es mostrar que manteniendo constante la tecnología de producción del bien colectivo, podemos encontrar grados muy diferentes de cooperación en función de la relevancia moral que tenga el bien colectivo para las personas que están afectadas por el problema de acción colectiva. En la sección siguiente desarrollo esta idea.

## **PREFERENCIAS Y PRINCIPIOS**

La teoría de la elección racional puede entenderse como una especie de consecuencialismo individualista, como consecuencialismo positivo en vez de prescriptivo. El consecuencialismo es una doctrina filosófica según la cual la persona elige aquella acción que contribuye a producir el mejor estado de cosas de acuerdo con un criterio de valoración impersonal (Scheffler, 1982). La variante mejor conocida de consecuencialismo es el utilitarismo, cuyo criterio de valoración impersonal es el bienestar social. En la medida en que el agente actúe para conseguir lo mejor no según un criterio impersonal, sino según su propio bienestar, el consecuencialismo se transforma en teoría de la elección racional, pasando de ser una teoría prescriptiva a una teoría positiva. Así, de la misma

manera que el consecuencialismo se opone al deontologismo, la teoría de la elección racional se opone a aquellas otras teorías que tratan de explicar la acción mediante reglas, normas o principios que el agente sigue al margen de las consecuencias que tengan para su bienestar.

Según Sen y Williams (1983: 5), los problemas del consecuencialismo derivan de una restricción artificial en la información que la teoría procesa sobre los agentes, pues de acuerdo con la teoría sólo importan las consecuencias de las acciones. La teoría pasa por alto el hecho de que las personas son distintas entre sí (Rawls, 1971), y que sus elecciones en cada caso dependen en buena parte de su carácter moral, es decir, del conjunto de proyectos, valores y compromisos que da sentido a sus acciones (Williams, 1981). Es fundamental dejar claro que dicha constitución moral no tiene por qué identificarse con la búsqueda de lo bueno, lo correcto o lo justo. La moral del imperativo categórico es sólo una posibilidad entre otras muchas. Cuando una persona actúa para defender su honor, aunque sea el honor de ser un miembro de la mafia o de una pandilla juvenil, cabe entender que está actuando en función de su constitución moral. La fuente de las obligaciones es completamente irrelevante. Lo que importa es que una vez que la persona hace suyo un cierto proyecto de vida (ser un miembro de la mafia, realizar en la tierra el reino kantiano de los fines), de tal proyecto se siguen unas obligaciones sobre qué tipo de decisiones ha de tomar la persona. Como ha señalado Williams (1981: Cap.3), el consecuencialismo no otorga valor alguno a la integridad: si la búsqueda de la mejor consecuencia exige actuar en contra de lo que establece el carácter moral de la persona, ésta ha de renunciar a su integridad moral en beneficio del mejor estado de cosas posible.

Pues bien, creo que la teoría de la elección racional opera en este sentido igual que el consecuencialismo: la información que recoge sobre el agente es demasiado pobre. De hecho, el agente que describe la teoría, al que Sen (1977) se refiere como un *rational fool*, es poco más que una ordenación de preferencias en función de las cuales actúa. No se incorpora consideración alguna sobre su carácter moral. Parte de las dificultades de la teoría para hacer frente a los dilemas de la acción proceden de esta descripción en exceso estilizada del agente.

La persona, a la hora de actuar, tiene en cuenta de qué modo su acción satisface sus preferencias, tal y como analiza la teoría de la elección racional; pero, además, la persona examina hasta qué punto la acción concuerda o se ajusta a las demandas que se derivan de su carácter moral. En multitud de ocasiones, el carácter moral puede que no intervenga: por ejemplo, cuando el agente decide usar sus ahorros en unos fondos de inversión. Pero en otras ocasiones puede ser crucial: por ejemplo, cuando el agente decide reciclar los restos de los bienes que consume. Mientras que la personalidad del agente rara vez queda comprometida por sus decisiones de inversión, sí es probable que intervenga en la cuestión del reciclaje, pues la decisión que tome a propósito de esta cuestión puede revelar qué tipo de persona es (si se preocupa o no por el medioambiente, si actúa en consonancia con sus preocupaciones, etc.).

Para ir precisando: del carácter moral del agente surgen obligaciones que pueden entrar en conflicto con las preferencias estándar de la elección racional. El agente, debido

a sus proyectos, valores y compromisos, desarrollará ciertas creencias sobre lo que debe hacer que no tienen por qué coincidir con lo que quiere hacer desde el punto de vista de sus preferencias. En el fondo, no se trata más que de recoger la idea de que el agente puede pasar por conflictos internos en los que chocan consideraciones sobre lo que cree que *debe hacer* y consideraciones sobre lo que *quiere hacer*. Este tipo de conflictos están ausentes en la teoría estándar de la elección racional: se supone, por ejemplo, que si un trabajador está a favor de una subida de sueldo y actúa como un *freerider* cuando sus compañeros llevan a cabo una huelga en demanda de una subida salarial, el trabajador estará en la mejor situación posible. Sin embargo, no hay que tener un sentido de la moralidad muy desarrollado para entender que no participar en la huelga mientras los compañeros lo hacen puede provocar una cierta incomodidad a la persona más allá de la presión social que el grupo ejerce sobre el individuo: la persona puede verse a sí misma como alguien insolidario, egoísta y mezquino por abandonar a sus compañeros cuando éstos luchan por algo de lo que él mismo se beneficiará.

De lo que se trata, por consiguiente, es de analizar el conflicto entre dos tipos de razones para actuar: razones que proceden del carácter moral de la persona y razones que reflejan sus preferencias materiales. Por ello, voy a distinguir dos tipos de utilidad, la utilidad que corresponde al auto-interés material y la utilidad que corresponde al conjunto de principios morales de la persona. Los principios pueden caracterizarse como criterios sobre qué elecciones hay que realizar en distintas situaciones. El principio de honestidad, por ejemplo, consiste en que en todos aquellos casos en que la persona tenga la oportunidad de elegir entre decir una verdad o decir una mentira, elija decir la verdad.

La utilidad moral refleja lo que podríamos llamar la dimensión “electiva” de las alternativas que constituyen el problema de decisión. En efecto, las alternativas, en la medida en que son alcanzables a partir de las elecciones o decisiones del agente, pueden ser evaluadas en virtud de su significado “electivo”, esto es, de las implicaciones que tiene el hecho de elegir una alternativa frente a otra dado el carácter moral del agente. Con otras palabras: la utilidad moral proviene de la conexión entre la elección de cada alternativa y los proyectos y valores del agente. De esta forma, se consigue internalizar el propio acto de elegir dentro del cálculo del agente (Sen, 2002: 161).

## UN MODELO DE MOTIVACIÓN MORAL

Sea un grupo de personas  $\{1, \dots, N\}$ . Hay un bien colectivo que proporciona una utilidad  $V$  a todas las personas del grupo, al margen de que hayan cooperado o no en su producción. Las estrategias son discretas: cooperar o defraudar. En concreto, la persona  $i$  realiza una acción  $C_i$ , que tiene los siguientes valores:

$$C_i = \begin{cases} 1 & \text{si el agente coopera} \\ 0 & \text{si el agente defrauda} \end{cases}$$

Siguiendo la literatura formal sobre acción colectiva, la parte de la función de utilidad que corresponde al auto-interés tiene tres elementos: (i) la función de producción del bien colectivo, que establece la cantidad del bien colectivo  $V$  del que disfruta el grupo si un cierto número de personas del grupo coopera; (ii) los costes de la cooperación y su posible disminución gracias a la existencia de incentivos selectivos; y (iii) las posibles consecuencias negativas de la cooperación.

En cuanto a la cantidad producida del bien colectivo, adopto tal cual la función de producción propuesta por Heckathorn (1996), multiplicada por el valor del bien colectivo:

$$\left( 1 - \left( \frac{N - \sum C_i}{N} \right)^F \right)_i V$$

Aquí la clave reside en el exponente  $F$ , responsable de que la función de producción sea acelerativa o decelerativa. Si  $F < 1$ , la función se acelera, es decir, las primeras contribuciones producen una cantidad muy pequeña del bien colectivo, mientras que las contribuciones posteriores van produciendo cada vez mayores cantidades. Si  $F > 1$ , la función se frena: cada contribución adicional tiene un impacto menor sobre la producción del bien colectivo. Si  $F = 1$ , la función es lineal o aditiva.

Como acabo de señalar, distingo dos tipos de costes. Por un lado, está el coste de la cooperación en términos de la renta o del tiempo que ha de sacrificar la persona por el hecho de cooperar. Este coste lo represento como  $KC_i$ ; es decir, si el agente defrauda no incurre en coste alguno, si coopera asume un coste  $K$ . Podemos además introducir un parámetro  $\beta \in [0, 1]$  que, multiplicado por  $KC_i$ , mida la presencia de incentivos selectivos positivos por cooperar. Cuanto menor es  $\beta$ , más grande es el incentivo selectivo, es decir, menor es el coste de cooperar para el agente. Si  $\beta = 1$ , no hay incentivos selectivos. En suma, el coste de la cooperación lo represento así:

$$-\beta KC_i$$

Por otro lado, tenemos un coste adicional  $J$  que tiene que ver con las consecuencias de la cooperación. Aparte del coste que supone cooperar en términos de renta o tiempo, el agente que coopera puede tener que pagar un coste adicional en ciertos contextos de acción colectiva: un trabajador que como consecuencia de su participación en una huelga es despedido, un ciudadano que como consecuencia de su protesta contra un régimen dictatorial es detenido, torturado o incluso asesinado. Ese coste se paga probabilísticamente: no es seguro que se vaya a realizar en la práctica. La probabilidad de incurrir en este coste depende en lo fundamental del número de cooperadores en el grupo, es decir, de la seguridad que proporciona el número de cooperadores (*safety in numbers*). Cuanta más gente coopera, menos probable es que este coste se materialice. Evidentemente, hay muchos bienes colectivos en los que este tipo de coste adicional no se plantea ( $J = 0$ ). Así sucede cuando la acción colectiva no entraña riesgos personales, como cuando un grupo

de interés presiona a un gobierno para conseguir una mejora fiscal. Este componente de la función de utilidad lo represento de este modo:

$$-\left(\frac{N - \sum C_i}{N}\right) J C_i$$

Estos tres elementos integran de forma muy esquemática los principales factores explicativos de la cooperación que se han manejado en la literatura formal de acción colectiva: miden los beneficios y costes de la cooperación en relación al auto-interés de la persona. Ahora quiero añadir un componente moral. Para ello, hay que resolver dos cuestiones preliminares:

(i) ¿Cómo operacionalizar la idea de que la persona actúa de acuerdo con principios morales? Los principios, por su propia naturaleza, son abstractos y generales. No está claro qué consecuencias prácticas tienen en una situación concreta. La persona necesita un punto de referencia para poder determinar si su nivel de cooperación es alto o bajo. Pensemos en preferencias morales medioambientales. El agente cree que debe cooperar en todo lo que corresponda a la preservación del medioambiente. ¿Qué implicaciones se siguen exactamente de estas preferencias? A mi juicio, lo más razonable en este caso es tomar como punto de referencia el nivel medio de cooperación en el conjunto de la sociedad, o si se quiere, en el grupo social en el que actúa principalmente el agente<sup>2</sup>. Las preferencias morales se concretan en el medio social en el que se desenvuelven las personas.

Tomando como punto de referencia la media social de cooperación con respecto al bien colectivo  $V$ , la utilidad moral sigue una pauta algo diferente a la utilidad que proviene del cumplimiento de las normas sociales (Bicchieri, 2006). Mientras que cuando hay una norma social, la persona obtiene el máximo de utilidad cuando su comportamiento es coherente con dicha norma, en el caso de los principios morales podemos decir que la persona simplemente cumple con su obligación cuando su cooperación coincide con la media social. En cambio, tiene una ganancia de utilidad moral cuando coopera por encima de la media y una pérdida cuando coopera por debajo de la misma. Si la persona llega más allá que el resto movida por sus convicciones morales, quiere decir que su comportamiento es más puro o íntegro que el del resto; por el contrario, si la persona no llega a la media social, su comportamiento será oportunista, mezquino o indigno, y la persona tendrá que pagar un coste moral derivado de la pérdida de integridad producida. Eamon Collins fue un miembro de la organización terrorista IRA. En sus memorias relata sus inicios en el movimiento Republicano en estos términos:

---

<sup>2</sup> No entro a discutir la cuestión esencial de si importan más lo vínculos sociales fuertes o débiles; véase por ejemplo Chwe (1999), Granovetter (1973), Macy (1991).

Aquellas marchas [se refiere a manifestaciones en apoyo a las campañas de protesta de los presos republicanos a finales de los 70] eran una especie de introducción continua al movimiento Republicano. Me hacían sentir culpa y vergüenza: me sentía culpable por llevar una vida tan cómoda y privilegiada; y me sentía también avergonzado por lo poco que estaba haciendo por la lucha (Collins, 1997: 64).

En esta ilustración puede verse claramente cómo un grado muy elevado de cooperación de los demás arrastra a la persona a tratar de ponerse a la misma altura. La utilidad moral, por tanto, depende de la diferencia entre la acción del agente y la media social de cooperación sin contar la propia acción del agente,  $C_i - \bar{C}$ , donde:

$$\bar{C} = \frac{\sum_{j \neq i} C_j}{N - 1}$$

Si la diferencia es positiva (coopera más que la media), el agente obtiene utilidad moral positiva; si es negativa (coopera menos que la media), paga un coste en términos de integridad moral.

Es interesante comparar esta especificación de la función de utilidad con la del principio de reciprocidad. En ambos casos la media social de cooperación es crucial, pero de manera diferente en cada caso. En el modelo inicial de reciprocidad de Sugden (1984), "el principio de reciprocidad nunca requiere que el agente contribuya más que la otra gente del 'grupo'" (p.775). La media de cooperación del grupo marca el nivel de cooperación que debe realizar la persona. En cambio, en el modelo que estoy presentando aquí, la media social constituye el punto a partir del cual se genera utilidad positiva o negativa, de tal manera que cuanto mayor es la desviación entre la cooperación de la persona y la cooperación del grupo, mayores las ganancias o las pérdidas de utilidad moral.

(ii) La diferencia entre el nivel de cooperación individual y el nivel social medio no es una medición absoluta de utilidad moral. Tan sólo refleja hasta qué punto la acción de la persona coincide con la del grupo. La importancia del grado de desviación de la persona con respecto al grupo depende de dos factores adicionales. Por una parte, de la relevancia moral que la persona atribuya al bien colectivo. Esto se mide mediante un parámetro  $\alpha_i \geq 0$ . Si  $\alpha_i = 0$ , la persona  $i$  no atribuye importancia moral alguna al bien colectivo y por tanto será irrelevante la desviación de  $i$  en relación al grupo.

Por otra parte, y quizá esto resulte más discutible, voy a considerar que la utilidad moral depende también de cuánto haya en juego. Si el bien colectivo tiene una importancia muy baja, la utilidad moral será por fuerza bastante limitada. La idea es que la moralidad desempeña un papel decisivo cuando lo que hay en juego es realmente importante. Tanto el heroísmo como su opuesto, la cobardía o la traición, surgen a propósito de bienes colectivos de gran relevancia, como una revolución, una protesta política en una dictadura, etc., pero no tanto en situaciones ordinarias. De hecho, podemos encontrarnos con el caso de alguien que no tenga reparo en mentir en pequeños asuntos sin mayor

trascendencia, y, sin embargo, cuando llega la hora de la verdad y lo que se ventila es un asunto fundamental, esa persona sea fiel al principio moral de ser sincero.

Un modo sencillo de representar esta idea consiste en multiplicar la distancia entre la cooperación individual y la del grupo por el valor  $V$  del bien colectivo. Más adelante discutiré una formulación alternativa algo más compleja que nos permitirá modelizar la idea de sacrificio personal, pero por el momento esta especificación general es suficiente.

Juntando por tanto los tres componentes de la utilidad moral, ésta puede formularse como sigue:

$$\alpha_i (C_i - \bar{C})V$$

Ahora podemos representar la función de utilidad completa:

$$(1) \quad U_i(C_i) = \left( 1 - \left( \frac{N - \sum C_i}{N} \right)^F \right) V - \beta K C_i - \frac{N - \sum C_i}{N} J C_i + \alpha_i (C_i - \bar{C})V$$

Dada la complejidad algebraica de (1), realizaré algunos supuestos simplificadores con el propósito de hacer más manejables e inteligibles las expresiones matemáticas que se van a utilizar. En adelante, consideraré  $J = 0$ , limitándome por tanto a bienes colectivos en los que no hay otro coste material que el coste oportunidad de cooperar. Si  $J > 0$ , los efectos que se analizan a continuación se dan igualmente, salvo que se produce un cambio de escala: cuanto más alto  $J$ , más intensas han de ser las convicciones morales para que una persona coopere cuando los demás no cooperan.

En términos generales,  $i$  cooperará cuando  $U_i(C_i = 1) > U_i(C_i = 0)$ , es decir, si se da el caso que

$$(2) \quad \left[ 1 - \left( \frac{N - \sum_{j \neq i} C_j - 1}{N} \right)^F \right] V - \beta K + \alpha_i V (1 - \bar{C}) > \left[ 1 - \left( \frac{N - \sum_{j \neq i} C_j}{N} \right)^F \right] V + \alpha_i V (0 - \bar{C})$$

Podemos expresar esta inecuación en términos del peso  $V_i$  de los principios morales del agente  $i$  con respecto al bien colectivo de valor  $V$  como sigue:

$$(3) \quad \alpha_i^* > \frac{\beta K}{V} - \left[ \left( \frac{N - \sum_{j \neq i} C_j}{N} \right)^F - \left( \frac{N - \sum_{j \neq i} C_j - 1}{N} \right)^F \right]$$

La expresión (3) especifica el umbral de relevancia moral a partir del cual  $i$  coopera. El umbral depende positivamente del coste de la cooperación y negativamente de la

importancia moral del bien colectivo. Por otro lado, cuanto mayor impacto en la producción del bien colectivo tenga la cooperación de  $i$ , menor podrá ser la importancia moral del bien colectivo. No obstante, la expresión (3) no es suficientemente precisa, ya que el valor de  $\alpha_i^*$  depende del número de cooperadores en el grupo y, además, el efecto del número de cooperadores depende a su vez de cuál sea la forma de la función de producción. La siguiente expresión resuelve este problema:

$$(4) \quad \alpha_i^* = \begin{cases} \beta K / V - \left(1 - \left(N - \frac{1}{N}\right)^F\right) & \text{si } F < 1 \\ \beta K / V - \frac{1}{N} & \text{si } F = 1 \\ \beta K / V - \left(\frac{1}{N}\right)^F & \text{si } F > 1 \end{cases}$$

Esta formulación del umbral, al no depender del número de cooperadores, proporciona los valores máximos de  $\alpha_i$  que, si se satisfacen, garantizan la cooperación de  $i$  al margen de lo que hagan los demás.

Resulta sencillo demostrar a continuación que mediante la función de utilidad que he propuesto cabe producir los cuatro juegos más relevantes para el análisis de la acción colectiva mediante variaciones en el parámetro  $\alpha_i$ , de manera similar a como hace Hec-kathorn (1996) con el parámetro  $F$  de la función de producción. Hay por tanto una cierta equivalencia observacional entre los dos modelos, en el sentido de que puede que las personas estén jugando un juego de la Seguridad y no un Dilema del Prisionero tanto a causa del parámetro  $F$  como del parámetro  $\alpha_i$ .

Supondré el caso más simple posible, en el que  $N = 2$ ,  $\beta = 1$  (no hay incentivos selectivos),  $J = 0$  (los costes tienen que ver solamente con el coste de la cooperación, no con el coste esperado de sus consecuencias) y la importancia moral del bien colectivo es la misma para ambos agentes,  $\alpha$ , con lo que no necesitamos subíndices para identificar al agente. Nótese que en estas circunstancias un tanto especiales, la media social de cooperación relevante para  $i$  es simplemente  $\bar{C} = C$ . Dado que las estrategias de los agentes son binarias, los resultados o alternativas posibles son cuatro: la cooperación mutua ( $CM$ ), la cooperación unilateral ( $CU$ ), la defección mutua ( $DM$ ) y la defección unilateral ( $DU$ ). En el cuadro 1 se han representado los cuatro resultados desde el punto de vista del jugador  $i$  (en filas).

Cuadro 1.  
El juego genérico de la cooperación

|   | C                             | D                               |
|---|-------------------------------|---------------------------------|
| C | Cooperación mutua ( $CM$ )    | Cooperación unilateral ( $CU$ ) |
| D | Defección unilateral ( $DU$ ) | Defección mutua ( $DM$ )        |

C=cooperar; D=defraudar

Supongamos en primer lugar que el bien colectivo que define el problema de elección no tiene implicaciones morales de ningún género,  $\alpha = 0$ . Tenemos un Dilema del Prisionero (DP) cuando  $DU > CM > DM > CU$ . Dada la función de producción del bien colectivo, esto requiere que el valor del bien colectivo, el coste de la cooperación y el exponente  $F$  de la función de producción sean tales que se cumpla esta inecuación:

$$(5) \quad V - K < \frac{V}{2^F}$$

Voy a suponer que (5) se cumple siempre. Con otras palabras, supondré que, por defecto, cuando no hay utilidad moral en la interacción entre los agentes, el juego de referencia es el DP. Es el juego menos favorable para la cooperación, y el que mejor se aproxima al problema de los bienes públicos. Lo que a continuación quiero demostrar es que manteniendo constante la parte “económica” del juego que genera un DP, la introducción de la utilidad moral puede transformar el DP en alguno de los tres juegos que se especifican en el cuadro 2, el juego de la Seguridad (*Assurance Game*), el juego del Gallina, o el juego Privilegiado. De lo que se trata de ver es de especificar cómo el DP original puede evolucionar hacia alguno de estos tres juegos en función de la utilidad moral.

Cuadro 2.  
*Cuatro juegos*

| Preferencias        | Equilibrios de Nash (con estrategias puras) | Juego                 |
|---------------------|---|-----------------------|
| $DU > CM > DM > CU$ | (D, D)                                      | Dilema del Prisionero |
| $CM > DU > DM > CU$ | (C, C) & (D, D)                             | Seguridad             |
| $DU > CM > CU > DM$ | (C, D) & (D, C)                             | Gallina               |
| $CM > DU > CU > DM$ | (C, C)                                      | Juego Privilegiado    |

Pues bien, el juego continúa siendo un DP a pesar de la introducción de utilidad moral cuando<sup>3</sup>:

$$(6) \quad \alpha < \min \left( \frac{K - V/2^F}{V}, \frac{K - V + V/2^F}{V} \right)$$

<sup>3</sup> El procedimiento es el siguiente: se calculan las utilidades correspondientes a las cuatro situaciones posibles,  $U_i(0, 0)$ ,  $U_i(1, 1)$ ,  $U_i(0, 1)$  y  $U_i(1, 0)$ , según la función de utilidad, y después se despeja con respecto a  $\alpha$  en la inecuación  $U_i(0, 1) > U_i(1, 1) > U_i(0, 0) > U_i(1, 0)$ . Se repite el mismo cálculo para las otras tres inecuaciones de los otros tres juegos.

Nótese que si la función de producción es lineal,  $F = 1$ , ambas cantidades son iguales. De hecho, el parámetro  $F$  determina que el juego pueda evolucionar hacia el Gallina o la Seguridad. Si  $F > 1$ , es evidente que

$$\frac{V}{2^F} < V - \frac{V}{2^F}$$

y viceversa. Depende por tanto de la función de producción del bien colectivo. En concreto, si  $F < 1$ , el juego puede transformarse en el de la Seguridad mediante la introducción de utilidad moral (puesto que la función de producción es acelerativa, hace falta mucha cooperación para conseguir el bien colectivo, tal y como recoge el juego de la Seguridad). Si, por el contrario,  $F > 1$ , el juego puede transformarse en Gallina (si la función es decelerativa, una pequeña contribución inicial producirá una gran cantidad del bien colectivo, con lo que el grupo será "privilegiado" en el sentido de Olson: compensará la cooperación unilateral y por tanto nos situaremos en el juego del Gallina). Con más detalle, el DP original se transforma en el juego de la Seguridad cuando:

$$(7) \quad \frac{K - V/2^F}{V} < \alpha < \frac{K - V + V/2^F}{V}; (F < 1)$$

Y se transforma en el Gallina cuando:

$$(8) \quad \frac{K - V + V/2^F}{V} < \alpha < \frac{K - V/2^F}{V}; (F > 1)$$

Finalmente, el juego resultante es Privilegiado cuando:

$$(9) \quad \alpha > \max \left( \frac{K - V/2^F}{V}, \frac{K - V + V/2^F}{V} \right)$$

Sirva el siguiente ejemplo para ilustrar estas transformaciones. Supongamos los siguientes parámetros de este juego: hay un bien colectivo cuyo valor es  $V = 10$ , el coste de cooperar es  $K = 8$ , y la función de producción es acelerativa,  $F = 0,8$ . Cuando no hay utilidad moral,  $\alpha = 0$ , el juego resultante es un DP, pues se cumple que

$$V - K < \frac{V}{2^F}$$

Los pagos de este juego aparecen en el cuadro 3.

Cuadro 3.  
Un DP con  $V = 10$ ,  $K = 8$ ,  $F = 0,8$ ,  $\alpha = 0$

|   | C           | D           |
|---|-------------|-------------|
| C | 2; 2        | -3,74; 4,26 |
| D | 4,26; -3,74 | 0; 0        |

¿De qué modo la utilidad moral puede transformar este juego en un juego de la Seguridad? De acuerdo con la inequación (7), cuando suceda lo siguiente:

$$\frac{8 - 10/2^{0,8}}{10} = 0,22 < \alpha < \frac{8 - 10 + 10/2^{0,8}}{10} = 0,37$$

Por tanto, siempre que la relevancia moral del bien colectivo esté entre 0,45 y 0,75, el juego será el de la Seguridad, con dos equilibrios de Nash, uno en el que los dos jugadores cooperan y otro en el que ambos defraudan. Supongamos por ejemplo que  $\alpha = 0,3$ . En tal caso, el DP del cuadro 3 pasa a ser el juego de la Seguridad del cuadro 4.

Cuadro 4.  
Un juego de la Seguridad con  $V = 10$ ,  $K = 8$ ,  $F = 0,8$ ,  $\alpha = 0,3$

|   | C           | D           |
|---|-------------|-------------|
| C | 2; 2        | -0,75; 1,25 |
| D | 1,25; -0,75 | 0; 0        |

Este ejemplo pone de manifiesto la relevancia de la utilidad moral. Manteniendo constantes las preferencias económicas y la función de producción del bien colectivo, los equilibrios del juego pueden variar según la relevancia moral que los agentes atribuyan al bien colectivo. Cuanta mayor importancia o relevancia moral tenga el bien colectivo, más probable es la cooperación.

En este modelo el hecho de que algunas personas estén dispuestas a cooperar si otros cooperan no es un punto de partida, sino un resultado, una consecuencia de introducir explícitamente motivaciones morales en el seno de la función de utilidad. Los análisis empíricos han demostrado que en muchos casos el juego que mejor representa el dilema de la acción colectiva es el de la Seguridad y no el Dilema del Prisionero. Esto requiere una explicación, pues si la situación se define en términos estrictos de auto-interés material, el juego debería ser el Dilema del Prisionero. Si no es así se debe a la función de producción y a los principios morales.

## EXTENSIONES

A continuación expongo dos extensiones del modelo que permiten integrar diversas cuestiones que han surgido en el análisis de la acción colectiva.

### Tipos de cooperación según el tipo de bien colectivo

El parámetro  $\alpha_i$  mide la relevancia moral del bien colectivo para cada persona. Se trata de una característica individual no observable. Por tanto, puede parecer que no hay manera de poner a prueba el modelo. Sin embargo, sabemos que el contexto social puede determinar ciertas variaciones en los valores de este parámetro, de forma que podamos establecer enunciados generales al respecto.

Podemos suponer que hay bienes colectivos con nula o muy baja relevancia moral para las personas (lo que Etzioni (1988: 101) llama zonas de indiferencia moral). Cuando así sea, se hará un uso intensivo de los incentivos selectivos. En cambio, en los bienes colectivos con fuerte carga moral, los incentivos selectivos tendrán menos importancia y en cambio los principios morales serán decisivos.

Se ha sugerido en la literatura (Frey, 1994; Wilson, 1995) que los bienes colectivos económicos tienden a tener baja relevancia moral, mientras que los bienes colectivos políticos tienden a tener media o alta relevancia moral. Por descontado, esto es una generalización o, si se prefiere, una idealización. Pero pensemos en bienes colectivos económicos como un arancel que beneficia a un grupo de empresas o una rebaja fiscal a favor de ciertos inversores. Es poco probable que las empresas o los inversores cooperen para conseguir este tipo de bienes colectivos movidos por consideraciones morales (sin perjuicio de que quizá normas sociales de reciprocidad o justicia puedan desempeñar un papel importante en las transacciones económicas). Si las empresas o los inversores consiguen movilizarse y actuar colectivamente, será gracias a incentivos selectivos y no a consideraciones morales. Si pasamos a bienes colectivos políticos, como la victoria electoral de un partido con respecto a sus seguidores, una protesta contra la guerra con respecto al grupo de los pacifistas, o la preservación del medio ambiente, es razonable suponer que la cooperación se conseguirá no principalmente a través de incentivos selectivos, sino mediante la apelación a valores morales que movilicen a la sociedad.

Esta distinción entre bienes colectivos económicos y políticos está relacionada con la literatura sobre el efecto *crowding out*, por el que motivaciones externas como incentivos selectivos materiales pueden desplazar motivaciones intrínsecas (entre las que se cuentan los principios morales) (Frey, 1994; Kreps, 1997; Miller y Whitford, 2002; Sansone y Harackiewicz, 2000). En los términos de la función de utilidad, el efecto *crowding out* se da cuando

$$\frac{d\alpha_i}{d\beta} > 0$$

(recuérdese que está medido al revés, de modo que cuanto más se acerca  $\beta$  a 1, menor es el incentivo selectivo). El mecanismo del efecto *crowding out* se puede describir sintéticamente: una vez que la persona lleva a cabo su acción motivada por un incentivo externo, se rompe la conexión original entre la acción y los principios morales que la persona tiene. Ésta ya no percibe su acción como resultado de su constitución moral, sino como consecuencia de incentivos externos.

Si hay un efecto *crowding out* sistemático, deberíamos observar una especie de polarización en las formas de acción colectiva según si el tipo de bien colectivo está asociado o no a una motivación intrínseca. Así, en bienes colectivos económicos dominarán los incentivos selectivos y en los políticos los principios morales.

Es interesante señalar que hay bienes colectivos que comparten características de las dos esferas, la económica y la política. Por ejemplo, los bienes colectivos laborales pueden plantearse en términos puramente “economicistas” (mejoras salariales, defensa de derechos adquiridos, etc.) o en términos políticos (enfrentamiento con el capital, desarrollo de la conciencia de clase, lucha contra la explotación, etc.). A veces los estudiosos se refieren a este asunto como el “problema de Lenin” (Levi, 2003): cómo conseguir que los trabajadores entiendan sus reivindicaciones no en términos económicos, sino políticos. En este caso, lo que se sigue del modelo presentado es que los sindicatos profesionales o gremiales se basarán fundamentalmente en incentivos selectivos, y sus demandas serán en lo esencial económicas, mientras que los sindicatos de clase o ideológicos darán más peso a las motivaciones morales y sus demandas se teñirán de una justificación política.

### Sacrificio y heroísmo

En su análisis de la acción colectiva, Hirschman (1982) habla del peregrinaje como un tipo de acción que no encuentra encaje en la teoría económica, pues el beneficio de la acción depende en parte del coste que uno esté dispuesto a asumir: cuanto mayor sea el coste, mayor el beneficio que aguarda al final del viaje. Si alguien realiza el Camino de Santiago en un lujoso coche con chofer, pernoctando en los mejores hoteles, no es realmente un peregrino. El peregrino ha de padecer sufrimiento antes de llegar a su destino. El extraño caso de Alvin Straight, llevado al cine por David Lynch (*The Straight Story*), muestra este mismo mecanismo en el que los costes aparecen unidos a los bene-

ficios (ahora en versión secular): Straight, de 73 años, con artritis y una vista deficiente, decidió cruzar Estados Unidos montado en una cortadora de césped para reunirse con un hermano del que llevaba largo tiempo distanciado. La penuria del viaje dio un valor sobresaliente a su acción.

En el ámbito de la acción colectiva hay casos similares en los que es difícil distinguir qué es coste y qué es beneficio. Julia Hill era una joven mujer cuando decidió subirse a una secuoya gigante para intentar evitar la tala de un bosque. La protesta se alargó y Hill permaneció 738 días subida al árbol, con grave riesgo para su salud y su vida en diversas ocasiones (Hill, 2000). El beneficio moral o existencial que Hill extrajo de aquella experiencia dependía crucialmente de los sufrimientos que hubo de sobrellevar durante su protesta.

Bobby Sands, el miembro del IRA que se hizo internacionalmente conocido por morir en una huelga de hambre en 1981 tras 66 días de ayuno en protesta por el trato que recibían los prisioneros del IRA en las cárceles británicas, dejó por escrito sus impresiones sobre las actividades que los presos llevaron a cabo en los años previos a la huelga de hambre: primero, la renuncia a vestirse con el uniforme carcelario, cubriéndose sólo con toallas o mantas, y luego, la renuncia a lavarse o a hacer sus necesidades fuera de la celda, hasta llegar a vivir por decisión propia en condiciones realmente infrahumanas. Sands lo cuenta en estos términos:

Me había vuelto esquelético. Pero nada realmente importaba, excepto permanecer firme. Me revolví continuamente en el suelo, sintiendo las punzadas del frío. No tienen nada en todo su arsenal imperial con lo que quebrar el espíritu de un prisionero de guerra Republicano que se niega a ceder, pensé. (Sands, 1997: 81)

Visto así, como un duelo entre los principios de Bobby Sands y el sufrimiento producido por el respeto a esos principios, es evidente que, en cierto sentido, cuanto más duras son las pruebas a las que se somete a la persona, mayor es la utilidad moral de superarlas. Es como si los costes materiales hicieran posible un aumento de la utilidad moral.

Sands llevó hasta el extremo este mecanismo, pues aceptó morir de hambre por la causa que defendía. Se trata de un caso de martirio o de heroísmo, similar por ejemplo al del monje vietnamita Thich Quang Duc, que se auto-inmoló en una pira de fuego en 1963. Puede parecer que se trata de conductas extremas e inusuales, que por estar tan alejadas de las formas habituales de cooperación no tienen otro valor que el anecdótico. Sin embargo, en algunos casos tienen un efecto contagioso evidente: tras la experiencia de Duc, varios cientos de personas decidieron seguir su ejemplo en los lugares más diversos del planeta (Biggs, 2005). Por otro lado, aunque se trate de acciones infrecuentes, pueden tener gran relevancia en la dinámica de la acción colectiva. La muerte de Bobby Sands y otros nueve compañeros suyos a lo largo de 1981 tuvo un efecto muy profundo en la dinámica del movimiento Republicano, abriendo paso a una mayor implicación en la política institucional que con el tiempo trabajaría a favor del abandono de la violencia terrorista.

La proliferación de atentados terroristas suicidas en el mundo a partir de 1982 es la mejor muestra de la importancia que pueden llegar a tener estos comportamientos tan extremos (Gambetta, 2005; Pape, 2005; Reuter, 2000). El fenómeno de los ataques suicidas apenas ha sido abordado en la literatura sobre acción colectiva, quizá por lo mucho que se aleja de las condiciones controladas de laboratorio. Con todo, se trata de un fenómeno que pone a prueba el funcionamiento de las teorías existentes, pues no puede explicarse fácilmente ni por los modelos tradicionales basados en el auto-interés ni por los modelos que incorporan normas sociales de reciprocidad y justicia (*fairness*).

Mediante una sencilla modificación de la función de utilidad propuesta es posible representar este sorprendente mecanismo motivacional en virtud del cual un aumento de los costes materiales puede producir un aumento aún mayor de la utilidad moral. En la función original (1), la utilidad moral depende, entre otras cosas, de cuánto haya en juego, medido a través del valor  $V$  del bien colectivo. Podemos sin embargo considerar que lo que hay en juego en un dilema de acción colectiva no sólo es el valor del bien colectivo, sino también el coste o sacrificio personal que supone la cooperación.

No obstante, no sería realista suponer que cuanto más alto sea el coste material de la cooperación, mayor la utilidad moral sin más. Si fuera así, las personas con fuertes convicciones morales tendrían una disposición ilimitada para asumir costes. Ni la forma más brutal de tortura, o la muerte de los familiares más directos, serían suficientes para disuadir a alguien con principios muy sólidos. Cabe pensar que el sufrimiento tiene rendimientos morales decrecientes, en el sentido de que puede llegar un momento en que un aumento de los costes materiales no sea ya compensado por un aumento equivalente de utilidad moral. Por eso, en la función de utilidad modificada (en la que sigue suponiéndose que  $J = 0$ ), lo que está moralmente en juego se modeliza como  $V \ln(K)$ :

$$(10) \quad U_i(C_i) = \left( 1 - \left( \frac{N - \sum C_i}{N} \right)^F \right) V - \beta K C_i + \alpha_i (C_i - \bar{C}) V \ln(K)$$

Esto significa que lo que la cooperación pone moralmente en juego es una combinación de la importancia del bien colectivo y del coste personal o sacrificio que la cooperación implica, teniendo en cuenta que la capacidad para soportar costes es limitada. Lo que nos interesa determinar ahora es cómo la utilidad varía con los costes:

$$(11) \quad \frac{dU_i}{dK} = -\beta C_i + \frac{\alpha_i (C_i - \bar{C})}{K}$$

Esta derivada sólo puede ser positiva cuando  $C_i = 1$ . En concreto, la utilidad crece cuando aumenta  $K$  siempre que se cumpla:

$$(12) \quad K < \frac{\alpha_i(1-\bar{C})V}{\beta}$$

La expresión (12) contiene un par de implicaciones interesantes. Por un lado, es evidente que cuanto menor es la media social de cooperación, mayor puede ser el sacrificio de la persona. Puesto que la utilidad moral aumenta cuanto mayor sea la diferencia entre la cooperación de la persona y la cooperación media del grupo, se sigue que cuanto menos cooperación haya en el grupo, mayor margen habrá para asumir costes elevados. Nótese que, en el límite, si todo el mundo coopera, es imposible que un aumento de los costes suponga un aumento de utilidad. Por otro lado, es también claro que cuanto mayor sea el valor del bien colectivo en juego, mayor será la capacidad de la persona para realizar sacrificios materiales.

La aplicación de estos resultados al caso de los ataques suicidas es casi inmediata. Si la utilidad de la vida de la persona está comprendida en el intervalo

$$\left( K, \frac{\alpha_i(1-\bar{C})V}{\beta} \right)$$

es racional para la persona suicidarse si la cooperación exige la entrega de la vida. Cooperar asumiendo un coste tan alto como el de la renunciar a la vida proporciona un incremento de utilidad. Esto requiere que  $V$  y  $\alpha_i$  sean valores muy altos y que la utilidad de la vida no sea demasiado elevada. De hecho, cuanto menos valore la persona la vida, más probable es que pueda haber un suicidio racional. Es decir, cuanto más desesperadas sean las condiciones de la existencia, más probable es que la utilidad de cooperar por un bien colectivo sea superior al coste de la vida.

Ricolfi (2005), en su penetrante análisis sobre los ataques suicidas palestinos, concluye que las condiciones sociales que favorecen estos ataques consisten en la ausencia de expectativas razonables de mejora y desarrollo personal. Dichas condiciones, combinadas con la radicalización religiosa y política (un  $\alpha$  muy elevado), permiten comenzar a entender las decisiones de los terroristas suicidas. Los ataques suicidas no se observan, por el contrario, en lugares en los que las personas encuentran oportunidades para llevar a cabo una vida razonable (Kalyvas y Sánchez-Cuenca, 2005).

En general, la inclusión de los costes de la cooperación como un elemento positivo en el componente moral de la función de utilidad permite entender comportamientos extremos como los atentados suicidas. El mecanismo es el siguiente: las personas pueden entender acciones muy costosas como una forma de poner a prueba sus convicciones morales, de forma que cuanto más exigente sea prueba, mayor sea la utilidad moral producida. Si esa utilidad es superior a la de vivir, dar la vida por la causa es racional.

## CONCLUSIONES

Blais (2000), al concluir su libro sobre cómo la teoría de la elección racional explica la acción de votar, hace esta confesión:

Como muchos de mis conciudadanos, siento que debo actuar en coherencia con los principios en los que creo. Dado que me veo a mí mismo como un demócrata, sería incongruente no votar. Voto, por tanto, porque quiero ser coherente con mis principios. Sí, me sentiría algo culpable si no votara. (p.143)

Este trabajo ha intentado recoger y desarrollar esta idea tan simple: ser *free-rider* no sale gratis cuando la persona tiene principios morales en virtud de los cuales siente la obligación de cooperar en la consecución de un bien colectivo. Por descontado, esa obligación moral no es absoluta ni incondicional. Depende de varios factores: de la intensidad de los principios morales, del valor del bien colectivo y del grado de cooperación que haya en el grupo.

Los principios reflejan el carácter moral del agente y han de conjugarse con sus preferencias económicas o materiales. He propuesto una nueva función de utilidad que combina lo que he llamado utilidad económica y moral. Esta función pretende recoger los principales factores que se han utilizado en la literatura basada en supuestos de auto-interés material para dar cuenta de la cooperación: los incentivos selectivos, la función de producción del bien colectivo y la seguridad que proporciona el número de cooperadores. A esos factores he añadido un término que describe la forma en la que se genera la utilidad moral.

En el modelo presentado, el juego de referencia para caracterizar un problema de acción colectiva cuando no hay principios morales es el Dilema del Prisionero, el juego más desfavorable para el surgimiento de la cooperación. Mediante principios morales y la función de producción del bien colectivo, el Dilema del Prisionero puede transformarse en otros juegos que hacen más probable el comportamiento cooperativo.

A partir de este modelo, he desarrollado algunas implicaciones y extensiones de la función de utilidad propuesta. Quizá la más importante sea la que tiene que ver con la posibilidad de explicar conductas extremas, pues son esas conductas justamente las que se escapan a los modelos de normas sociales que se han desarrollado en la economía del comportamiento. El modelo hace inteligible la posibilidad de un sacrificio de la persona en beneficio del grupo que persigue el bien colectivo, incluyendo como caso límite que se sacrifique la propia vida. Para llegar a este resultado es necesario hacer depender la utilidad moral de la capacidad de soportar costes materiales.

## REFERENCIAS BIBLIOGRÁFICAS

- BICCHIERI, C. (2006), *The Grammar of Society. The Nature and Dynamics of Social Norms*, Cambridge, Cambridge University Press.
- BIGGS, M. (2005), "Dying Without Killing: Self-Immolations, 1963-2002", en Diego Gambetta (ed.) *Making Sense of Suicide Missions*, Oxford, Oxford University Press, pp. 173-208.
- BLAIS, A. (2000), *To Vote or Not to Vote. The Merits and Limits of Rational Choice Theory*, Pittsburgh, The University of Pittsburgh Press.
- BOWLES, S. (2004), *Microeconomics. Behavior, Institutions, and Evolution*, Princeton, Princeton University Press.
- CAMERER, C.F. (2003), *Behavioral Game Theory. Experiments in Strategic Interaction*, Princeton, Princeton University Press.
- CAMERER, C.F., G. LOEWENSTEIN y M. RABIN (eds.) (2004), *Advances in Behavioral Economics*, Princeton, Princeton University Press.
- CHONG, D. (1991), *Collective Action and the Civil Rights Movement*, Chicago, The University of Chicago Press.
- CHWE, M.S. (1999), "Structure and Strategy in Collective Action", *American Journal of Sociology*, vol. 105, pp. 128-156.
- COLLINS, E. (1997), *Killing Rage*, Londres, Granta.
- COX, G.W. (1999), "The Empirical Content of Rational Choice Theory", *Journal of Theoretical Politics*, vol. 11, pp. 147-169.
- ETZIONI, A. (1988), *The Moral Dimension. Towards a New Economics*, Nueva York, The Free Press.
- FREY, B.S. (1994), "How Intrinsic Motivation Is Crowded Out and In", *Rationality and Society*, vol. 6, pp. 334-352.
- FRIEDMAN, J. (ed.) (1996), *The Rational Choice Controversy*, New Haven, Yale University Press.
- GAMBETTA, D. (ed.) (2005), *Making Sense of Suicide Missions*, Oxford, Oxford University Press.
- GRANOVETTER, M. (1973), "The Strength of Weak Ties", *American Journal of Sociology*, vol. 78, pp. 1360-1380.
- (1978), "Threshold Models of Collective Behavior", *American Journal of Sociology*, vol. 83, pp. 1420-1443.
- HARDIN, R. (1982), *Collective Action*, Londres, John Hopkins University Press.

- HECKATHORN, D. D. (1996), "The Dynamics and Dilemmas of Collective Action", *American Economic Review*, vol. 6, pp. 250-277.
- HILL, J.B. (2000), *The Legacy of Luna*, Nueva York, Harper SanFrancisco.
- HIRSCHMAN, A. O. (1982), *Shifting Involvements. Private Interest and Public Action*, Princeton, Princeton University Press.
- KALYVAS, S. e I. SÁNCHEZ-CUENCA (2005), "Killing Without Dying: The Absence of Suicide Missions", en D. Gambetta (ed.), *Making Sense of Suicide Missions*, Oxford, Oxford University Press, pp. 209-232.
- KREPS, D.M. (1997), "Intrinsic Motivation and Extrinsic Incentives", *The American Economic Review*, vol. 87, pp. 359-364.
- LEVI, M. (2003), "Inducing Preferences within Organizations: The Case of Unions", Manuscrito, Universidad de Washington.
- LEWIN, L. (1991), *Self-Interest and Public Interest in Western Politics*, Oxford, Oxford University Press.
- LICHBACH, M. I. (1996), *The Cooperator's Dilemma*, Ann Arbor, The University of Michigan Press.
- LOHMANN, S. (1994), "The Dynamics of Informational Cascades", *World Politics*, vol. 47, pp. 42-101.
- MACDONALD, P. K. (2003), "Useful Fiction or Miracle Maker: The Competing Epistemological Foundations of Rational Choice Theory", *American Political Science Review*, vol. 97, pp. 551-565.
- MACKIE, G. (2003), *Democracy Defended*, Cambridge, Cambridge University Press.
- MACY, M. (1991), "Chains of Cooperation: Threshold Effects in Collective Action", *American Sociological Review*, vol. 56, pp. 730-747.
- MARWELL, G. y P. OLIVER (1993), *The Critical Mass in Collective Action: A Micro-Social Theory*, Cambridge, Cambridge University Press.
- MILLER, G. y A. B. WHITFORD (2002), "Trust and Incentives in Principal-Agent Negotiations", *Journal of Theoretical Politics*, vol. 14, pp. 231-267.
- OLSON, M. (1965), *The Logic of Collective Action*, Cambridge (Mass.), Harvard University Press.
- PAPE, R. A. (2005), *Dying to Win. The Strategic Logic of Suicide Terrorism*, Nueva York, Random House.
- RAWLS, J. (1971), *A Theory of Justice*, Cambridge (Mass.), Harvard University Press.
- REUTER, Ch. (2004), *My Life Is a Weapon. A Modern History of Suicide Bombing*, Princeton, Princeton University Press.
- RICOLFI, L. (2005), "Palestinians, 1981-2003", en D. Gambetta (ed.) *Making Sense of Suicide Missions*, Oxford, Oxford University Press, pp. 77-129.

- SANDLER, T. (1992), *Collective Action. Theory and Applications*, Ann Arbor, The University of Michigan Press.
- SANDS, B. (1997), *Writings from Prison*, Boulder, Roberts Rinehart Publishers.
- SANSONE, C. y J. M. HARACKIEWICZ (eds.) (2000), *Intrinsic and Extrinsic Motivation. The Search for Optimal Motivation and Performance*, San Diego, Academic Press.
- SCHEFFLER, S. (1982), *The Rejection of Consequentialism*, Oxford, Oxford University Press.
- SHELLING, Th. C. (1978), *Micromotives and Macrobehavior*, Nueva York, Norton.
- SEN, A. (1977), "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy & Public Affairs*, vol. 6, pp. 317-344.
- (2002), *Rationality and Freedom*, Cambridge (Mass.), Harvard University Press.
- SEN, A. y B. WILLIAMS (eds.) (1982), *Utilitarianism and Beyond*, Cambridge, Cambridge University Press.
- SOBEL, J. (2005), "Interdependent Preferences and Reciprocity", *Journal of Economic Literature*, vol. 43, pp. 392-436.
- SUGDEN, R. (1984), "Reciprocity: The Supply of Public Goods Through Voluntary Contributions", *The Economic Journal*, vol. 94, pp. 772-87.
- WHEWELL, W. (1984), *Selected Writings on the History of Science* (edición de Y.Elkana), Chicago, University of Chicago Press.
- WILLIAMS, B. (1981), *Moral Luck*, Cambridge, Cambridge University Press.
- WILSON, J.P. (1995), *Political Organizations*, Princeton, Princeton University Press.

**RECIBIDO: 3/12/05**  
**ACEPTADO: 26/04/06**