

El pluralismo motivacional en la especie humana. Aportaciones recientes de la ciencia social experimental¹

Jordi Tena Sánchez

Universitat Autònoma de Barcelona. Departament de Sociologia
jordi.tena@uab.cat

Recibido: 07-10-2008

Aceptado: 26-03-2009

Resumen

La teoría de la elección racional ha realizado aportaciones decisivas para el estudio de los fenómenos sociales. No obstante, el uso de supuestos falsos o excesivamente incompletos supone un importante lastre en este sentido. El presente trabajo parte del rechazo a la adopción rutinaria de un punto de vista monista motivacional, según el cual el egoísmo es el único tipo de supuesto motivacional capaz de dar cuenta de la acción social.

En estas páginas se trata, primero, de esbozar una panorámica general de los tipos de motivación humana más importantes identificados por la bibliografía especializada, para, posteriormente, centrarse específicamente en el cuantitativamente más importante de todos ellos, la reciprocidad fuerte. El trabajo concluye con una discusión sobre la posibilidad de incluir una pluralidad de motivaciones en los modelos formales.

Palabras clave: teoría de la elección racional, motivación, egoísmo, reciprocidad fuerte.

Abstract. *Motivational Pluralism in Human Species. Recent contributions from experimental social science*

Rational Choice Theory has made crucial contributions for the study of social phenomena. However, the utilization of false or too simple assumptions supposes an important limitation in this sense. This paper begins with a refusal of the common adoption of a motivationally monist point of view in which self-interest is the only motivation capable of explaining social action.

In this paper, first, I offer a general overview of the most important kinds of human motivation identified by the specialized literature. Then, I focus on the quantitatively most important one: strong reciprocity. The paper finishes with a discussion about the possibility of incorporating plural motivations in formal models.

Key words: Rational Choice Theory, motivation, self-interest, strong reciprocity.

1. El presente trabajo se ha desarrollado en el marco de un proyecto del Plan Nacional de I+D+i financiado por el Ministerio de Ciencia e Innovación (MICINN), con referencia CSO2009-09890. El autor expresa su agradecimiento a José A. Noguera y a Francisco J. León por sus comentarios.

Sumario

| | |
|--|--|
| La teoría de la elección racional y el supuesto de egoísmo | La reciprocidad débil y la reciprocidad fuerte |
| El pluramismo motivacional en la especie humana | Conclusiones |
| | Referencias bibliográficas |

La teoría de la elección racional y el supuesto de egoísmo

No parece exagerado sostener que la teoría de la elección racional (TER en adelante) constituye la escuela teórica más fértil en el campo de las ciencias sociales. Su relevancia no se deriva, empero, o no principalmente, de su popularidad entre los científicos sociales (no cabe duda de que la TER no es la corriente mayoritaria en ciencias sociales), sino más bien de sus decisivas contribuciones al estudio de los fenómenos sociales.

En su versión estándar (TER-e en adelante), la teoría de la elección racional supone que todas las acciones sociales se explican en base a las decisiones adoptadas por un individuo o un conjunto de individuos en un contexto determinado. Se asume que dichos individuos tratan de maximizar su utilidad. A saber, ante un cierto conjunto de oportunidad, se supone que el individuo escogerá la opción que le reportará mayor beneficio (Aguiar y De Francisco, 2008).

En el modelo estándar, de modo particular, se asume que los individuos son egoístas. La TER-e adopta, de este modo, un punto de vista *monista motivacional*: se asume que la acción puede ser explicada a través de un único tipo de motivación (el egoísmo). Que el egoísmo es el único tipo de motivación contemplado por la teoría de la elección racional ha sido discutido por muchos autores. En numerosas ocasiones se sostiene que esto no tiene por qué ser así, que la TER no necesita suponer que los agentes sean egoístas, sino únicamente «autointeresados» (Sánchez-Cuenca, 2007). Es decir, sus preferencias no tienen por qué referirse exclusivamente a su propio interés estrecho. Desde mi punto de vista, este argumento es totalmente correcto. No cabe duda de que el de egoísmo no constituye un requisito lógico de la TER y que hay muchos autores que utilizan modelos formales que podemos, claramente, catalogar como modelos de teoría de la elección racional, y que no emplean ese supuesto. No obstante, no cabe duda tampoco de que el de egoísmo es el supuesto habitual entre la mayoría de los autores de la TER y que es uno de los elementos que mejor define a la versión estándar de la misma².

La TER-e ha sido criticada en innumerables ocasiones por la falta de realismo y/o por lo incompleto de sus supuestos³. Entre los más criticados se

2. Casos paradigmáticos en este sentido serían Becker (1986), Coleman (1988), Friedman (1953), Hardin (1982), Olson (1965), Taylor (1982, 1987, 1991) y Ullmann-Margalit (1977).
3. Véase, por ejemplo, Boudon (2003, 2008), Camerer, Loewenstein y Rabin (2004), Elster (2007), Fehr y Falk (2002a, 2002b), Gintis (2005), Hedström (2005), Sánchez-Cuenca (2008).

encuentran, por ejemplo, además del propio egoísmo, el hecho de concebir a los agentes como dotados de una capacidad perfecta de procesamiento de información y como maximizadores de utilidad esperada, o el hecho de suponer que el descuento temporal de los individuos aumenta de forma exponencial (Loewe, 2006).

En el presente trabajo, me centraré exclusivamente en discutir el uso del supuesto de egoísmo, aunque mucho de lo que se dirá en estas páginas resulta, a mi juicio, generalizable a todos los casos de supuestos falsos o excesivamente simples. Pese a que la TER ofrece herramientas teóricas con un enorme potencial explicativo y predictivo de los fenómenos sociales, el uso reiterado de supuestos falsos o excesivamente incompletos limita sobremanera el avance de la teoría, así como su capacidad para dar cuenta de los fenómenos estudiados.

El uso del supuesto de egoísmo coincide, como veremos, con el segundo tipo de problema, el de la utilización de supuestos excesivamente simples. Suponer que los seres humanos tenemos una capacidad perfecta de procesamiento de información constituye un ejemplo de supuesto falso. En cambio, no es falso que el egoísmo sea una motivación humana, al contrario, es una de las más importantes. Aquí el problema consiste en que el supuesto es demasiado incompleto, los seres humanos tenemos otras motivaciones además del egoísmo y algunas de ellas son más importantes que aquél. «Más importantes» tiene aquí un sentido cuantitativo. Hay motivaciones que explican una parte mayor de la conducta que el egoísmo y en casos muy relevantes.

Los defensores de la utilización del egoísmo como único supuesto motivacional acostumbran a manejar dos tipos distintos de argumento. En ocasiones se afirma que, si bien los seres humanos pueden tener motivaciones plurales, el egoísmo es la más importante de todas ellas, de manera que las demás tienen poco o nulo poder explicativo. En esta línea suele sostenerse también que dichas motivaciones se anulan entre ellas y que, por tanto, en aras de respetar el requisito de *parsimonia*⁴, lo mejor es dejarlas fuera de los modelos explicativos (Marí-Klose, 2000). Pero este argumento es inadecuado. Existen cantidades ingentes de evidencia empírica que muestran que el egoísmo, pese a ser una motivación importante en la especie humana, no es la principal. En este sentido, existen motivaciones diferentes del egoísmo que tienen un importante poder explicativo en muchas y muy relevantes situaciones, y que en modo alguno se anulan entre ellas⁵. En la próxima sección, trataré de ofrecer una panorámica general de las más destacables.

4. El *desiderata* de parsimonia nos dice que, a igualdad de condiciones, la explicación más sencilla siempre es preferible. Sobre la importancia de que las explicaciones, los conceptos, las hipótesis, etc. que se utilizan en ciencia cumplan con este requisito, véase Noguera (2006).
5. Véase, por ejemplo, Bowles (2004, 2005), Bowles y Choi (2003), Bowles y Gintis (2000a, 2000b, 2001a, 2001b, 2002, 2003), Bowles y Naidu (2005), Camerer, Loewenstein y Rabin (2004), Dawes y Thales (1988), Fehr y Falk (2002a, 2002b), Gintis (2000a, 2000b, 2001a, 2003, 2005, 2006, 2007), Gintis, Bowles, Boyd y Fehr (2008), Henrich, Boyd, Bowles, Camerer, Fehr, Gintis (2005), Mansbridge (1990), Sally (1995), Yung-An Hu y Day-Yang Liu (2003).

El segundo tipo de respuesta que ofrecen los defensores del monismo motivacional pasa por argumentar que es un error tratar de que un modelo formal capture características de la realidad. Los modelos formales deben usarse, desde este punto de vista, de forma instrumental, por su potencial predictivo. En la que quizá sea la formulación más célebre de este punto de vista, M. Friedman (1953) argumenta que no tiene sentido demandar que los supuestos en que se basa una teoría se ajusten a la realidad, debido a que, dado el requisito de parsimonia, una teoría o una hipótesis científica deben tratar de explicar el máximo posible de realidad con el mínimo posible de supuestos. De este modo, cuanto mejor sea una teoría, muy probablemente, menos realistas serán sus supuestos. En definitiva, según Friedman, no son esos supuestos lo que debemos testar, sino las predicciones que se derivan de la teoría.

Pero, nuevamente, el argumento, desde mi punto de vista, resulta inadecuado. De modo general, el uso instrumental de la teoría podría ser legítimo si se cumpliera alguno de los siguientes requisitos: la existencia de un mecanismo que nos permitiera dar cuenta de por qué nuestro modelo funciona, o bien que el grado de acierto en la predicción fuese tan elevado que no nos quedara otra opción que reconocer la bondad del modelo, aunque no tuviésemos ni la más mínima idea de las razones de la misma (Elster, 2007: 464). Pero en los modelos que utilizamos en ciencias sociales, no se cumple ninguno de los dos requisitos. Por una parte, no contamos con un mecanismo que nos permita explicar por qué el modelo funciona a pesar de estar basado en supuestos erróneos. En biología, por ejemplo, suelen utilizarse modelos de teoría de juegos que permiten predecir que aquellos organismos que desarrollen una determinada estrategia tendrán éxito evolutivo. Los modelos se construyen «como si» los organismos fuesen agentes racionales. En este caso, es obvio que los agentes no son racionales, pero disponemos de un mecanismo, la selección natural, que nos permite comprender por qué el modelo funciona y por qué los organismos que tienen éxito son los que actúan «como si» siguiesen una determinada estrategia. Pero en ciencias sociales no disponemos de mecanismos de ese tipo. En segundo lugar, en muchos casos relevantes, el modelo simplemente no funciona y el supuesto potencial predictivo de los modelos TER-e instrumentales no se da⁶.

En los dos apartados siguientes, en primer lugar, trataré de dibujar un esbozo general de los tipos de motivación humana más importantes identificados por la bibliografía especializada. Posteriormente, me centraré específicamente en el cuantitativamente más importante de esos tipos de motivación, la *strong*

6. Véase, por ejemplo, Bowles (2004, 2005), Bowles y Choi (2003), Bowles y Gintis (2000a, 2000b, 2001a, 2001b, 2002, 2003), Bowles y Naidu (2005), Camerer, Loewenstein y Rabin (2004), Dawes y Thales (1988), Fehr y Falk (2002a, 2002b), Gintis (2000a, 2000b, 2001a, 2003, 2005, 2006, 2007), Gintis, Bowles, Boyd y Fehr (2008), Henrich, Boyd, Bowles, Camerer, Fehr y Gintis (2005), Mansbridge (1990) y Sally (1995). Sobre los problemas del uso instrumental de la teoría, además del citado Elster (2007), pueden consultarse también Boudon (2003), Hedström (2005), Ovejero (2008) o Sen (1980).

reciprocity ('reciprocidad fuerte'). Por *reciprocidad fuerte* entenderé «una propensión a cooperar y compartir con aquellos que tienen una predisposición similar y una voluntad de castigar a aquellos que violan la cooperación y otras normas sociales, aun cuando el hecho de compartir y el castigo conlleven costos personales» (Bowles y Gintis, 2001b: 173).

Una aclaración resulta importante antes de continuar. En lo que sigue, dejaré de lado un importante tipo de motivación humana, la motivación para seguir normas. Pese a que las normas están relacionadas con un tipo de motivación como la reciprocidad fuerte, la complejidad del tema desaconseja abordar aquí esa cuestión.

El pluralismo motivacional en la especie humana

El ser humano es un animal capaz de albergar una amplia variedad de motivaciones diferentes. En la presente sección, repasaré brevemente algunas de las más destacadas por la literatura, para, en la siguiente, centrarme ya exclusivamente en la reciprocidad fuerte. Antes de empezar, empero, vale la pena apuntar que sería bastante pobre suponer que existe una clase de individuos que siempre está motivada de una determinada forma (que siempre es egoísta, por ejemplo), mientras que existe otra clase de individuos que siempre lo está de una forma diferente (que siempre es altruista, por ejemplo) y así sucesivamente. En primer lugar, los diferentes motivos pueden variar de intensidad. En segundo lugar, de modo más importante, todos nosotros albergamos diferentes tipos de motivaciones.

Podemos distinguir entre preferencias sociales y no sociales (Fehr y Fischbacher, 2005). El caso paradigmático de preferencias no sociales es el de las preferencias egoístas. Cuando un agente es egoísta, sus preferencias se refieren únicamente a su propio bienestar y se muestra indiferente ante el bienestar de los demás. El egoísmo goza de buena salud entre los seres humanos y es un tipo de motivación muy importante en términos cuantitativos.

Las motivaciones sociales son aquellas que se dirigen hacia otras personas, no son, por tanto, indiferentes respecto de las preferencias o los intereses de los demás⁷. Un primer tipo de preferencia social es la *aversión a la inequidad* (*inequity aversion*) (Fehr y Fischbacher, 2005). Las personas motivadas de esta forma quieren alcanzar una distribución equitativa de los recursos. Agentes motivados de este modo serán altruistas con los más desfavorecidos, pero también desearán un reparto más igualitario de los bienes de los más afortunados. En muchas ocasiones, agentes motivados de este modo se comportarán de la misma forma que los motivados por la reciprocidad fuerte. Por ejemplo, ambos tipos de agentes desearán reducir los pagos obtenidos por otro agente si ese

7. Podría argumentarse que la conducta egoísta también se dirige en ocasiones hacia los demás. La diferencia radica en que, cuando la conducta egoísta se dirige hacia los demás, lo hace sólo instrumentalmente en aras de alcanzar el beneficio propio, mientras que en las conductas sociales los demás son importantes por sí mismos (Jenks, 1990).

agente ha actuado de un modo tal que ha obtenido unos pagos mucho más altos que los logrados por el agente averso a la inequidad o motivado por la reciprocidad fuerte. Sin embargo, la reciprocidad fuerte es un motivo mucho más importante en términos cuantitativos (Fehr y Fischbacher, 2005).

Algunos autores (Charness y Rabin, 2002) han encontrado evidencia a favor de la existencia de una motivación para ayudar a los más débiles. No obstante, dicha motivación es muy difícil de distinguir en la práctica de la aversión a la inequidad (Fehr y Fischbacher, 2005).

Otro tipo de preferencia social que podemos encontrar en la literatura especializada es el *altruismo incondicional* (Fehr y Fischbacher, 2005). El altruista incondicional trata de satisfacer las preferencias o los intereses de los demás. Se trata de una conducta incondicional y no recíproca, en el sentido de que no se espera que los destinatarios de nuestro altruismo nos correspondan en un futuro más o menos cercano. El altruismo incondicional es diferente de la aversión a la inequidad, debido a que el agente que es altruista lo será también con aquellos agentes socialmente más favorecidos, mientras que el averso a la inequidad deseará un reparto más igualitario de los bienes de éstos. Por otra parte, el altruismo incondicional es diferente de la reciprocidad fuerte debido a que el agente motivado por la reciprocidad fuerte coopera sólo de forma condicional y, además, sanciona a aquellos otros agentes que no cooperan.

Uno de los primeros y más relevantes críticos con el uso exclusivo del supuesto de egoísmo en los modelos formales fue Amartya Sen. En un célebre trabajo (1986), Sen distingue entre dos tipos de motivaciones más o menos distintas del egoísmo que pueden dar cuenta de la acción: la *simpatía* y el *compromiso*. En estas páginas, me centro exclusivamente en motivaciones identificadas en trabajos de tipo experimental, por esa razón estoy dejando de lado aquí aportaciones provenientes de trabajos de tipo teórico. No obstante, dada la relevancia de las aportaciones de Sen a este debate, quisiera decir algo aquí acerca del encaje de su concepto de simpatía en el marco de los tipos de motivaciones a los que se está haciendo referencia en el presente trabajo⁸.

Según Sen, podemos, por ejemplo, solidarizarnos con una determinada persona que está sufriendo, por simpatía, porque nos hace sufrir su sufrimiento, por así decirlo. Para sostener esto, no necesitamos apartarnos de la senda de la conducta racional-egoísta. No estaríamos más que en una variante particular de la misma, aquella en que asumimos como propias las preferencias de otros individuos y tratamos de maximizar su utilidad. En este sentido, la maximización de la utilidad de los otros no es más que un medio para maximizar la nuestra. Sin embargo, desde mi punto de vista, Sen incurre aquí en una falacia argumentativa. Según Sen, detrás de la simpatía existen emociones muy intensas. De este modo, en un caso de simpatía, tratamos de paliar el sufri-

8. Según Sen, nos encontramos ante un caso de *compromiso* cuando el agente actúa motivado por consideraciones acerca del deber. Este tipo de motivación nos lleva inmediatamente a la cuestión de la motivación para seguir normas que, como se dijo, no será abordada en este trabajo.

miento de los demás porque el conocimiento de su existencia nos enferma. Si no actuásemos, nos sentiríamos culpables, mientras que hacerlo nos puede producir emociones positivas como el orgullo. Pero la evitación o el logro de esas emociones no pueden ser las razones de la acción, ya que se trata de estados mentales que sólo pueden ser alcanzados como subproductos de actividades que se llevan a cabo con otros objetivos (Elster, 2002). De este modo, pienso que lo que Sen trata de capturar con su concepto de simpatía se captura mejor con el concepto de altruismo tal y como se ha definido aquí, como un tipo de motivación distinta del egoísmo y en modo alguno reducible a aquél.

Merece la pena destacar que, independientemente de cuestiones cuantitativas, el egoísmo goza de prioridad lógica sobre el altruismo (Elster, 1990a). La razón es que, si bien nos resulta posible concebir algo así como un estado de naturaleza en la que todos los individuos se comportan única y exclusivamente de modo egoísta, no nos es posible hacer lo contrario. El altruista asume como propias las preferencias egoístas de un segundo. Si no hubiese individuos con preferencias egoístas, tampoco podría haber otros individuos, altruistas, que se esforzasen en procurar utilidad a los primeros. Por tanto, para que el altruismo pueda existir, es necesario que también exista el egoísmo, cosa que no sucede a la inversa.

La evidencia experimental también muestra que un importante tipo de preferencias sociales son de carácter malicioso (*spiteful*) y envidioso (Fehr y Fischbacher, 2005). Un agente motivado de esta forma siempre valora de manera negativa que los demás agentes obtengan pagos positivos y trata de perjudicarlos si es posible, aún a costa de asumir él mismo costes por hacerlo.

En una aproximación a la cuestión muy diferente, Jon Elster (1991a, 2000) sitúa a la envidia como uno de los pilares del orden social. Las motivaciones envidiosas (como las maliciosas) no son racionales en el sentido técnico del término, pero sí son sociales. Elster cita una fábula popular que ejemplifica bien esta cuestión. En una ocasión, hace ya muchos años, un genio mágico se apareció a un campesino que se encontraba trabajando la tierra. El genio le dijo al campesino que le pidiese cualquier deseo y que él se lo concedería. La única condición que le puso fue que concedería a su vecino el doble de lo que le concediese a él. Después de pensarlo por unos instantes, el campesino le pidió al genio que le arrancase un ojo.

En la literatura especializada, pueden encontrarse algunos tipos de motivación más, así como un tratamiento más detallado de los tipos referidos. No obstante, las mencionadas son las formas cuantitativamente más importantes de motivación humana. En el cuadro 1, se recogen conjuntamente de manera esquemática. Todas ellas (aisladamente o de forma conjunta) permiten explicar partes más o menos importantes de la conducta humana en situaciones relevantes. La reciprocidad fuerte juega un rol central de manera particular en la aplicación de sanciones a los *free-riders* que se abstienen de cumplir las reglas de cooperación. En cambio, por lo que se refiere a las conductas «positivas», como la aplicación de recompensas o la ayuda a los necesitados, pese a que la reciprocidad fuerte sigue siendo el motivo más importante, otros motivos como

Cuadro 1. Principales formas de motivación humana en términos cuantitativos.**Motivaciones no sociales:**

Egoísmo

- Cuando un individuo es egoísta, sus preferencias se refieren únicamente a su propio bienestar y se muestra indiferente ante el bienestar de los demás.

Motivaciones sociales:

Aversión a la inequidad

- Las personas motivadas de esta forma quieren alcanzar una distribución equitativa de los recursos. Agentes motivados de este modo serán altruistas con los más desfavorecidos, pero también desearán un reparto más igualitario de los bienes de los más afortunados.

Reciprocidad fuerte

- Propensión a cooperar y a compartir con aquellos que tienen una predisposición similar y una voluntad de castigar a aquellos que violan la cooperación y otras normas sociales, aun cuando el hecho de compartir y el castigo conlleven costos personales.

Altruismo incondicional

- El altruista incondicional trata de satisfacer las preferencias o los intereses de los demás. Se trata de una conducta incondicional y no recíproca, en el sentido de que no se espera que los destinatarios de nuestro altruismo nos correspondan en un futuro más o menos cercano.

Malicia y envidia

- Un agente motivado de esta forma siempre valora de manera negativa que los demás agentes obtengan pagos positivos y trata de perjudicarlos si es posible, aún a costa de asumir el mismo costes.

Fuente: elaboración propia.

el altruismo o la aversión a la inequidad tienen también un importante poder explicativo (Fehr y Fischbacher, 2005).

Así pues, podemos ver como, al elaborar modelos formales que permitan explicar causalmente fenómenos sociales, resulta adecuado adoptar un enfoque plural desde el punto de vista motivacional, esto es, asumir que para explicar la acción social necesitamos tener en cuenta diferentes tipos de motivación en los agentes. En el siguiente apartado, trataré de forma específica el tipo de motivación cuantitativamente más importante en el ser humano: la reciprocidad fuerte⁹.

9. La reciprocidad fuerte es el tipo de motivación cuantitativamente más importante, tanto entre los hombres como entre las mujeres. No obstante, diversos estudios experimentales han mostrado que las mujeres parecen tener una mayor tendencia hacia las conductas pro-sociales que los hombres (véase, por ejemplo, Aguiar, Brañas-Garza, Cobo-Reyes, Jiménez y Miller (2007), Andreoni y Vesterlund (2001), Brañas-Garza (2006) o Muren y Dufwenberg

La reciprocidad débil y la reciprocidad fuerte

La reciprocidad débil y sus limitaciones

Por reciprocidad débil entenderé la conducta recíproca basada en motivaciones egoístas. Las explicaciones sobre la cooperación que propone la psicología evolucionaria tienen como precedente propuestas respecto a comportamientos cooperativos como el altruismo y la reciprocidad generados desde la biología evolucionaria (Cosmides y Tooby, 1992). Hamilton (1963) postuló a la *inclusive fitness* ('aptitud inclusiva') posteriormente llamada por Maynard-Smith (1978) *kin selection* ('selección de parientes'), que argumenta que existen comportamientos altruistas entre individuos parientes de una especie, que pueden incluso sacrificar su vida para salvar a un pariente, si esto permite que el número de copias de los genes del individuo altruista se transmita a través de sus parientes, hecho que no ocurriría si el comportamiento altruista de sacrificio no se hubiera producido (García, 2005).

Por su parte, Trivers (1971) acuñó el concepto de *altruismo recíproco* para referirse a aquellos actos de altruismo que los individuos realizan con terceros a cambio de ser correspondidos por éstos en el futuro. Este comportamiento de altruismo recíproco fue seleccionado evolutivamente porque es adaptativo. Se trata de un rasgo diseñado por la selección natural y que guía las acciones de los individuos hacia el altruismo mutuo. Los postulados del altruismo recíproco fueron adoptados por Axelrod (1986a), quien, desde la teoría de juegos, particularmente a partir de la búsqueda de soluciones al dilema del prisionero iterado, encontró que la estrategia llamada *tit for tat*¹⁰ ('toma y daca') permite que la cooperación fundada en la reciprocidad sea evolutivamente estable¹¹.

Este tipo de estrategias casan bien con el supuesto de egoísmo. Nótese que el denominado «altruismo» recíproco, en realidad no es tal altruismo, sino que, según la caracterización ofrecida en la sección anterior, en realidad es egoísmo. Axelrod demostró que, si se dan determinadas circunstancias, en una situa-

(2006)). No obstante, algunas tesis clásicas de autoras feministas como Gilligan (1982), quien tratara de dar cuenta de las diferencias entre la conducta moral, social y política de hombres y mujeres a partir de la *gender socialization*, no parecen confirmarse a la luz de la evidencia empírica disponible (véase, por ejemplo, Scott, Matland, Michelbach y Bornstein (2001)).

10. Diseñada por Rapoport (véase Axelrod, 1986a) y que indica que se debe iniciar el juego cooperando y luego tomar la misma decisión —cooperar o no cooperar— que haya realizado el otro jugador en el movimiento previo.
11. Posteriormente (1997), Axelrod también mostró que la reciprocidad débil no necesita ser inmediata y centrada exclusivamente entre dos agentes, tal y como prevén los modelos más sencillos basados en el *tit for tat*, sino que ésta también puede tener un carácter difuso y diferido. El agente A puede ser recíproco con el agente B en un momento presente, y el agente B puede devolverle la reciprocidad en algún momento lejano del futuro, en lugar de hacerlo inmediatamente, esto es, la reciprocidad puede ser diferida. Por su parte, la reciprocidad será difusa cuando no sé dé entre los agentes A y B, sino que el agente A coopere con el agente B, B con C, C nuevamente con A o con D y así sucesivamente. Dejo aquí de lado estas complicaciones.

ción de dilema del prisionero iterado con dos jugadores egoístas, la cooperación mutua puede ser un equilibrio estable si ambos siguen una estrategia de *tit for tat*. Las condiciones mencionadas por Axelrod consisten en que la «sombra del futuro» sea lo suficientemente importante para los individuos, así como que éstos tengan memoria y capacidad de reconocimiento. Dicho de otro modo, es necesario que los individuos otorguen un importante valor a las futuras interacciones y que sean capaces de reconocerse unos a otros, así como de recordar cuál fue la conducta pasada del individuo con el que van a interactuar ahora.

De este modo, se han realizado importantes avances en la investigación de la emergencia de equilibrios cooperativos aplicando el instrumental de la teoría de juegos y de la TER-e. Sin embargo, el *tit for tat* fracasa al tratar de explicar la emergencia de dichos equilibrios¹² en la mayoría de situaciones importantes, y existe numerosa evidencia sobre sus limitaciones. El mismo Axelrod (1986b, 1997), sin ir más lejos, demuestra que el *tit for tat* no funciona cuando se encuentran más de un puñado de jugadores en un dilema del prisionero iterado jugado por agentes egoístas. En un contexto así, para que la cooperación pueda convertirse en una estrategia estable, se necesitará una metanorma que prescriba sancionar tanto a los que violan la norma de primer orden, como a aquéllos que se abstienen de sancionar a éstos últimos. Me referiré aquí brevemente a alguna evidencia empírica disponible que cuestiona la viabilidad de dar cuenta de la cooperación social a partir de la reciprocidad débil. Existe, por ejemplo, numerosa evidencia de tipo experimental basada en la utilización de juegos del ultimátum en diferentes culturas y con la participación de miles de individuos en los que sólo una exigua minoría se comporta según lo previsto por la TER-e¹³. Como es bien sabido, en el juego del ultimátum hay dos jugadores y algún bien valioso y divisible a repartir (en los experimentos citados, 10 dólares). Una de las partes (el ofertante) le hace una oferta de repartición a la otra parte, el receptor, que sólo puede aceptar o rechazar la oferta. Si el receptor acepta, el dinero se reparte según lo acordado, si rechaza, ambos lo pierden todo. Los experimentos citados, además, se llevaron a cabo en condiciones de anonimato entre las partes, con lo cual aumentaron los incentivos para defraudar. En general, los proponentes hicieron ofertas mucho más altas de lo previsto y, de modo mucho más importante, los receptores rechazaron ofertas de cuantías considerables si no las percibían como justas. Sería de esperar que un ofertante racional ofreciera un reparto 9-1 y que un receptor también racional lo aceptara. Se puede tratar de explicar la conducta de los ofertantes apelando a que no tienen manera de saber si el receptor es racional, pero no

12. En la literatura, suele sostenerse que esos equilibrios cooperativos son normas sociales. No obstante, comparto el punto de vista de Jon Elster (2007, 2008a), según el cual un equilibrio cooperativo fruto de la iteración de un dilema social no es una norma social.

13. Véase, por ejemplo, Bowles (2004, 2005), Bowles y Choi (2003), Bowles y Gintis (2000a, 2000b, 2001a, 2001b, 2002, 2003), Bowles y Naidu (2005), Camerer, Loewenstein y Rabin (2004), Gintis (2000a, 2000b, 2001a, 2003, 2005, 2006, 2007), Gintis, Bowles, Boyd y Fehr (2008), Henrich, Boyd, Bowles, Camerer, Fehr, Gintis (2005) y Mansbridge (1990).

se puede explicar la conducta de los receptores sin salirse de los supuestos del *Homo economicus*.

En los mismos experimentos, se cambiaron las reglas y se pasó a jugar juegos del dictador. La diferencia entre este juego y el del ultimátum consiste en que en este tipo de juego el receptor no cuenta con la opción de rechazar la oferta, de modo que el ofertante puede hacer la distribución que le parezca sin tener que preocuparse por la racionalidad de la otra parte.

Tampoco en este caso, en el que el diseño institucional ofrece unos incentivos aun más importantes para la conducta egoísta, la conducta prevista por la teoría económica fue mayoritaria entre los ofertantes. Puede descartarse también, por tanto, que en los juegos del ultimátum los ofertantes no ofrezcan 9-1 simplemente para evitar que sus ofertas sean rechazadas.

He destacado esta evidencia experimental simplemente a modo de ejemplo. En las referencias citadas, pueden encontrarse muchos más experimentos con resultados igualmente inequívocos.

Pero la evidencia contra la plausibilidad de la reciprocidad débil como única base de la cooperación en las sociedades humanas no se limita a resultados experimentales. La reciprocidad débil puede explicar por qué los individuos cooperan cuando existen expectativas de futuras interacciones, pero tiene muchas más dificultades para explicar por qué se coopera en situaciones con características de última ronda o de jugadas de ronda única. En el mundo social fuera de contextos experimentales, la reciprocidad débil no puede dar cuenta de cómo pudo sostenerse la cooperación en los numerosos momentos a lo largo de nuestra historia evolucionaria en los que el *Homo sapiens* se ha visto enfrentado a profundas crisis que pudieron llevarnos a la desaparición. Paradójicamente, por tanto, la estrategia no sirvió para asegurar la cooperación en los momentos en que era más importante que lo hiciese (Bowles y Gintis, 2003). De hecho, es prácticamente imposible que la especie humana hubiese tenido éxito evolucionario si hubiese albergado únicamente motivaciones egoístas (Boyd, Gintis, Bowles y Richerson, 2005; Jenks, 1990). Esto fue especialmente así durante el pleistoceno, que, como es bien sabido, ha sido el período más largo de nuestra historia evolutiva. Durante ese período, además, se desarrollaron las principales características de nuestra mente (por ejemplo, nuestras motivaciones), que apenas ha evolucionado desde entonces (Miller, 2008). La tesis según la cual la especie humana no habría podido sobrevivir durante el pleistoceno si hubiese albergado únicamente motivaciones egoístas, ha sido convincente apoyada por modelos formales de simulación multiagente¹⁴ en reiteradas ocasiones. Dichos modelos reproducen las características básicas de las poblaciones humanas de aquel momento¹⁵ (que conocemos gracias a los descubrimientos arqueológicos sobre la materia).

14. Sobre simulación multiagente, véase, por ejemplo, Gilbert y Troitzsch (2006).

15. Véase, por ejemplo, Bowles (2005), Bowles y Choi (2003), Bowles y Gintis (2001a, 2003), Bowles y Naidu (2005), Boyd, Gintis, Bowles y Fehr (2005), Gintis, Bowles, Boyd y Fehr (2005, 2008), Kaplan y Gurven (2005).

Además de lo mencionado, en muchos de aquellos casos de crucial importancia en nuestra historia evolucionaria que toman la forma de juegos de dilemas iterados con n jugadores y en los que la estabilidad de la cooperación podría ser explicada a través de la interacción entre individuos egoístas, la cooperación basada exclusivamente en el *tit for tat* constituye un equilibrio altamente inestable (Bowles y Gintis, 2003).

Finalmente, existe evidencia de innumerables e importantes conductas pro-sociales en la vida cotidiana que no pueden ser explicadas en términos de reciprocidad débil (Bowles y Gintis, 2003). La gente coopera habitualmente en situaciones del día a día que pueden ser modelizadas en términos de dilema social y en las que, según los supuestos de la TER-e, el equilibrio debería ser la no cooperación mutua. Los ciudadanos y las ciudadanas participan más o menos activamente en organizaciones y campañas cívicas de todo tipo, acostumbran a cumplir la ley y lo requerido por las normas de cooperación incluso en aquellos casos en los que no son observados, actúan de forma altruista en multitud de situaciones en las que tienen la oportunidad de ayudar a otros semejantes que lo necesitan, etc. En palabras de Gintis (2006: 108):

While nearly everyone behaves amorally on some occasions, and some behave amorally much of the time, there is normally a sufficient reserve of moral behavior, including the motivation to punish the moral transgressions of others, to maintain a high level of conformity with group morality.

La reciprocidad fuerte

La reciprocidad fuerte, la motivación para cooperar condicionalmente con los demás cooperadores y sancionar incondicionalmente a aquellos que violen las reglas de cooperación, es la forma de motivación más importante en la especie humana y se encuentra englobada dentro de lo que he denominado *motivaciones sociales*. Es la motivación cuantitativamente más importante que se identifica en condiciones experimentales y explica satisfactoriamente los resultados de los experimentos citados anteriormente. Por ejemplo, en el caso del juego del ultimátum, el rechazo de los receptores de ofertas consideradas injustas se explica fácilmente como la sanción que el receptor impone al proponentor (incluso incurriendo en costes personales) por haber violado una regla de equidad. El hecho de que una amplia mayoría de los proponedores efectúen ofertas generosas incluso en juegos del dictador, puede explicarse igual de fácilmente por la voluntad de estos de actuar conforme a dichas normas¹⁶.

En la reciprocidad fuerte, la disposición a cooperar es condicional, mientras que la predisposición a sancionar es altruista¹⁷. Que la cooperación es con-

16. Véase, por ejemplo, Bowles (2004, 2005), Bowles y Choi (2003), Bowles y Gintis (2000a, 2000b, 2001a, 2001b, 2002, 2003), Bowles y Naidu (2005), Gintis (2000a, 2000b, 2001a, 2003, 2006, 2007), Gintis, Bowles, Boyd y Fehr (2008).

17. Véase la nota anterior.

dicional, resulta bastante evidente, puesto que sólo se coopera con aquellos que también lo hacen. Pero el sancionar es incondicional y altruista. Se sanciona en todos los casos, con lo cual se genera un beneficio colectivo, aunque al hacerlo se incurra en costes individuales y no existan expectativas de ganancia futura. Un acto tan irracional (en el sentido técnico del término) como la venganza puede interpretarse fácilmente en clave de reciprocidad fuerte (Bowles y Gintis, 2001b).

Los cooperadores fuertes sancionan principalmente las motivaciones no cooperativas independientemente de cuales sean sus resultados, aunque los malos actos con buenas motivaciones también suelen ser sancionados, pero con menor intensidad (Gintis, 2000a; Gintis, Bowles, Boyd y Fehr, 2005). Sí, por ejemplo, el agente A trata de ayudar al agente B de buena fe, pero, debido a su torpeza, le termina causando un perjuicio, puede que B sancione a A aunque no en todos los casos y con una intensidad menor que si (B creyese que) A hubiese tenido realmente la intención de causar un perjuicio. Por el contrario, si el agente A trata de aprovecharse del agente B en el transcurso de una interacción, pero es tan torpe que termina por generar un beneficio para B y un perjuicio para sí mismo, cabe esperar que A recibirá la represalia de B, no su gratitud.

Un argumento adicional a favor de que los agentes motivados por la reciprocidad fuerte sancionan principalmente las motivaciones no cooperativas viene dado por el hecho de que, en los experimentos citados, cuando la oferta no equitativa en un juego del ultimátum no era hecha por un ofertante humano, sino por un ordenador de forma aleatoria, los receptores tendían a aceptar todas las ofertas fuese cual fuese su monto. Esta reacción por parte de los receptores tiene sentido desde el punto de vista de la reciprocidad fuerte. En este caso, el receptor no puede atribuir al ordenador la intención de explotarlo o de aprovecharse de él, de modo que no se activan emociones como la indignación o el enfado y no se produce sanción alguna.

Un agente motivado por la reciprocidad fuerte lo estará para seguir las reglas de la cooperación vigentes en una sociedad determinada. La reciprocidad fuerte ofrece, por tanto, un mecanismo para la aplicación descentralizada de normas sociales (Sethi y Somanathan, 2005). De este modo, la presencia de agentes de este tipo sirve como elemento de disuasión para potenciales gorriones y convierte en estable la cooperación (Fehr y Fischbacher, 2005; Sethi y Somanathan, 2005).

El cooperador fuerte no es indiferente al bienestar de los demás ni a los procesos que determinan los resultados sociales. Esos procesos deben ser equitativos (Bowles y Gintis, 2001b). En tiempos pasados, cuando la sociedad estaba dividida en estamentos sociales que contaban con privilegios basados en concepciones religiosas o ciertas ideas por el estilo, una situación justa podía ser altamente desigual. Sin embargo, en las sociedades modernas, la justicia suele entenderse como equidad. Suele considerarse que todos los individuos deben contar con las mismas oportunidades y que los resultados del azar o de procesos fuera de nuestro control deberían también repartirse de forma igualitaria.

No sucede lo mismo con los beneficios obtenidos a través del mérito y del esfuerzo (Bowles y Gintis, 2001b). Por ejemplo, en los experimentos referidos en este trabajo, los proponentes tendían a hacer ofertas más generosas cuando percibían que estaban en la posición de ofertantes por pura casualidad que cuando tenían que ganársela a través de algún proceso previo. Por tanto, la reciprocidad fuerte incorpora una clara concepción de la equidad y de la solidaridad como motores rectores de la vida social. Entre todos los iguales, debe existir un balance ajustado de derechos y de obligaciones capaz de regular el intercambio social.

Conclusiones

El presente trabajo parte de la premisa según la cual el hecho de que los modelos de la TER-e adopten de forma habitual supuestos falsos o excesivamente incompletos supone una enorme limitación para la capacidad explicativa y predictiva de la teoría. Esa premisa general puede concretarse en una más específica: concretamente, este trabajo ha partido de la idea de que el hecho de asumir de forma casi rutinaria el monismo motivacional y adoptar el egoísmo como el único tipo de motivación que puede dar cuenta de la acción social, supone un enorme lastre para los modelos de teoría de la elección racional.

A lo largo de estas páginas, se ha tratado de mostrar que existe una pluralidad de motivaciones que, junto con el egoísmo, dan cuenta de una porción importante de la conducta humana y que no existe ninguna justificación para excluirlas sistemáticamente de los modelos. Se ha prestado una atención especial a la motivación cuantitativamente más importante en el ser humano: la reciprocidad fuerte.

Los autores de la TER-e argumentan que, si relajamos el supuesto de egoísmo, la teoría pierde parsimonia, poder explicativo e informatividad, y corre el riesgo de volverse tautológica. En primer lugar, se podría responder que, aunque eso fuese cierto, de ningún modo justificaría la elección rutinaria del supuesto de egoísmo (Elster, 2007). Si realmente fuese cierto que el supuesto de egoísmo es el único susceptible de ser modelizado formalmente y, al mismo tiempo, como hemos visto en este trabajo, el uso exclusivo de dicho supuesto da lugar a malas explicaciones y predicciones en multitud de casos relevantes, la conclusión es lamentable para la ciencia social, a saber, sería imposible construir modelos formales que den cuenta de lo que acontece en multitud de casos relevantes.

Pero, pese a que la crítica debe tomarse en serio, no resulta insalvable. No existe ningún problema fundamental que impida construir modelos formales que contemplen una pluralidad de supuestos motivacionales, siempre y cuando se identifiquen claramente *a priori* el papel y la intensidad de las mismas, de modo que podamos construir hipótesis que cumplan con el requisito de ser falsables. Muchas de las referencias citadas en estas páginas corresponden a trabajos de autores que construyen modelos formales que no sólo contemplan

motivaciones egoístas, sino también otras motivaciones como el altruismo¹⁸. No obstante, en el caso concreto de la reciprocidad fuerte, es cierto que, hasta el momento, los intentos de elaborar funciones de utilidad basadas en dicha motivación han tenido como resultado funciones extremadamente complejas y difíciles de manejar¹⁹. Ha habido también intentos de capturar la noción de reciprocidad fuerte a partir de modelos más sencillos. Sin embargo, su éxito ha sido relativo, además se trataba de modelos que contienen cajas negras (Fehr y Fischbacher, 2005).

No obstante, es importante también tener en cuenta que no todo el trabajo riguroso en ciencias sociales pasa necesariamente por la utilización de modelos matemáticos con funciones de utilidad. Es perfectamente posible construir explicaciones basadas en modelos de mecanismos que van mucho más allá de los supuestos contemplados en este trabajo y que incluyen motivaciones tan variadas como las emociones o las normas²⁰. En este sentido, los modelos de mecanismos constituyen una herramienta idónea para incluir motivaciones distintas del egoísmo.

Referencias bibliográficas

- AGUIAR, F.; BRAÑAS-GARZA, P.; COBO-REYES, R.; JIMÉNEZ, N. y MILLER, L. (2007). *Gender Based Prescriptions. Evidence for altruism*. IESA Working Paper Series. [www.iesaa.csic.es, consulta: 16 de marzo de 2009]
- AGUIAR, F. y FRANCISCO, A. de (2008). «Can rational choice cope with identity?». *ISA Forum of Sociology*. Barcelona, 5-8 de septiembre.
- ANDREONI, J. y VESTERLUND, L. (2001). «Which is the fair sex? Gender Differences in Altruism». *Quarterly Journal of Economics*, 116 (1), 293-312.
- ARNSPERGER, C. y VAROUFAKIS, Y. (2003). «Toward a theory of solidarity». *Erkenntnis*, 59, 157-188.
- AXELROD, R. (1986a). *La evolución de la cooperación*. Madrid: Alianza.
- (1986b). «An evolutionary approach to norms». *American Political Science Review*, 80 (4).
- (1997). *La complejidad de la cooperación: Modelos de cooperación y colaboración basados en los agentes*. México, D.F.: Fondo de Cultura Económica, 2004.
- BARBERA, F. (2004). *Meccanismi sociali: elementi di sociologia analitica*. Bolonia: Il Mulino.
- BECKER, G. (1986). «The economic approach to human behavior». En: ELSTER, J. (comp.) (1986). *Rational choice*. Oxford: Basil Blackwell.
- BOUDON, R. (1981). *La lógica de lo social: Una introducción al análisis sociológico*. Madrid: Rialp.
- (2003). *Raison, bonnes raisons*. París: Presses Universitaires de France.

18. Véase, por ejemplo, Arnsperger y Varoufakis (2003), Kolm (1983), Mansbridge (1990), Margolis (1990), Sánchez-Cuenca (2007) y Wolfelsperger (2001).

19. A este respecto, véase Falk y Fischbacher (2005), Fehr y Fischbacher (2005) y Gintis (2000b).

20. Véase, por ejemplo, Barbera (2004), Boudon (1981, 2003, 2008), Cherkaoui (2005), Elster (1988, 1990a, 1990b, 1990c, 1990d, 1991a, 1995, 1997, 1998, 2001, 2002, 2006, 2007, 2008a, 2008b), Hedström (2005), Hedström y Swedberg (1998).

- (2008). *Essais sur la théorie générale de la rationalité*. París: PUF.
- BOWLES, S. (2004). *Microeconomics: Behavior, institutions and evolution*. Nueva Jersey: Princeton University Press.
- (2005). «Equality's fate: toward a natural history». *The Co-evolution of Behaviors and Institutions*. Santa Fe: Santa Fe Institute.
- BOWLES, S. y CHOI, J-K (2003). «The first property rights revolution». *The Co-evolution of Behaviors and Institutions*. Santa Fe: Santa Fe Institute.
- BOWLES, S. y GINTIS, H. (2000a). *The evolution of reciprocal preferences*. Working Paper. Santa Fe: Santa Fe Institute.
- (2000b). «Walrasian economic in retrospect». *Quarterly Journal of Economics*, 115 (4), 1411-1439.
- (2001a). *In search of homo economicus: behavioral experiments in 15 small-scale societies*. Working Paper. Santa Fe: Santa Fe Institute.
- (2001b). «¿Ha pasado de moda la igualdad? El *homo reciprocans* y el futuro de las políticas igualitarias». En: GARGARELLA, R. y OVEJERO, F. (2001). *Razones para el socialismo*. Barcelona: Paidós.
- (2002). *Prosocial emotions*. Working Paper. Santa Fe: Santa Fe Institute.
- (2003). *The evolution of cooperation in heterogeneous population*. Working Paper. Santa Fe: Santa Fe Institute.
- (2005). «Social capital, moral sentiments and community governance». En: GINTIS, H. et al. (eds.) (2005). *Moral sentiments and material interests: the foundations of cooperation in economic life*. The MIT Press.
- BOWLES, S. y NAIDU, S. (2005). «Evolutionary dynamics of class structure». *The Co-evolution of Behaviors and Institutions*. Santa Fe: Santa Fe Institute.
- BOYD, R.; GINTIS, H.; BOWLES, S. y RICHEERSON, P. (2005). «The evolution of altruistic punishment». En: GINTIS, H. et al. (eds.) (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. The MIT Press.
- BRANAS-GARZA, P. (2006). «Why gender based game theory». *The Papers 06/08*. Universidad de Granada.
- CAMERER, C.; LOEWENSTEIN, G. y RABIN, M. (2004). *Advances in behavioral economics*. Princeton: Princeton University Press.
- CHARNESS, G. y RABIN, M. (2002). «Understanding social preferences with simple test». *Quarterly Journal of Economics*, 117, 817-869.
- CHERKAOUI, M. (2005). *Invisible codes: Essays on generative mechanisms*. Oxford: The Bardwell Press.
- COLEMAN, J. (1990). *Foundations of Social Theory*. Cambridge: Harvard University Press.
- COSMIDES, L. y TOOBY, J. (1992). *The adapted mind*. Nueva York: Oxford University Press.
- DAWES, R. y THALER, R. (1988). «Cooperation». *The Journal of Economic Perspectives*, 2 (3), 187- 197.
- ELSTER, J. (1988). *Uvas amargas: Sobre la subversión de la racionalidad*. Barcelona: Península.
- (1990a). «Selfishness and altruism». En: MANSBRIDGE, J. (1990). *Beyond self-interest*. Chicago: University of Chicago.
- (1990b). *El cambio tecnológico: Investigaciones sobre la racionalidad y la transformación social*. Barcelona: Gedisa.
- (1990c). *Tuercas y tornillos: Una introducción a los conceptos básicos de las ciencias sociales*. Barcelona: Gedisa.

- (1990d). «Racionalidad y normas sociales». *Revista Colombiana de Sociología*, 1 (1) (enero-julio). Bogotá: Universidad Nacional de Colombia.
- (1991a). *El cemento de la sociedad: las paradojas del orden social*. Barcelona: Gedisa.
- (1991b). «Racionalidad, moralidad y acción colectiva». En: AGUIAR, F. (1991). *Intereses individuales y acción colectiva*. Madrid: Pablo Iglesias.
- (1995). *Psicología política*. Barcelona: Gedisa.
- (1997). *Egonomics: análisis de la interacción entre racionalidad, emoción, preferencias y normas sociales en la economía de la acción individual y sus desviaciones; y el relato autobiográfico «Going to Chicago»*. Barcelona: Gedisa.
- (1998). «A plea for mechanisms». En: HEDSTRÖM, P. y SWEDWERG, R. (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge Press.
- (2000). «La envidia en la vida social». En: ELSTER, J. (2000). *Las limitaciones del paradigma de la elección racional: Las ciencias sociales en la encrucijada*. Valencia: Institució Alfons el Magnànim.
- (2001). *Sobre las pasiones: Emoción, adicción y conducta humana*. Barcelona: Paidós.
- (2002). *Alquimias de la mente: La racionalidad y las emociones*. Barcelona: Paidós / El Roure.
- (2006). *Raison et raisons*. Lecciones inaugurales del Collège de France. París: Collège de France / Fayard.
- (2007). *Explaining Social Behavior: More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
- (2008a). «Social Norms and the Explanation of Behavior». En: HEDSTRÖM, P. (2009) (ed.). *The Oxford Handbook of Analytical Sociology*.
- (2008b). «Emotions». En: HEDSTRÖM, P. (2009) (ed.). *The Oxford Handbook of Analytical Sociology*.
- FALK, A. y FISCHBACHER, U. (2005). «Modeling strong reciprocity». En: GINTIS et al. (eds.) (2005). *Moral sentiments and material interests: the foundations of cooperation in economic life*. The MIT Press.
- FEHR, E. y FALK, A. (2002a). *Physiological foundations of incentives*. Working Paper. [http://ssrn.com/abstract_id=294287, consulta: marzo de 2008]
- (2002b). «Physiological foundations of incentives». *European Economic Review*, 46, 687-724.
- FEHR, E. y FISCHBACHER, U. (2005). «The economics of strong reciprocity». En: GINTIS, H. et al. (eds.) (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. The MIT Press.
- FRIEDMAN, M. (1953). *Essays in positive economics*. Chicago: University of Chicago Press.
- GARCÍA, M. (2005). *Racionalidad, psicología evolucionaria y cooperación* [texto inédito].
- GILBERT, N. y TROITZSCH, K.G. (2006). *Simulación para las ciencias sociales: Una guía práctica para explorar cuestiones sociales mediante el uso de simulaciones informáticas*. Madrid: McGraw-Hill.
- GILLIGAN, C. (1982). *In a different voice*. Cambridge: Harvard University Press.
- GINTIS, H. (2000a). *Strong reciprocity and human sociability*. [www.ideallibrary.com]
- (2000b). *Game theory evolving. A problem-centered introduction to modelling strategic interaction*. Princeton: Princeton University Press.
- (2001). *The puzzle of prosociality*. [www-unix.oit.umass.edu/~gintis/]
- (2003). *Towards a unity of human behavioral sciences*. Working Paper. Santa Fe: Santa Fe Institute.

- (2006). «Towards a unity of the human behavioral sciences». *Papers: Revista de Sociología*, 80.
- (2007). *Bayesian rationality and social norms*. Working Paper. Santa Fe: Santa Fe Institute.
- GINTIS, H. et al. (eds.) (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. The MIT Press.
- GINTIS, H.; BOWLES, S.; BOYD, R. y FEHR, E. (2005). «Moral sentiments and material interests: Origins, evidence and consequences». En: GINTIS, H. et al. (eds.). *Moral sentiments and material interests: The foundations of cooperation in economic life*. The MIT Press.
- (2008). «Strong reciprocity and the roots of human morality». *Social Justice Research*, 21, 241-253.
- HAMILTON, W. (1963). «The evolution of altruistic behavior». *American Naturalist*, 97.
- HARDIN, R. (1982). *Collective action*. Baltimore: John Hopkins University Press.
- HEDSTRÖM, P. (2005). *Dissecting the social*. Cambridge: Cambridge University Press.
- HEDSTRÖM, P. y SWEDWERG, R. (1998). *Social mechanisms. An analytical approach to social theory*. Cambridge: Cambridge Press.
- HENRICH, J.; BOYD, R.; BOWLES, S.; CAMERER, C.; FEHR, E. y GINTIS, H. et al. (2005). «“Economic man” in cross-cultural perspective: behavioral experiments in 15 small-scale societies». *Behavioral and Brain Sciences*, 28, 795-855.
- JENKS, C. (1990). «Varieties of altruism». En: MANSBRIDGE, J. (1990). *Beyond self-interest*. Chicago: University of Chicago.
- KAPLAN, H. y GURVEN, M. (2005). «The natural history of human food sharing and cooperation: A review and a new multi-individual approach to the negotiation of norms». En: GINTIS, H. et al. (eds.) (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. The MIT Press.
- KOLM, S. (1983). «Altruism and efficiency». *Ethics*, 94 (1), 18-65.
- LOEWE, G. (2006). «The development of a theory of rational intertemporal choice». *Papers: Revista de Sociología*, 80, 195-221.
- MANSBRIDGE, J. (1990). *Beyond self-interest*. Chicago: University of Chicago.
- (1990). «Expanding the range of formal modelling». En: MANSBRIDGE, J. (1990). *Beyond self-interest*. Chicago: University of Chicago.
- MARGOLIS, H. (1990). «Dual Utilities and Rational Choice». En: MANSBRIDGE, J. *Beyond self-interest*. Chicago: University of Chicago.
- MARÍ-KLOSE, P. (2000). *Elección racional*. Madrid: Centro de Investigaciones Sociológicas.
- MAYNARD-SMITH, J. (1978). «La evolución del comportamiento». *Investigación y Ciencia*, 26, 116-127.
- MILLER, L. (2008). «Coordinación y convención». En: AGUIAR, F.; BARRAGÁN, J. y LARA, N. (2008) (ed.). *Economía, sociedad y teoría de juegos*. Madrid: Mc Graw Hill.
- NOGUERA, J.A. (2006). «Why We Need an Analytical Sociological Theory». *Papers: Revista de Sociología*, 80, 7-28.
- OLSON, M. (1965). *La lógica de la acción colectiva: bienes públicos y la teoría de grupos*. México D.F.: Limusa. Grupo Noriega, 1992.
- OVEJERO, F. (2008). «La teoría de juegos: ¿una teoría social?». En: AGUIAR, F.; BARRAGÁN, J. y LARA, N. (2008). *Economía, sociedad y teoría de juegos*. Madrid: Mc Graw Hill.
- SALLY, D. (1995). «Conversation and co-operation in social dilemmas». *Rationality and Society*, 7, 58-92.

- SÁNCHEZ-CUENCA, I. (2007). «Cooperar por principio». *Revista Internacional de Sociología*, 46, 11-35.
- (2008). «A preference for selfish preferences. The problem of motivations in rational choice political science». *Philosophy of the Social Sciences*, 38 (3), 361-378.
- SCOTT, J. T.; MATLAND, R.E.; MICHELBACH, P.A. y BORNSTEIN, B.H. (2001). «Just Deserts: An experimental study of distributive justice norms». *American Journal of Political Science*, 45 (4), 749-767.
- SEN, A. (1980). «Description as choice». *Oxford Economic Papers*, 32, 353-369.
- (1986). «Los tontos racionales: una crítica de los fundamentos conductistas de la teoría económica». En: HAHN, F. y HOLLIS, M. (comps.). *Filosofía y teoría económica*. México: Fondo de Cultura Económica.
- SETHI, R. y SOMANATHAN, E. (2005). «Norm compliance and strong reciprocity». En: GINTIS, H. et al. (eds.) (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life*. The MIT Press.
- TAYLOR, M. (1982). *Community, anarchy and liberty*. Cambridge: Cambridge University Press.
- (1987). *The possibility of cooperation*. Cambridge: Cambridge University Press.
- (1991). «Racionalidad y acción colectiva revolucionaria». En: AGUIAR, F. *Intereses individuales y acción colectiva*. Madrid: Pablo Iglesias.
- TRIVERS, R. (1971). «The evolution of reciprocal altruism». *Quarterly Review of Biology*, 46, 35-57.
- ULLMANN-MARGALIT, E. (1977). *The emergence of norms*. Oxford: Clarendon Press.
- WOLFELSPERGER, A. (2001). «La modelisation économique de la rationalité axiologique: Des sentiments moraux aux mécanismes sociaux de la moralité». En: BOUDON, P.; DEMEULENAERE, P. y VIALE, R. (2001). *L'explication des normes sociales*. Paris: Presses Universitaires de France.
- YUNG-AN HU y DAY-YANG LIU (2003). «Altruism versus egoism in human behavior of mixed motives». *American Journal of Economics and Sociology*, 62 (4).