

# *Preferencias sociales y egoísmo racional*

## *Social preferences and rational egoism*

ALEJANDRO ROSAS

*Universidad Nacional de Colombia (Colombia)*

Recibido: 22-1-2013

Aprobado definitivamente: 1-2-2013

### RESUMEN

Los economistas experimentales han demostrado que algunos humanos cooperan con base en preferencias sociales, pero el modelo de agente racional conocido como *Homo oeconomicus* no incluye preferencias sociales como causas de la conducta cooperativa. ¿Qué implica su existencia para el egoísmo racional defendido en los modelos microeconómicos dominantes y en la teoría de juegos? Se describen tres posibles respuestas: 1) preferencias sociales como mecanismos redundantes de reserva; 2) preferencias sociales para remediar la racionalidad imperfecta y 3) preferencias sociales como motor primario de la cooperación, porque el egoísmo racional no recomienda cooperar en dilemas de prisionero iterado con información imperfecta. Defiendo la tercera opción: agentes que carecen de preferencias sociales ven el engaño y la coerción como opciones racionales.

### PALABRAS CLAVE

COOPERACIÓN, DILEMA DE PRISIONEROS, EGOÍSMO RACIONAL, EXPLICACIÓN Y JUSTIFICACIÓN

### ABSTRACT

Experimental economics provides evidence that social preferences drive human cooperation in the lab, but the dominant microeconomic model of a rational agent, *Homo oeconomicus*, denies such preferences. Assuming the evidence is cogent, what follows for the claim that humans cooperate on the basis of rational egoism? I describe three possible answers: 1) social

preferences are backup mechanisms for rational egoism; 2) social preferences are required to remedy for imperfect rationality; and 3) social preferences are the primary motives for cooperation, because rational egoism does not recommend cooperation in the iterated prisoner's dilemma with imperfect or private information. I argue for option 3): rational egoists without social preferences see deception and coercion as rational options.

KEYWORDS:

COOPERATION, PRISONER'S DILEMMA, RATIONAL EGOISM, EXPLANATION AND JUSTIFICATION

### I. PREFERENCIAS SOCIALES COMO RETO AL *HOMO OECOMICUS*

EN ESTE ARTÍCULO DEFIENDO QUE LAS PREFERENCIAS SOCIALES (PS) no son simplemente un mecanismo alternativo o remedial para apoyar lo que el egoísmo racional recomienda independientemente de ellas. La tesis es que las preferencias sociales son indispensables, y que sin ellas, la interacción en un dilema de prisioneros (DP) repetido no conduciría a la cooperación entre agentes racionales egoístas. Postular que las PS son indispensables para la cooperación contradice el modelo económico dominante del agente racional, defendida también en la teoría de juegos. Su modelo es el de un agente que carece de PS y que coopera porque el egoísmo racional lo recomienda o exige. Formulada de este modo, la tesis es descriptiva y normativa al mismo tiempo. Hay dos versiones de la tesis normativa en la literatura. La teoría de juegos sostiene que un agente racional egoísta maximiza su utilidad cooperando porque cooperar es la mejor jugada en los DP iterados. Este postulado, conocido como el teorema «folk», ha inspirado varias pruebas formales que muestran la existencia de numerosos equilibrios de Nash con estrategias cooperativas en el DP iterado (Bowles y Gintis 2011, cap. 5). La otra versión de la tesis normativa es la del filósofo moral contractualista David Gauthier. Gauthier sostiene en *Morals by Agreement* (1986, 170) que en un DP, incluso en uno de una sola jugada, la mejor respuesta racional-egoísta es cooperar cuando los agentes son translúcidos o transparentes, es decir, cuando la probabilidad de leer sus intenciones en relación a la cooperación es mayor al mero azar.

La tesis normativa ha servido de base a la tesis descriptiva-explicativa: los seres humanos cooperamos en DP iterados porque somos capaces, como agentes racionales, de actuar de acuerdo con las normas de racionalidad. Pero ambas tesis, la normativa y la explicativa, han sido cuestionadas por científicos sociales que señalan el «lado oscuro» del egoísmo racional (Bowles & Gintis 2011, 5-6).

Estos autores reclaman la inclusión de las PS en el modelo. La pregunta que planteo aquí es: ¿qué pasa con el modelo clásico de agente racional una vez se incluyen las PS? Esta pregunta se puede descomponer en varias interrogantes. ¿Sigue siendo verdad que cooperar es la mejor opción para agentes racionales egoístas en un DP? Pero entonces, ¿por qué tenemos PS? ¿Realmente las necesitamos? Y si las necesitamos, ¿podemos mantener la tesis normativa según la cual el egoísmo racional recomienda cooperar en un DP, o debemos por el contrario abandonarla?

Para defender que las PS son indispensables, no voy a recurrir a teoremas o formalizaciones. Aunque hay desarrollos formales sobre juegos iterados con información imperfecta o privada que sugieren que las PS son indispensables (Bowles y Gintis 2011, cap 5), aquí doy un argumento de tipo no-formal, basándome en consideraciones plausibles sobre la psicología humana.

Voy a asumir que, al menos algunos humanos, deciden con base en PS en juegos económicos que se implementan en el laboratorio. Quizás no todo investigador en el área esté dispuesto a aceptar esto (Binmore y Shaked 2010). Sólo mencionaré en relación a este punto que los resultados experimentales que vienen acumulándose desde los primeros juegos de ultimátum en los años 80s parecen exigir una interpretación que recurre al concepto de PS. Una revisión reciente puede leerse en el capítulo 3 de Bowles y Gintis (2011). Para los propósitos presentes defino la PS como un interés, positivo o negativo, en el bienestar de otro(s) agente(s). Ese interés debe cumplir con una condición adicional: no puede ser instrumental, es decir, debe mantenerse incluso si el sujeto descubre que su propio bienestar no sería promovido, o sería perjudicado en una medida tolerable, por la satisfacción de ese interés. El interés positivo y no instrumental en el bienestar de otro se denomina también altruismo psicológico.

Algunos filósofos han sostenido que no es correcto confundir las PS positivas (es decir, el altruismo psicológico) con motivaciones propiamente morales. La motivaciones propiamente morales se fundan en juicios morales del tipo «X es indebido o incorrecto (wrong)» (Joyce 2006). Creo que esa postura es correcta. Pero eso no quita que las PS altruistas sean elementos necesarios para la moral, aunque no siendo suficientes se las pueda distinguir de la motivación propiamente moral. Decir que son necesarias equivale a decir que no podríamos hacer juicios morales sin ellas, aunque se necesiten además otros elementos. De Waal ha defendido la existencia de motivaciones pro-sociales en chimpancés y probablemente en nuestros ancestros hace seis millones de años (De Waal 2006). Varios autores interesados en la evolución humana han sostenido que la existencia simultánea de motivaciones pro-sociales y egoístas lleva a un tipo de vida social compleja, pero limitada por un alto grado de conflicto intra-grupal y, al parecer, incapaz de desarrollar cooperación a gran escala (Joyce 2006; Tomasello 2009; Kitcher 2011). Presumiblemente, la vida social de nuestros ancestros

era de ese estilo. Gracias a que evolucionamos una motivación propiamente moral, pudimos dar el salto a un tipo de cooperación más intensa. No sabemos exactamente como hizo la selección natural para evolucionar una motivación moral, pero es razonable suponer que de algún modo usó las motivaciones pro-sociales pre-existentes. Intuitivamente, dado que las motivaciones pro-sociales que evolucionaron por selección natural estaban inicialmente condicionadas al parentesco, y que la moralidad incorpora imparcialidad, alcanzarla requirió una identificación del individuo con el grupo de referencia, independientemente de relaciones de parentesco. Tuvo que emerger la conciencia de un «nosotros» más abarcadora que la que se requiere para una motivación pro-social con parientes cercanos. Pero sin un componente de altruismo psicológico, sólo quedaría el egoísmo racional como apoyo de la imparcialidad, y la moral se identificaría con él. Esta equivalencia entre moralidad y egoísmo racional es la tesis que cuestiono en este ensayo. Una tercera alternativa sostendría que las reglas morales no se basen ni en preferencias sociales ni en el egoísmo racional, sino en la «razón pura», estilo Kant. Comparto con Kant la idea de que la moral no puede reducirse al egoísmo racional (a lo Hobbes). Pero su idea de «razón pura» postula que todo agente racional es un fin en sí mismo. En un mundo humeano donde el valor es dependiente de los deseos subjetivos, ese estatus se otorga a un agente sólo con base en emociones empáticas que tenemos por evolución. Sería una distorsión equipararlas con una capacidad racional.

## II. TRES CONCEPCIONES ALTERNATIVAS DEL ROL DE LAS PREFERENCIAS SOCIALES

Asumiendo que al menos algunos seres humanos tienen preferencias sociales, preguntamos qué consecuencias tiene ese hecho para la tesis que dice que el egoísmo racional basta para explicar la cooperación humana. Hay tres respuestas posibles a esta interrogante.

La primera respuesta comparte la tesis que el comportamiento moral se puede *justificar y explicar* en la racionalidad como es concebida en la teoría de juegos. Habría un argumento, normativo y explicativo a la vez, que deriva de la racionalidad egoísta las estrategias morales de cooperación en dilemas de prisionero. En esta concepción las preferencias sociales y las emociones que las acompañan son redundantes. Podrían entonces funcionar como mecanismos de reserva.

Una segunda concepción está de acuerdo con la primera en que la racionalidad egoísta proporciona una *justificación* normativa de las estrategias morales y cooperativas, pero niega que el egoísmo racional explique la cooperación. Sostiene en cambio que los humanos no somos capaces, por defectos de principio en nuestra racionalidad, de actuar conforme al argumento normativo. En esta concepción las preferencias sociales y las emociones que las acompañan

tienen una función remedial: son «muletas» que nos ayudan a comportarnos como lo exige el egoísmo racional.

La tercera concepción va más allá de la segunda: rechaza incluso que la racionalidad egoísta, ella sola, en ausencia de preferencias sociales, justifique la cooperación en dilemas de prisionero iterado. Así, las preferencias sociales dejan de ser meras «muletas» que nos hacen capaces de seguir los mandatos de la racionalidad egoísta; más bien se necesitan para justificar normativamente la cooperación. En esta concepción la racionalidad egoísta no puede ni explicar, ni justificar, la cooperación a gran escala o moral que parece constituir un núcleo muy importante de la cultura humana. Esta es la concepción que defenderé en la sección 3. En lo que sigue expondré con algún detalle las dos primeras concepciones intentando mostrar por qué son insuficientes y por qué exigen dar el paso radical que nos lleva a la tercera concepción.

### II.1. PREFERENCIAS SOCIALES COMO MECANISMOS DE RESERVA O REDUNDANTES

La idea que la moral o el punto de vista moral se puede derivar del egoísmo racional es muy antigua. En cierto modo Platón fue su primer defensor cuando sugirió que una vida injusta suponía un alma psicológicamente desajustada, carente de armonía, propensa a un desarreglo en el modo de vida, y finalmente expuesta a frustración e infelicidad. Algunas doctrinas de los sofistas defendieron una idea semejante, aunque menos centrada en la armonía interna del alma. Su argumento señalaba que una vida de acuerdo a la justicia era materialmente preferible a la alternativa, consistente en intentar dominar a los demás por la fuerza y ponerlos al servicio de los propios intereses. Hobbes defendió esta idea en la época moderna, argumentado que una vida de acuerdo a las «leyes naturales» evita que nuestras vidas se conviertan en una guerra de todos contra todos y sean «*nasty, brutish and short*» (Hobbes 1981). Hume sostuvo una posición más ambigua, pero en algunos pasajes sugiere que una vida de acuerdo con la justicia es la opción racional basada en el interés egoísta o prudente de cada cual (Hume, 2000).

Entre los filósofos morales contemporáneos, David Gauthier se destaca como defensor de esta posición, que se conoce a veces como el «proyecto de reconciliación», por su intento de reconciliar la moral con la prudencia o el egoísmo de largo plazo (Kavka 1986; Hampton 1986). Los autores que participan en este proyecto defienden que la estructura básica de interacción humana, en donde la moral tiene sentido y cumple una función indispensable, es lo que en teoría de juegos se conoce como un dilema de prisioneros. En la versión de dos jugadores, la utilidad por cooperar ambos es mayor que la utilidad si ambos desertan, pero un desertor obtiene mayor utilidad «explotando» a un cooperador. Cuando el juego es de una sola interacción, el equilibrio de Nash, es decir, el

par de estrategias del cual ningún jugador debe desviarse de manera unilateral si quiere maximizar su utilidad, es la mutua deserción.

Gauthier (1986) cuestionó esta solución para los DP de una sola jugada (1986, 170). Su argumento es que los seres humanos podemos predecir la intención y la jugada de la contraparte con un acierto superior al azar. Eso es lo que él ha llamado la translucidez de las intenciones cooperativas (1986, 174). La translucidez produce un efecto nuevo en la estructura de la interacción. La intención de desertar en cualquiera de los jugadores causaría en la contraparte una deserción y ambos jugadores obtendrían la paga sub-óptima de la mutua deserción. Si la contraparte puede inferir mi jugada con un acierto superior al azar, lo que me conviene, lo que el egoísmo racional recomienda, es cooperar con aquellos que según mi predicción van a cooperar también y que leen mis intenciones así como yo leo las suyas. Así me aseguro de obtener la recompensa por cooperar en lugar del castigo por mutua deserción.

Si la cooperación es la mejor respuesta, es decir, la respuesta racional-egoísta, en un DP e incluso en DP de una sola jugada, ¿qué papel juegan las preferencias sociales, asumiendo que al menos algunos humanos las tienen? Como el egoísmo racional es suficiente para cooperar, las PS no pueden jugar sino un papel redundante. Los ingenieros a menudo diseñan en sus sistemas dos o más mecanismos para cumplir una misma función importante. El propósito de esta redundancia es asegurar que el sistema funcione: si un mecanismo falla, lo que nunca puede descartarse, el mecanismo de reserva lo reemplaza. Las PS podrían jugar el papel de un mecanismo de reserva en el ser humano para asegurar la cooperación, si el egoísmo racional llega a fallar.

Para los autores que defienden el egoísmo racional como suficiente para explicar y justificar la cooperación, las PS sólo podrían cumplir ese papel redundante, en caso de que acepten su existencia. De hecho, no es posible citar autores que defiendan esta posición, pero ello se debe a que no se han planteado el problema que suscita la coexistencia de egoísmo racional y preferencias sociales. Si miramos por ejemplo el caso de Gauthier, encontramos una posición peculiar. No sólo acepta que existen las PS, sino que reconoce que la mayoría de los seres humanos no cooperan por egoísmo racional. Se guían, más bien, por PS. En sus reflexiones, las PS aparecen como «*tuistic concerns*» o «*tuistic interests*» (Gauthier 1986, 327). Esta admisión lo obliga a modificar el papel del egoísmo racional: ya no es causa eficiente de la cooperación, al menos no en la mayoría de los seres humanos. Es, en cambio, un recurso para justificar la racionalidad de las verdaderas causas, las PS. Le otorga mucha importancia a ese papel justificante, que proporciona una «reconstrucción racional» de cooperación en dilemas de prisionero, porque de lo contrario recaería sobre ella una sospecha de irracionalidad (1986, 339). Con ello se salva el valor normativo del egoísmo racional y se sacrifica su papel explicativo, al menos para la mayoría de los hu-

manos. El egoísmo racional no desempeña el papel de motivo o causa habitual de la cooperación, que podría además fallar, si acaso, ocasionalmente. Por esta razón Gauthier no parece ser del todo un buen ejemplo del primer modelo. Esto suscita una duda: si el egoísmo racional no es el motivo habitual que nos impulsa a cooperar, ¿será que falla por principio, en lugar de ocasionalmente.

## II.2. RACIONALIDAD IMPERFECTA

La segunda explicación toma en serio la observación de que el egoísmo racional no es la causa común o habitual de la cooperación en los humanos. Además niega explícitamente que el egoísmo racional falle sólo ocasionalmente. Más bien, los egoístas racionales sufren bajo una imperfección de diseño que los incapacita para seguir lo que recomienda su naturaleza racional egoísta. La imperfección consiste en un descuento excesivo del futuro, es decir, las recompensas y los costos futuros se devalúan en comparación con los presentes. Si fuésemos agentes racionales perfectos, acataríamos el egoísmo racional y sus recomendaciones. Pero como descontamos el futuro de manera excesiva, no podemos cooperar basados en el egoísmo racional cuando la cooperación supone intercambio de beneficios diferidos en el tiempo. Las PS remedian este problema de diseño y alinean nuestra conducta con el egoísmo racional.

El descuento de los costos o recompensas futuras comparadas con las presentes tiene sentido para cualquier organismo diseñado por la selección natural. Desde el punto de vista de la supervivencia, el presente es más importante que el futuro y tiene prioridad, pues no hay futuro si no se sobrevive al presente. Pero la cooperación humana se basa en intercambios diferidos en el tiempo. Pongamos un ejemplo: si B me ayuda hoy con la expectativa de que yo lo ayude dentro de un mes, y amenaza con no volver a ayudarme si no cumplo cuando llegue el momento, aun así me puede tentar la idea de no ayudar a B y de renunciar a su ayuda futura. Cuando llegue el momento esperado por B, cumplir su expectativa significará para mí un costo presente, mientras que el beneficio de volver a contar con su ayuda es futuro y se puede ver como incierto. Descontamos el daño futuro que pueda acarreararnos no cumplir y en cambio el costo presente nos parece considerable. La decisión puede ser imprudente, y lo es más aún si sumamos el costo de una mala reputación por incumplimiento. B puede desquitarse arruinando mi reputación en los círculos sociales compartidos. Eso implica que podría perder no sólo la ayuda futura de B, sino la oportunidad de obtener beneficios cooperando con otros, pues ningún agente racional se arriesga a iniciar empresas cooperativas con agentes que no gozan de buena reputación. Pero aunque la decisión de actuar de ese modo es irracional e imprudente por la amenaza que pende sobre mi reputación, no puedo evitar, por el descuento del futuro, que el costo presente de ayudar a B me parezca mayor que el costo futuro de no ayudarlo.

Si la estructura de intercambios diferidos es esencial a la cooperación humana, el descuento del futuro es un obstáculo para la emergencia de cooperación. Ésta no puede evolucionar en un organismo cuya mente o cuya psicología esté dominada por un descuento del futuro. Hay incluso razones para pensar que ese tipo de cooperación es escasa, si no inexistente, en animales no humanos. Cabría que decir que, por su misma naturaleza, la cooperación típicamente humana se basa en valorar el futuro y el presente de manera semejante y en contradecir la tendencia de la selección natural a priorizar el presente, al menos en contextos de cooperación. Pero si el descuento del futuro está incorporado a nuestra racionalidad egoísta, ¿cómo hemos logrado contradecir esa tendencia? La respuesta es que la misma selección natural nos dotó con PS.

Este argumento fue utilizado por Frank (1988) para explicar la función de las emociones morales en la evolución del altruismo recíproco (Trivers 1971) como estrategia típica de la cooperación humana. Dado que la reciprocidad entre individuos no emparentados es beneficiosa, pero enfrenta el obstáculo del descuento del futuro, su evolución requirió de mecanismos adicionales al egoísmo racional, a saber, las emociones morales (simpatía, culpa, arrepentimiento, indignación). Ellas fueron necesarias para extender el altruismo hacia miembros no emparentados del grupo social. De ahí que autores como Joyce (2006) o también Kitcher (2011) distingan el altruismo psicológico que existe entre parientes del que puede existir entre individuos no emparentados de un mismo grupo. En este caso existe la tentación de hacer trampa cuando los beneficios son diferidos y la estructura de interacción es un dilema de prisioneros. Si todos ceden a esa tentación, la cooperación a gran escala no evoluciona.

En esta explicación las preferencias sociales evolucionan respecto de todos los miembros del grupo social para remediar el obstáculo causado por el descuento del futuro a las relaciones cooperativas con intercambio diferido de beneficios. El egoísmo racional sigue recomendando la cooperación en interacciones con la estructura de un dilema de prisioneros, pero no es suficiente para garantizar la conducta cooperativa en organismos que descuentan el futuro. Se trata de un obstáculo en el diseño mismo de los agentes cooperativos. Las PS remedian ese problema de diseño. La explicación se basa, entonces, en una premisa que dice que descontar el futuro es irracional en contextos de cooperación, incluso en agentes que carecen de PS.

La tercera explicación de la función de las PS, que exponemos en la sección siguiente, cuestiona esta premisa. Según ella, la cooperación no siempre es la estrategia racional para agentes racionales egoístas que carecen de PS. A veces puede pagar no-cooperar en juegos con la estructura de un dilema de prisionero iterado. ¿Cómo se puede defender esta postura, en vistas de las pruebas formales de la existencia de equilibrios de Nash en dilemas de prisionero iterados? Abordamos esta pregunta en la sección que sigue.



### III. EGOÍSMO RACIONAL Y ENGAÑO

Estrictamente hablando, es posible mostrar formalmente que existen equilibrios de Nash con estrategias cooperativas en dilemas de prisionero iterado (Bowles & Gintis 2011, cap. 5). Lo que sucede es que se hace difícil creer que los agentes que interactúan en condiciones reales puedan coordinar su comportamiento de modo que coincidan en sus jugadas cooperativas y obtengan así la recompensa por mutua cooperación (R), que es mayor en esos dilemas que el castigo por mutua deserción (P) o que alternar entre la paga de la tentación y la paga por ser explotado  $[(T + S)/2]$ . La dificultad consiste en que, en condiciones reales de interacción, la información es imperfecta o privada. El caso es especialmente grave cuando la información es privada, pues en ese caso los agentes no concuerdan en su representación de lo que cada uno hizo en interacciones pasadas y no pueden coordinar sus estrategias.

Esto afecta la tesis que sostiene que el egoísmo racional recomienda la cooperación. Para entender cómo la afecta, debemos preguntar por las causas de que la información sea imperfecta o privada en las interacciones sociales. De manera un tanto gruesa distingo dos tipos de causas: una es lo que podríamos llamar simplemente ruido en el entorno. Lo importante aquí para el caso humano es que el ruido se deriva de la complejidad del entorno y es independiente de las actitudes y propósitos de los agentes. El otro tipo de causas, en cambio, no es independiente de las actitudes de los agentes. Examinemos esta última causa primero. Una de las razones más frecuentes por las que la información es privada en las interacciones humanas es que los mismos agentes distorsionan la información o la esconden con el propósito de engañar a sus contrapartes. El engaño sucede en el mundo biológico de manera pre-programada. Pero en algunos organismos —con certeza en el caso humano y con alguna probabilidad en otros primates— el engaño es una conducta deliberada que el individuo astutamente diseña. En el contexto de la cooperación en dilemas de prisionero, el propósito del engaño es hacerle creer a la contraparte que tengo intenciones cooperativas, cuando en realidad lo que busco es explotar su jugada cooperativa. Si el engaño tiene éxito, accedo a la paga más alta en un dilema de prisioneros, a saber T, la paga del desertor que explota a un cooperador. La estrategia de explotar la cooperación del otro es en este caso adaptativa, aumenta el éxito reproductivo y, desde el punto de vista de la racionalidad egoísta es racional. En otras palabras, si el engaño es posible, y la experiencia humana cotidiana sugiere que a menudo lo es, no puede sostenerse que el egoísmo racional recomienda unívocamente cooperar en interacciones con la estructura de un dilema de prisioneros iterado. Esta no es una tesis nueva. Ha sido defendida por Mary

Gibson (1977), por ejemplo, en el contexto de las democracias modernas, en donde la información sobre la estrategia de algunos agentes sólo está disponible para la gran mayoría a través de los medios de comunicación, susceptibles de ser manipulados por intereses particulares. En general, no es descabellado suponer que varias formas de coerción y engaño son opciones racionales para agentes que no tienen preferencias sociales.

Por esta vía llegamos a una conclusión importante: cuando la información es imperfecta o privada debido a estrategias de engaño, la existencia de equilibrios de Nash con estrategias cooperativas en juegos de información perfecta no es aplicable. Aquí el egoísmo racional no favorece la emergencia de la cooperación. Pues un agente racional, con suficiente destreza en prácticas mendaces, puede obtener más utilidad desertando que cooperando. Así, un agente racional egoísta sin preferencias sociales no necesariamente ve la cooperación como la vía más expedita para maximizar sus utilidades. Un problema para la estrategia basada en el engaño es que no maximiza la utilidad si todo agente la adopta (esto recuerda al argumento kantiano). La estrategia funciona sólo como parasitaria de estrategias cooperativas (y al parásito le tiene sin cuidado que su estrategia no sea universalizable). Es decir, su existencia depende de la existencia previa de estrategias cooperativas que se convierten en blanco de explotación. Por otra parte, no es completamente imposible que la cooperación surja con estrategias basadas en el engaño. Sería en todo caso una cooperación con distribución inequitativa de costos y beneficios. El engaño sería una manera de lograr que los participantes desfavorecidos se conformen con la distribución inequitativa, y podría complementarse con distintas formas de coerción. Pero en ese caso se requeriría también de agentes dispuestos a cooperar con una distribución equitativa de costos y beneficios, y que se reclutan para interacciones inequitativas en virtud de manipulación y engaño. Estos agentes genuinamente cooperativos, a su vez, no pueden basar su disposición en el mero egoísmo racional, pues así sucumbirían igualmente al aliciente del engaño. Es razonable suponer que su diseño contemple mecanismos distintos al mero egoísmo racional para valorar la cooperación equitativa.

Si nos enfocamos ahora en la primera causa de la existencia de información imperfecta, el ruido en el entorno, tenemos dos maneras de llegar a la misma conclusión: la primera es que la existencia de ruido es aprovechada por algunos agentes para engañar de manera eficiente, con la opción adicional de escudarse en el ruido del ambiente para ocultar su manipulación deliberada de la información. Esta vía termina confundiendo con la anterior. Es decir, los agentes ven el engaño como una estrategia racional egoísta y se aprovechan del ruido en el entorno para confundir a los agentes que quieren explotar. Alternativamente, si asumimos que el ruido causa problemas en la información sin la intervención de actitudes engañosas, entonces planteamos una pregunta adicional. ¿Mediante

qué mecanismos logran los grupos humanos dar solución al problema de la información imperfecta debido a ruido en el entorno? Uno de los mecanismos universales parece ser el diseño de reglas y normas de conducta que garanticen la coordinación de los agentes involucrados. En este caso basta darse cuenta de que una estrategia de este tipo supone una meta-regla que exhorta a los agentes a seguir las reglas acordadas porque conducen a beneficios para el grupo. Pero esa meta-regla se basa en realidad en una previa disposición de cada individuo a comprometerse con el bienestar general. Este compromiso es una preferencia social, porque presupone que el bienestar general es una meta tan válida para el individuo como su propio bienestar. Es decir, aun asumiendo que ningún agente tiene el deseo o propósito de engañar a sus vecinos, necesitan PS para coordinar sus estrategias siguiendo las normas acordadas en el grupo. Si no es posible resolver el problema de la información imperfecta por ruido en el entorno sin tener PS, entonces, también por esta vía concluimos que el egoísmo racional no recomienda la cooperación en un mundo con información imperfecta o privada. Sólo si los agentes previamente tienen PS, entonces la racionalidad recomienda cooperar para satisfacer esas preferencias. Las PS permiten resolver el problema de la información imperfecta en las dos versiones aquí exploradas.

Sin embargo, no sobra señalar que la tendencia a la conducta egoísta o nepotista es un rasgo predecible en organismos diseñados por la selección natural. En ese orden de ideas es también plausible que la estrategias basadas en el engaño sean el origen principal, y quizás el único, de interacciones con información imperfecta o privada. Esas estrategias, como señalamos arriba, pueden llevar a formas de cooperación inequitativa, dónde unos pocos agentes se llevan la porción leonina del producto de la cooperación. En la literatura dominante sobre la evolución de la cooperación, el término «cooperación» se ha reservado para aquellas interacciones en las que la distribución es equitativa (Trivers 1971; Axelrod and Hamilton 1981; Sober & Wilson 1998; Bowles & Gintis 2011). Trivers, por ejemplo, señaló que cuando hay diferencias de poder entre los jugadores la cooperación no evoluciona, aunque él sabía que los primates que viven en grupos jerárquicos cooperan y distribuyen los beneficios inequitativamente. Todos ganan, pero los dominantes ganan más que los subordinados. Pero en la literatura sobre la evolución de la cooperación una distribución inequitativa no se «honra» con el término «cooperación». «Cooperación» inequitativa puede también evolucionar si algunos agentes carecen de PS y tienen la destreza suficiente en el engaño para explotar a los cooperadores y llevarse así la porción leonina del producto cooperativo. Sin embargo, como señalamos arriba, al menos algunos agentes deben ser genuinos cooperadores, en el sentido de preferir una distribución equitativa y ser aversos a la explotación de otros cooperadores.

## IV. CONCLUSIÓN

La combinación de egoísmo racional y engaño es la objeción crucial contra la idea de que cooperar es racional para agentes egoístas sin PS. Hume estuvo muy cerca de esta tesis en su enfrentamiento con el carácter que denominó el «bribón sensato» (sensible Knave). El bribón obra según la máxima: «Aunque la honestidad es una buena política general, está sujeta a excepciones, de modo que el sabio [egoísta, AR] es el que cumple la regla en general pero incumple en todas las excepciones» (Hume 1902, §232). Esta máxima presupone que el egoísmo racional no recomienda cooperar en todos los casos. Hume sostuvo oficialmente que la justicia es un motivo artificial que se deriva del auto-interés dirigido racionalmente al largo plazo. Sin embargo, Hume dijo al enfrentarse a la máxima del bribón: «If his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villainy or baseness, he has indeed lost a considerable motive to virtue» (EPM 9.22). Aunque la frase es ambigua, en el contexto parece conceder que el bribón no carece de racionalidad, sino de un motivo primitivo hacia la justicia, o en la terminología aquí utilizada, carece de PS positivas o altruistas. Al carecer de ellas, ve la deserción como una opción racional en todos aquellos casos que no acarreen consecuencias dañinas para su reputación y para su continuada aceptación en una sociedad de cooperadores. La moralidad genuina requiere, por tanto, de una disposición primitiva a tratar a otros con equidad y a renunciar a las oportunidades de explotar su conducta cooperativa.

## REFERENCIAS BIBLIOGRÁFICAS

- AXELROD, R., HAMILTON W.D. 1981: «The Evolution of Cooperation», *Science* 211, pp. 1390-1396.
- BINMORE, K., SHAKED, A. 2010: «Experimental economics: Where next», *Journal of Economic Behavior and Organization* 73, pp. 87-100.
- BOWLES S., GINTIS H. 2011: *A Cooperative Species. Human reciprocity and its evolution*. Princeton and Oxford: Princeton University Press.
- De WAAL, F. 2006. *Primates and Philosophers*. Princeton and Oxford: Princeton University Press.
- FRANK, R. 1988: *Passions within Reason*, New York: W.W. Norton.
- GAUTHIER, D. 1986: *Morals by Agreement*. Oxford: Oxford University Press.
- GIBSON, M. 1977: «Rationality», *Philosophy and Public Affairs* 6, pp. 193-225.
- HAMPTON, J. 1991. «Two faces of contractarian thought», en P. Vallentyne (ed.) *Contractarianism and Rational Choice*. New York: Cambridge University Press, pp. 31-55.

- HOBBS, T. 1981 [1651]: *Leviathan*, C.B. Macpherson (ed). London: Penguin.
- HUME, D. (1739, 2000) *A Treatise of Human Nature*, David F. Norton (ed.). Oxford: Oxford University Press.
- \_\_\_\_\_, 1902 [1777]: *Enquiries concerning the human understanding and concerning the principles of morals*, L. Selby-Bigge (ed). Oxford, Oxford University Press.
- JOYCE, R. 2006: *The Evolution of Morality*. Cambridge MA and London: MIT Press,
- KITCHER, P. 2011: *The Ethical Project*. Cambridge MA: Harvard University Press.
- SAYRE-McCORD, G. 1991: «Deception and reasons to be moral», en P. Vallentyne (ed.) *Contractarianism and Rational Choice*. New York: Cambridge University Press, pp. 181-95.
- SOBER, E., WILSON, D.S. 1999: *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge MA: Harvard University Press.
- TRIVERS, R. 1971: «The Evolution of Reciprocal Altruism», *Quarterly Review of Biology* 46, pp. 35–57

ALEJANDRO ROSAS es doctor en Filosofía por la Universidad de Münster, Alemania. Profesor titular en el Departamento de Filosofía de la Universidad Nacional de Colombia.

Líneas de investigación:

Es director del Grupo de investigación Ética, comportamiento y evolución, cuyos objetivos de investigación comprenden la psicología y filosofía moral, y la evolución de la mente y la cooperación.

Publicaciones recientes:

«Disentangling social preferences from group selection», *Biological Theory* (2012) 6 (2): 169-175;

«Towards a unified theory of reciprocity». Commentary on F. Guala: Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate, *Behavioral and Brain Sciences* (2012) 35, 36-37.

Correo electrónico: arosasl@unal.edu.co