Luis Enrique Segoviano Contreras,* Mario Alberto Morales Sánchez**

# El principio de interés propio en el análisis y el diseño económico

# The self-interest principle in economic design and analysis

**Abstract** | The aim of this paper is to analyze the explicative and descriptive function of self–interest assumption as one of the most important elements of the behavioral model of rational choice in economics. From experimental evidence in behavioral sciences on prosocial behavior and social preferences, has been developing a critical approach about the way human agency is modeled in economic theory, specifically, about empirical and theoretical scope of behavioral assumptions as self–interest. Setting out that the self–interest assumption has two functions —one representational in modelling and the other normative in economic design—, we argue what aspects in the economic practice required radical changes and what others not, taking into account such experimental evidence. Our main thesis is that there is a difference between the self–interest assumption's validity to represent human behavior and, on the other side, its application in economic design, because the latter entails a vision that legitimizes only incentives and markets mechanisms, which could undermine our understanding about how such instruments crowding out norms and people's social preferences.

**Keywords** | self–interest, behavioral sciences, economic design, incentives, social preferences.

**Resumen** | El objetivo de este trabajo es analizar el carácter descriptivo y explicativo del *principio de interés propio* como uno de los supuestos fundamentales del modelo conductual de elección racional, tal como se emplea en economía. A partir de la evidencia experimental en las ciencias del comportamiento sobre preferencias sociales y prosocialidad, se ha venido desarrollando una visión crítica sobre la forma en que se modela la agencia humana en la teoría económica, en particular sobre el alcance teórico y empírico de supuestos conductuales como el del interés propio. Partiendo de que el *principio de interés propio*



Recibido: 19 de octubre de 2020. Aceptado: 3 de febrero de 2021.

* Doctor en Filosofía de la Ciencia, Facultad de Economía, UNAM.
** Doctor en Economía, Facultad de Economía, UNAM.
**Correos electrónicos:** luis.segoviano@live.com.mx | almoralessanchez@gmail.com

Segoviano Contreras, Luis Enrique, Mario Alberto Morales Sánchez. «El principio de interés propio en el análisis y el diseño económico.» *Interdisciplina* 9, nº 25 (septiembre–diciembre 2021): 185-208.
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


cumple dos funciones —una de carácter representacional y otra normativa en el diseño económico—, presentamos una línea de argumentación para señalar qué aspectos de la práctica económica requieren modificaciones tomando en consideración dicha evidencia experimental. Nuestra aserción principal es que hay una diferencia entre la validez del *principio* para representar el comportamiento y su aplicación en el diseño económico, puesto que esto último conlleva una visión que legitima únicamente instrumentos de intervención basados en incentivos y mecanismos de mercado, lo cual podría socavar nuestra comprensión sobre cómo tales instrumentos llegan a desplazar normas y las preferencias sociales de las personas.
**Palabras clave** | interés propio, ciencias del comportamiento, diseño económico, incentivos, preferencias sociales.

## Introducción

EL OBJETIVO DE ESTE TRABAJO es presentar un análisis del carácter normativo y explicativo del *principio de interés propio* como uno de los supuestos fundamentales del modelo conductual de elección racional. Desde una perspectiva económica, este principio dicta que el objeto último de la acción del individuo es la satisfacción de sus necesidades y su bienestar propio (Kirchgässner 2014; Cropanzano, Goldman, y Folger 2005). A partir de esta caracterización, se ha establecido una manera de entender la motivación y la toma de decisiones ampliamente aceptada para describir y explicar la conducta humana como objeto de estudio de la ciencia económica y otras disciplinas sociales (Kirchgässner 2008). No obstante, con el auge de la vertiente de investigación experimental en las ciencias del comportamiento se ha venido desarrollando una visión crítica sobre la forma en que se ha modelado la agencia humana en la teoría económica (Samson 2014; Thaler 2016; Angner y Loewenstein 2012). Hay una larga lista de resultados experimentales que apuntan a que la concepción de la motivación basada en el interés propio resulta insuficiente para captar motivaciones y preferencias sociales que las personas demuestran en escenarios de laboratorio y campo (Gintis 2000; Bowles y Gintis 2011; Van Dijk 2015). Esto ha llevado a una serie de cuestionamientos y objeciones sobre su alcance metodológico y validez empírica como parte de los supuestos que conforman el modelo conductual de agencia racional (Thaler 2000; Van Dijk 2015).

En este trabajo, se presenta una evaluación crítica del *principio de interés propio* demostrando que su aplicación en el análisis económico cumple dos funciones que, aunque íntimamente conectadas, pueden diferenciarse: una representacional en la modelación y otra normativa en el diseño económico. A partir de esta propuesta, se pretende clarificar una serie de malentendidos sobre lo que los resultados experimentales en ciencias del comportamiento implican para el análisis

económico, y los cambios que requieren realizarse en la práctica económica para alcanzar una convergencia disciplinar apropiada entre ambas vertientes. Nuestra principal aserción es que, más que centrarnos en la discusión teórica de la dimensión representacional de elección racional, el área que requiere mayor atención es la aplicación normativa del *principio de interés propio* en el diseño económico, debido a que los supuestos canónicos de agencia racional conllevan una visión que legitima únicamente instrumentos de intervención basados en incentivos y mecanismos de mercado, lo cual podría estar mermando la forma en que comprendemos cómo las personas responden a normas sociales y cómo se generan otras formas de cooperación social. Desde nuestro punto de vista, la importancia de ampliar y enriquecer los supuestos conductuales, a partir de los cuales se caracteriza el comportamiento humano, no reside en el carácter realista de los modelos de agencia —ya que en ocasiones funcionan muy bien con supuestos altamente idealizados—, sino en el plano del diseño económico e institucional, en el cual hay repercusiones importantes sobre la forma de desarrollar y legitimar ciertas medidas de intervención conductual.

En la primera parte, se comienza con una revisión conceptual de este principio, su carácter representacional como parte de los supuestos conductuales de los modelos de agencia, y se discute su importancia teórica y explicativa para captar rasgos fundamentales del comportamiento humano. Aquí se analiza la pretensión de realismo que comúnmente se critica de la modelización económica y se extraen algunas consecuencias sobre el papel de los supuestos conductuales en tales modelos, con particular atención en teoría de juegos. En la segunda parte, se presentan y discuten algunas implicaciones de la evidencia experimental sobre motivación y conducta prosocial —i.e. acciones en que las personas procuran el bienestar e interés de otros— sobre la forma en que actualmente se comprende la agencia humana en la ciencia económica. En particular, se discute si ello conlleva a una revisión crítica del *principio de interés propio* como supuesto empíricamente válido de la motivación humana en la concepción más canónica de agencia racional. Parte de la discusión que aquí se presenta gira en torno a la tensión de que la interpretación de agencia racional incluya o sea consistente con los resultados experimentales que apuntan a motivaciones y formas de comportamiento en que las personas muestran una consideración genuina por el interés y bienestar de los demás, y qué cambios en la perspectiva convencional requerirían realizarse. En la tercera parte, se presenta nuestra propuesta sobre el alcance normativo con que se aplica el principio analizando una serie de implicaciones para el desarrollo de instrumentos y medidas de intervención dentro del área del diseño económico. Se analizará cómo este principio no solamente entraña una suposición factible para el enfoque disciplinar de la ciencia económica, sino que trata, fundamentalmente, de una perspectiva normativa para jus-

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


tificar y legitimar la aplicación de incentivos para el desarrollo de estrategias de intervención social y organizacional, lo cual podría estar llevando a limitaciones muy importantes sobre lo que entendemos como cambio de comportamiento. En la cuarta parte, se cierran conclusiones con los resultados del análisis presentado.

## La dimensión representacional del principio de interés propio

El *principio de interés propio* es un supuesto conductual que establece que las personas solo actúan para satisfacer sus necesidades y su bienestar individual (Kirchgässner 2008; Cropanzano, Goldman, y Folger 2005). Más en específico, este principio hace equivalente la satisfacción del interés propio con el bienestar material (Kirchgässner 2014).[1] Tal formulación tiene una larga historia que data del *dictum* de Smith (1776/1994) sobre la mano invisible, la concepción del así denominado *homo economicus* de Mill (1848/1951), los tipos ideales de Weber (1913/1997), hasta la caracterización formalizada del agente racional económico de la vertiente neoclásica dominante a mediados del siglo XX (Morgan 2006; Angner y Loewenstein 2012). Aunque la forma en que se han estudiado y comprendido las motivaciones e intereses que caracterizan el comportamiento humano ha variado enormemente a través de las diversas tradiciones del pensamiento económico, se puede rastrear una concepción heredada que ha venido a conformarse bajo el enfoque de elección racional (Kirchgässner 2008).

En la práctica económica moderna, la suposición de interés propio ha pasado a conformar una aserción empírica simple que sirve para modelar el comportamiento del consumidor: los seres humanos nos movemos por aquello que nos genera un bienestar material. En la teoría moderna del consumidor, por ejemplo, el objetivo es conocer cómo deciden los consumidores los bienes que compran dada su renta limitada y los precios de los bienes (Varian 2010). El *principio* se inserta como parte de un problema de maximización de utilidad o satisfacción a través del ordenamiento de cestas de consumo de acuerdo con las preferencias del agente y los niveles de satisfacción que obtiene de cada una. De esta manera, el análisis ulterior en la teoría del consumidor presupone a un agente buscando satisfacer su propio interés material enfrentado a una restricción presupuestaria que determina la cantidad máxima que puede obtener de los bienes que desea. Tal aserción sobre el interés material resulta metodológica y empíricamente vá-

**1** Aunque en su formulación teórica, el *principio de interés propio* no conlleva ninguna implicación específica sobre lo que conforma el interés de un agente —sea adquirir bienes materiales o salvar el mundo de la hambruna—, también es cierto que en la práctica convencional los economistas priorizan hacer equivalente el interés propio con el bienestar material. Aquí seguimos dicha práctica.

lida, y podemos constatarlo desde nuestra experiencia propia. Compramos y adquirimos bienes que nos proporcionan satisfacción y bienestar y, generalmente, nuestra vida cotidiana gira en torno al consumo e intercambio de productos directamente relacionados con nuestro propio interés. Como parte de los supuestos de la toma de decisiones, este principio permite una representación apropiada, en muchas circunstancias, de la motivación y del comportamiento humano. Aunque requerimos asentar un análisis más exacto en cada caso para establecer cómo la búsqueda del bienestar material influye nuestras decisiones y se ve determinada por nuestras preferencias, tal suposición sobre el interés propio permite trazar un aspecto constitutivo fundamental de la naturaleza humana. Esta simple constatación debe ser suficiente para justificar que el *principio de interés propio* resulta un eje metodológico y descriptivo clave para comprender el comportamiento humano como parte del objeto de estudio de la ciencia económica. La objeción a esta simple caracterización no está, por supuesto, en que tal búsqueda del bienestar material no sea parte del comportamiento, lo cual sería complicado rechazar, sino que sea suficiente para comprender todo lo que resulta importante del mismo. Ciertamente, el interés propio interpretado como la búsqueda del bienestar material propio no es en absoluto exhaustivo para comprender el carácter de la motivación humana y, en muchas ocasiones, trazar inferencias a partir de este tipo de supuestos conductuales se suele considerar como idealizaciones e incluso distorsiones sobre el comportamiento humano (Morgan y Knuuttila 2012; Tittenbrun 2013).

Ha sido una crítica recurrente de investigadores en la línea de la economía del comportamiento que un mayor realismo de los supuestos psicológicos es lo más apropiado para ampliar nuestra visión de la toma de decisiones (Jolls, Sunstein y Thaler 1998; Camerer 1999; Thaler 2000; Samson 2014). Usualmente, señalan que simplificaciones sobre información, racionalidad, y motivación, resultan limitantes para comprender *lo que realmente* hacen los seres humanos (Mullainathan y Thaler 2000; Angner y Loewenstein 2012). Esta exigencia requiere ser evaluada de manera más precisa. Si lo que se busca establecer es que nuestra comprensión del comportamiento mejora al ampliar el conjunto de factores que explican la toma de decisiones, tal aserción es correcta. Pero, en contraste, si lo que se pretende derivar es que los modelos de agencia asemejan caricaturas o idealizaciones que requieren ser remplazadas con supuestos más realistas, esto no es necesariamente lo más factible en muchas ocasiones. No es obvio que la pretensión de mayor realismo sea siempre una razón epistémica y metodológica suficiente para modificar los supuestos conductuales y robustecer los modelos de agencia y toma de decisiones empleados en economía. Para ver tal punto, se requiere explorar el papel representacional de tales modelos y de los supuestos que los constituyen.

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


Para demostrar cuál es el papel metodológico y explicativo que juega el *principio de interés propio* —junto con otros supuestos conductuales de la toma de decisiones— vamos a explicar cómo se utiliza en los modelos de comportamiento estratégico de teoría de juegos para representar problemas de cooperación y acción colectiva. Con ello, se pretende establecer que la exigencia de mayor realismo en los modelos de agencia no siempre garantiza que sean más explicativos, sino que su función depende crucialmente de identificar cuáles son los factores críticos que determinan el espectro de alternativas de decisión y los incentivos de los que disponen los agentes en un escenario estratégico, y que brindan un marco de referencia para el estudio de muchas situaciones económicas que enfrentan los seres humanos en el mundo real (Grüne-Yanoff y Schweinzer 2008; Grüne-Yanoff y Lehtinen 2012).

En un *dilema social* se estudia la tensión entre lo que resulta mejor individualmente, pero que, a un nivel colectivo, conduce a todos a un peor resultado (Kollock 1998; Ostrom 1998). Provisión de bienes públicos como sistemas de alumbrado, infraestructura vial, o el manejo y cuidado de los recursos naturales —agua, zonas forestales, aire limpio— son ejemplo de situaciones en las que emergen conflictos de interés entre muchos agentes sobre la adquisición, producción, y administración de los medios de asignación de tales bienes y recursos. Cada ciudadano se ve beneficiado por la provisión y protección de estos bienes; no obstante, cada uno puede estar aun en una mejor posición si los demás contribuyen, pero se evita el costo de producirlo o protegerlo. Debido a que resulta sumamente difícil y costoso excluir a aquellos que no contribuyen, cada uno está tentado a disfrutar del bien público sin asumir el costo de producirlo. Dado que todos se encuentran en la misma situación, el resultado de esto es que nadie contribuye para su protección o provisión y, por ende, quedan en peor situación que de haber cooperado. En estos casos, como en muchos otros similares, lo que parece racional y conveniente desde un punto de vista del individuo —utilizar o consumir tanto como le sea posible— conlleva a una situación socialmente subóptima en la que todos quedan en peor condición que si se hubiesen decidido a cooperar (Kollock 1998; Bowles y Gintis 2011).[2] Veamos ahora cómo esta narrativa se modeliza utilizando la teoría de juegos.

La figura 1 es la representación gráfica del modelo formal de un juego de bienes públicos que emplea una función de producción lineal. El eje horizontal representa el número de jugadores que cooperan ($N > 2$), y el eje vertical el pago $\pi$

**2** Tal es la advertencia —y predicción— que hizo Hardin (1968) sobre la *tragedia de los comunes* y que ha establecido una larga tradición la representación de problemas de bienes públicos y de acción colectiva en la teoría económica.

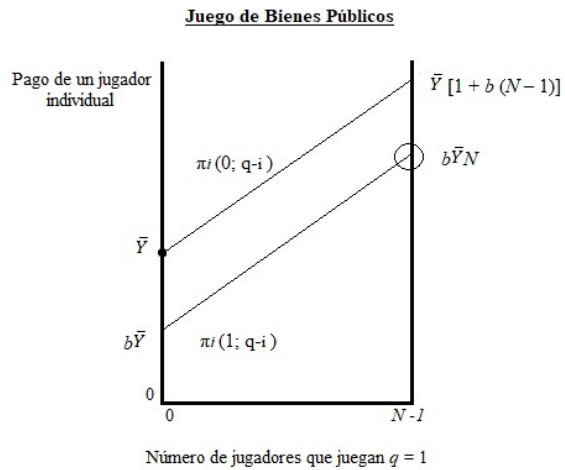
que recibe un jugador individual *i* con respecto al total de contribuciones *q* que dicho agente realiza al bien público. Cada jugador debe elegir $qi = \{0,1\}$ para maximizar

$$\pi i = \bar{Y}\,[b(qi + q{-}i) + 1 - qi]$$

- donde $\bar{Y}$ es un parámetro que representa la sumatoria de la contribución de los jugadores al bien público
- $\bar{Y}\,(b-1)$ es el beneficio neto para *i* de cooperar,
- $q - i = \sum j,\ j \neq i\ qj$ denota, desde la perspectiva de *i*, el número de otros jugadores que cooperan.
- Se asume que $bN > 1 > b > 0$ donde bN representa el beneficio grupal superior de contribución sobre el privado $qi = 1$, pero el beneficio individual privado es mayor que *b*, que representa el valor de una sola contribución al bien público.[3]

Asumiendo que $\pi\,(1;\ q-i\,)$ es el pago que recibe el jugador *i* cuando contribuye al bien público —y $\pi\,(0;\ q-i)$ cuando no contribuye—, se puede explicar por qué no-cooperar es la mejor respuesta estratégica, independientemente de lo que los demás hagan. Cada jugador i está mejor jugando $qi = 0$, puesto que, si todos contribuyen, obtiene el máximo beneficio de disfrutar del bien público sin asumir el costo de producirlo, que está dado por el punto $\bar{Y}\,[1 + b\,(N-1)]$.

**Figura 1.** Juego de bienes públicos.




Fuente: Barrett (2016), traducido con autorización.

**3** Elaborado a partir de Barrett (2016).

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


Ahora bien, si ninguno contribuye, también es mejor no cooperar, ya que su resultado es $\bar{Y}$ en el cual preserva su dotación inicial de 1 y es un mejor resultado que $b\bar{Y}$, lo cual sería lo que obtendría de ser el único en haber contribuido. Dado que se trata de un juego simétrico, se puede demostrar que el equilibrio de Nash es $\bar{Y}$ —donde todos optan por no cooperar ($qi = 0 \forall i$), que corresponde al punto del círculo negro en la figura 1—, pues es el único conjunto de estrategias dominantes para todos los jugadores. Con esto, se puede demostrar lo señalado arriba acerca de los *dilemas sociales*: aunque todos vean que están mejor cooperando, agentes actuando por su interés propio terminarán en una peor situación de lo que podrían haber sacado colectivamente —($qi = 0 \forall i$) que corresponde al círculo blanco de la figura 1—.

El *principio de interés propio* es un supuesto conductual fundamental para representar y explicar el resultado subóptimo de los modelos de juego utilizados para representar dilemas de acción colectiva. En esta función representacional, el *principio de interés propio*, junto con otros supuestos sobre información y comportamiento estratégico, nos permite establecer una serie de factores críticos sobre la situación bajo estudio, la estructura de pagos, y las opciones de decisión de que dispone cada agente en función de lo que harán los demás jugadores, y que se replican en muchos de los problemas que enfrentamos en el mundo real. Nuestro primer punto aquí es que su alcance representacional *no depende* de que todos los problemas de acción colectiva resulten en un estado social subóptimo, es decir, que prevalezca un carácter predictivo del modelo, sino que funciona como un marco de referencia tanto para derivar explicaciones de por qué falla la cooperación humana y también cuando resulta exitosa. Ahora bien, si las personas cooperan y no caen en las trampas señaladas en tales dilemas, ¿deja de funcionar el modelo de juego y, por lo tanto, requieren agregarse supuestos más realistas —i.e. agregar otras variables que permitan representar que las personas no actúan *solo* por su bienestar material? Nuestra respuesta es: no necesariamente.

Siendo el caso de situaciones de interacción social en que las personas logran resolver sus problemas de cooperación, el modelo de juego sigue siendo válido en su carácter representacional porque nos permite estudiar cuáles fueron las barreras estratégicas y los costos materiales que tuvieron que superar quienes enfrentaban el conflicto para alcanzar, justamente, un resultado social o colaborativamente favorable. Que haya casos empíricamente corroborados que no terminan en el resultado social subóptimo como queda predicho en el modelo, no invalida en absoluto la función explicativa del mismo. Sea por la implementación exitosa de nuevas instituciones, cambios en los canales de comunicación, o cualquier otro factor que garantice el establecimiento de normas de la cooperación, el modelo de juego sigue teniendo una aplicación metodológica

crucial para identificar los factores críticos que imponen este tipo de problemas y que vuelven tan inestable la cooperación.[4] Esto no invalida la suposición de interés propio, dado que no se trata de considerar que las personas dejan de buscar su bienestar material, sino que la constatación de otros factores de carácter social u organizacional resulta necesaria para explicar lo que ocurre en un caso de estudio específico y que resulta en una desviación a la predicción formal del modelo de juego. En tal caso, *ampliamos* el campo de aplicación del modelo introduciendo nuevas condiciones para investigar cada nuevo problema bajo estudio. Es un error pretender que los supuestos de agencia racional deban sustituirse simplemente porque los modelos no representan de manera más realista otros aspectos constitutivos de la interacción y el comportamiento humano. Los modelos de agencia y comportamiento estratégico siguen siendo válidos, tanto en lo que motiva a los agentes como en las barreras de interacción que enfrentan, aunque no son exhaustivos. Esta cuestión es parte de la confusión que se tiene con respecto a la función de los supuestos de racionalidad, entre ellos el de interés propio, y su contraste con la extensa acumulación de resultados experimentales sobre comportamiento humano que permiten constatar la existencia de preferencias sociales y motivaciones no–económicas (Henrich *et al*. 2001; Gintis *et al*. 2005; Tyler 2011).

En la mayoría de las ocasiones, los modelos no harán mejor trabajo si simplemente incluimos más y más suposiciones conductuales que parezcan brindarnos una visión más completa de la realidad. Que los modelos de agencia empleados en la ciencia económica presenten la motivación y el comportamiento bajo esta caracterización, no puede tratarse como una limitación simplemente debido a la falta de realismo de sus supuestos. Ciertamente, la percepción y la motivación humana integran muchos otros aspectos que no son tomados en cuenta en la modelación económica, pero los criterios por los cuales esas simplificaciones funcionan al nivel de la explicación y la interpretación no pueden rechazarse meramente por su carácter idealizado o altamente abstracto. Los modelos no tienen pretensión de exhaustividad y funcionan, justamente, porque permiten aislar aspectos clave que están bajo consideración del objeto de estudio (Rodrik 2015; Grüne-Yanoff y Schweinzer 2008). Modelos de juego como el discutido arriba, son robustos, precisamente, porque permiten representar un amplio número de situaciones de interacción social que mantienen una estructura de pagos semejante. Sea que hablemos sobre cómo proveer un alumbrado pú-

**4** Para entender el uso de los modelos formales aplicados a investigaciones experimentales en el estudio de problemas de cooperación, véase Ostrom (2005 y 2010). Algunos trabajos clásicos sobre el estudio de la cooperación en juegos de bienes públicos en laboratorio y de campo se pueden revisar en Fehr y Gätcher (2000), Henrich *et al*. (2004); Ensminger y Henrich (2014).

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


blico, una infraestructura vial, o quizás administrar un sistema de seguridad pública, todos estos problemas pueden ser representados bajo un mismo dominio de modelos de juego dado que captan aspectos estructurales que les subyacen a cada uno, y que nos proporcionan un enfoque general para su estudio.

Si queremos que nuestros modelos funcionen, necesitamos que aíslen de manera precisa los factores que están bajo análisis, y no meramente que nos provean de una imagen completa del comportamiento que termine siendo estéril a nuestros objetivos de investigación. Modelar implica un intento de captar aspectos constitutivos de la realidad, mientras se omiten otros aspectos no esenciales, al menos para el objeto de estudio (Rodrik 2015). Es parte del análisis y la aplicación del modelo lo que permitirá distinguir aquellos casos en los cuales las omisiones resultan o no relevantes. Parte del reto, en este sentido, es establecer qué simplificaciones de los modelos pueden contravenir a una mejor comprensión del objeto de estudio. Y esto es crucial, pero entonces no se debe meramente a la falta de realismo por el cual no funcionan los modelos, sino en establecer en qué condiciones, o para qué tareas de ciencia aplicada, un cambio o adición de supuestos conductuales resulta necesario para mejorar el poder explicativo de nuestras herramientas de análisis. Esta cuestión es la que nos conduce al estudio de la evidencia experimental sobre prosocialidad humana que vamos a tratar a continuación.



## El *principio de interés propio* y motivaciones prosociales

A lo largo de las últimas décadas, se ha ido acumulando una extensa evidencia experimental que demuestra que los seres humanos no siempre toman decisiones que correspondan con la suposición canónica de interés propio derivada del modelo de elección racional, sino que actúan tomando en consideración el bienestar y el interés de otras personas (Camerer y Fehr 2004; Gintis *et al.* 2005; Henrich *et al.* 2004). Hay resultados en estas áreas de investigación que demuestran que las personas están dispuestas a castigar o recompensar a otros, aunque ello implique incurrir en un costo personal (Fehr y Gintis 2007; Van Dijk 2015). Asimismo, las personas cooperan, siguen normas, y buscan resultados equitativos, y todo ello en situaciones de interacción montadas en escenarios experimentales que ponen en juego ganancias materiales para ellos y para otros (Bowles y Gintis 2011). Esta evidencia ha sido interpretada para demostrar la prosocialidad en los seres humanos.[5]

**5** La prosocialidad se define como una serie de rasgos de la motivación y la conducta en la que el individuo asume un costo, sea monetario, en tiempo, esfuerzo, o algún otro recurso, para beneficiar a otros (Schroeder y Graziano 2015; Henrich y Henrich 2007). A partir de resultados experimentales en el estudio del comportamiento humano, se ha comprendido el carácter prosocial como un conjunto de rasgos de la motivación y la conducta de las perso-

Esta evidencia sobre la existencia de *rasgos prosociales* ha sido utilizada para cuestionar la validez y alcance de los supuestos conductuales del enfoque de agencia racional y, en particular, del supuesto de interés propio. Si los seres humanos no somos tan egoístas y mostramos una preocupación por el interés de los demás, al parecer dicho principio, junto con otros supuestos conductuales, requiere una revisión seria sobre el papel y función que tiene dentro de la teoría económica. Cómo llevar a cabo dicha revisión no es una tarea concluida. Hay quienes abogan por un remplazo de la concepción de agencia racional (Van Lange *et al*. 2007; Tittenbrun 2013; Gowdy y Polimeni 2005). Aunque quizás se había considerado que tal evidencia acumulada en ciencias del comportamiento mostraba las inconsistencias más severas sobre la naturaleza y la aplicación de los supuestos de agencia racional, los resultados experimentales se han mostrado, en realidad, como un reto sobre la forma en que entendemos la modelación y la función de los modelos en la ciencia económica. Asimismo, como se ha señalado anteriormente, la mera pretensión de realismo ya no se considera condición suficiente para sustituir el supuesto de interés propio dado que la evidencia experimental no invalida el carácter representacional y explicativo de la visión de agencia racional.

Parte del problema consiste en determinar cómo esta evidencia experimental sobre prosocialidad contrasta con la visión de agencia racional. La constatación experimental sobre los *rasgos prosociales* de la motivación y el comportamiento humano amplía nuestra comprensión empírica de la naturaleza humana, pero esto no demuestra, en sentido estricto, que el supuesto de interés propio sea falso. No es fácil rechazar cuáles aspectos de la conducta más centrados en el interés y bienestar propio siempre juegan un papel importante en la mayoría de las decisiones de nuestra vida, incluso en el plano social. Tal pretensión sería errónea. La evidencia experimental amplía y extiende nuestra comprensión sobre cuándo y en qué circunstancias, las personas están dispuestas a sacrificar parte su interés material por el bienestar de otros, así como respetar normas y realizar acciones colaborativas que el enfoque de elección racional no nos permite predecir. Esto, ciertamente, amplía nuestra comprensión de factores explicativos sobre lo que los seres humanos hacemos u omitimos en espacios de interacción y competencia económica. Con ello, se extiende empíricamente nuestra comprensión inicial sobre la búsqueda del interés y el alcance de los supuestos conductuales del análisis económico, pero no se sustituye.

nas orientadas al bienestar de los demás, a pesar de que ello no conlleve, en muchas ocasiones, un beneficio material y represente un costo neto hacerlo (Van Dijk 2015). Hay un largo recorrido en esta literatura que atraviesa una intersección muy importante con la comprensión evolutiva del comportamiento moral (Henrich y Henrich 2006; Viciana 2014).

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


Para asentar el punto anterior, exploremos brevemente uno de los descubrimientos más robustos cuando se han empleado modelos de juego de bienes públicos, como el presentado en la sección 1, para realizar experimentos (Fehr y Gätcher 2000; Fehr y Gintis 2007): la ejecución de un castigo costoso. Desde el trabajo de Fehr y Gätcher (2000), que se considera uno de los pioneros en esta área, y que ha sido replicado en diversas ocasiones, uno de los resultados más consistentes es que los jugadores están dispuestos a castigar a otros cuando se les da la oportunidad, a pesar de que tengan que asumir un costo material al hacerlo. Aun cuando imponer una sanción sobre los no–cooperadores implica asumir un costo personal, un número considerable de participantes están dispuestos a sancionar a aquellos que hacen bajas contribuciones a un bien público. La ejecución de un *castigo costoso* es un comportamiento también identificado en otros reportes experimentales utilizando otros juegos como el juego del ultimátum (Bowles y Gintis 2011). Esto contrasta, por supuesto, con el análisis económico estándar en teoría de juegos, dado que se representa al agente racional que no realizará una acción a un alto costo personal que no conlleve a una ganancia material directa. Además, aplicar un castigo costoso para sancionar la transgresión a una norma representa un dilema de segundo orden: quienes castigan asumen un costo personal, pero el resto puede disfrutar del beneficio de sancionar al transgresor y evadiendo el costo de hacerlo. Si vemos el modelo de juego de bienes públicos, la oportunidad de castigar no debe cambiar el interés de los agentes si todos actúan solo por su interés propio ya que ninguno realizará una acción que represente una pérdida neta de su ganancia material. Si todos los participantes intentaran maximizar su beneficio individual, y, paralelamente, dejan que sean otros los que apliquen el castigo costoso, no habría diferencias significativas entre el modelo de juego y los resultados experimentales reportados, puesto que racionalmente ninguno tendría un aliciente para ejecutar un castigo que le lleve a una pérdida neta de su dotación inicial. Pero la evidencia demuestra que esta predicción es falsa. Los participantes no solo se preocupan de la ganancia material que pueden obtener, sino que también muestran consideración sobre el comportamiento equitativo o justo que otros jugadores lleven a cabo. En este sentido, la aplicación de un castigo representa uno de los descubrimientos más relevantes de las investigaciones experimentales y permite evaluar cómo los participantes responden al comportamiento oportunista de otros y están dispuestos a sacrificar parte de su ganancia con el objetivo de reducir la ganancia que otros buscan obtener, incluso en aquellos casos en que la interacción sea de un solo encuentro (Fehr y Gintis 2007).

¿Qué podemos concluir de esta discrepancia, y de otras similares constatadas en las investigaciones experimentales, entre la predicción del análisis formal del juego y el comportamiento efectivo de las personas en juegos económicos?

La evidencia experimental recabada sobre prosocialidad ha conducido a una nueva comprensión de la naturaleza y complejidad del comportamiento humano sintetizada a partir de resultados específicos sobre el papel de las normas, el deber cívico, la confianza, entre otros rasgos y actitudes morales, que resultan en un conjunto de factores clave para explicar cómo las personas orientan sus expectativas y decisiones de convivencia social e intercambio económico (Aguiar, Gaitán, y Viciana 2020). En la parte de estudio de las preferencias sociales y la cooperación, se ha permitido asentar evidencia para constatar empíricamente ese conjunto de rasgos de la psicología moral y social en los seres humanos que regulan nuestros motivos de conducta y, en consecuencia, las decisiones que llegamos a tomar cuando entran en conflicto beneficios materiales para nosotros y el bienestar de otras personas. Esto ha sido de la mayor relevancia para la forma en que se están estudiando muchos problemas de carácter económico y social hoy en día (Bicchieri 2017; Atkins, Wilson y Hayes 2019).

Como hemos señalado previamente, la evidencia empírica amplía nuestra visión de la toma de decisiones, pero no sustituye los supuestos de agencia racional como el de interés propio. Podemos establecer que las desviaciones con respecto a la obtención del beneficio material amplían nuestra comprensión sobre otros factores conductuales que pudieran jugar un papel en situaciones de mercado reales, como la equidad o la confianza, y que, claramente, no quedan especificados en la representación formal. Pero los modelos no son exhaustivos, y justo en esto viene nuestro reto sobre cómo deben ser integrados todos estos descubrimientos experimentales en el estudio de la cooperación humana. Quizás una de las principales advertencias que se siguen de esto es no caer en la falsa generalización de principios conductuales, como el *principio de interés propio*, es decir, que de manera ingenua se asuma que siempre la ganancia material es el factor principal de motivación humana, pero tal cuestión no radica en una discusión teórica, sino que lleva a una revisión del ejercicio y la práctica de ciencia aplicada que los economistas y los científicos sociales hacen a partir de sus modelos.

Una parte muy importante de la discusión sobre agencia racional y toma de decisiones se ha centrado en una discusión teórica sobre el alcance de los modelos, su idealización, y generalizaciones en la formulación axiomática del análisis económico estándar y su contraste con la visión empírica emergente de las ciencias del comportamiento (Chetty 2015). Lo que hemos señalado hasta este punto, es que tal discusión está mal encaminada. En nuestra opinión, el terreno en el cual buscar esta síntesis y complementariedad radica en el diseño institucional y en la parte de la intervención. No es problema teórico sobre modelos y cuáles tienen supuestos más realistas, sino en cómo todo ello impacta al momento de establecer un enfoque de diseño para intervenir la conducta de las personas y

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


que se logren alcanzar determinados objetivos sociales y organizacionales. Desde nuestro punto de vista, el área de investigación prioritaria para lograr esta síntesis y convergencia está en el diseño económico. Más que revolverse en el terreno teórico y conceptual, las repercusiones, tanto positivas como negativas, del enfoque de agencia que utilicemos las afrontamos más directamente cuando se trata de proveer los medios para influenciar las preferencias y actitudes de otras personas, es decir, buscando repercutir su comportamiento, y es aquí donde se ponen a prueba tales supuestos conductuales. Esto es lo que vamos a defender a continuación.

## La dimensión normativa del principio en el área del diseño económico

En esta última parte, vamos a estudiar la función normativa del *principio de interés propio* en el diseño económico para asentar algunas de sus implicaciones para validar el desarrollo y aplicación de ciertos instrumentos de cambio de comportamiento. A diferencia de lo que se ha señalado anteriormente, con respecto a su función representacional, aquí hay una serie de cuestiones que resultan de la mayor urgencia tratar sobre el alcance y las consecuencias de los supuestos conductuales para dar forma y legitimidad a la perspectiva canónica basada en incentivos y otros medios de motivación que hacen caso omiso de factores conductuales como normas y preferencias sociales.

### *La aserción principal radica en el siguiente argumento*

A partir del *principio de interés propio* se ha ido desarrollando en el análisis económico una visión de agencia humana que representa a los seres humanos como *bribones y oportunistas* de la norma (Ferraro, Pfeffer, y Sutton 2005; Bowles 2008 y 2016). Dado que el análisis económico ha partido de supuestos conductuales que giran en torno a una visión de agente actuando por su interés material, tal visión se ha tomado como eje metodológico y normativo para justificar la implementación de incentivos materiales y de mecanismos de mercado para modificar la conducta humana, y ello podría estar incurriendo en efectos colaterales que socavan, tanto motivaciones no-económicas, como el desarrollo de otros medios de intervención basados en normas sociales y de deber cívico. Si esto es así, la dimensión normativa del diseño requiere una reformulación radical en los supuestos de agencia que se emplean dados los resultados experimentales sobre los efectos perniciosos que llegan a causar los incentivos y otras herramientas de carácter económico.

Desde el enfoque del diseño económico, los supuestos conductuales que conforman la concepción de agencia racional adquieren la función de legitimar

una serie de condiciones, no solo de analizar y comprender, sino principalmente una forma de *intervenir* los problemas de cooperación social y organizacional. Aquí, reconstruimos tal proceso de la siguiente forma. Primero, se parte de una visión de agencia y de toma de decisiones estratégica en la que, como premisa *de facto*, se representa a las personas atrapadas en sus propios dilemas de cooperación y acción colectiva (Ostrom 2005). Segundo, derivado de lo anterior, se establecen supuestos *ex ante* sobre la motivación específica de las personas y la forma en que van a responden a cambios en los costos y beneficios materiales asociados con su repertorio de acciones posibles. Tercero, se elabora e implementa una solución centrada en el problema de la *deserción racional* dado que, si los incentivos y otros instrumentos de intervención conductual hacen su trabajo, entonces tenemos una visión específica de lo que significa que las cosas vayan bien, su impacto y el alcance de los cambios propugnados. Pasemos a comentar en secuencia cada uno de estos puntos para sustentar la aserción señalada anteriormente.

Hemos mostrado que, a partir del supuesto de interés propio, se puede derivar una explicación del resultado subóptimo al que se llega en un dilema de acción colectiva en el cual la mutua deserción es el Equilibrio de Nash. Siguiendo la suposición de interés propio, y de que cada uno busca maximizar su beneficio, se puede concluir que la no-contribución es el conjunto de las estrategias dominantes y racionales para todos los jugadores. Tal interpretación es robusta con respecto a la forma de representar y analizar la estructura de pagos y los factores críticos que los seres humanos enfrentamos para un amplio número de situaciones que puedan ser estudiadas a partir de este modelo. Y como se ha señalado, una de sus virtudes en el plano representacional de la modelación en teoría de juegos. Un giro de la función representacional del modelo al utilizarlo como instrumento para la intervención se encuentra en Olson (1965):

> Indeed, unless the number of individuals in a group is quite small, or unless there is coercion or some other special device to make individuals act in their common interest, rational, *self-interested individuals will not act to achieve their common or group interests*. (Olson 1965, 2)[6]

Esta aserción ha sido denominada la tesis de la contribución cero (TCC) (Ostrom 2000). Desde este punto de vista, cualquier política de diseño institucional u organizacional debe ser elaborada bajo el supuesto de que los ciudadanos

**6** "En realidad, a menos que el número de individuos en un grupo sea bastante pequeño, o que haya coerción o algún otro dispositivo que haga que los individuos actúen en por interés común, *individuos racionales, actuando por interés propio, no actuarán para lograr sus intereses en común o de grupo*". (Traducción propia, en cursiva original).

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


como agentes racionales no harán ninguna contribución voluntaria en miras al bienestar público u organizacional. Bajo la suposición de que actuarán por su interés propio, la TCC plantea que, sin la posibilidad de que se establezca un mecanismo de control y supervisión que garantice la participación de todos, los ciudadanos optarán racionalmente por no contribuir pese a que todos estarían mejor si lo hicieran. Olson (1965) está haciendo equivalente aquí actuar por interés propio y actuar racionalmente, en el sentido de que la motivación e interés de cada individuo excluye la posibilidad de estar dispuestos a hacer algo por el interés o preferencias de otros. Actuar por su interés propio implica que no existe ninguna consideración por el bienestar de los demás —i.e. los intereses que haya en común— de manera que, en aquellos casos donde se requiera una contribución que implique costos no asociados con un beneficio material propio, se asume que solo mediante mecanismos de motivación extrínseca se podrá lograr el cambio conductual deseado.

De acuerdo con el enfoque de elección racional, dado que los ciudadanos son vistos como agentes racionales que no tienen una preocupación genuina ni un interés moral por los otros, y que en ningún caso estarán dispuestos a realizar algo que vaya en detrimento de su interés propio, la aplicación de medidas e instrumentos de intervención mediante incentivos materiales resulta el enfoque de diseño apropiado para lograr el cambio conductual. Esto constituye la visión económica convencional para el diseño institucional. Si lo que se busca es aumentar la productividad del empleado, el administrador se vale de una extensa gama de incentivos en forma de bonos, pagos, compensaciones, que le garanticen inducir un cambio en el desempeño esperado. Si lo que nos interesa es que los ciudadanos respeten las señales de tránsito, un aumento en las multas y penalizaciones facilitará disuadir las transgresiones a la norma. Sea en el ámbito organizacional, en el gobierno, o en las universidades, la visión dominante para el desarrollo de estrategias cuyo objetivo es intervenir e influir sobre el comportamiento humano se ha dirigido a introducir, cambiar, o reforzar los incentivos, y es parte de lo que en la actualidad se ha convertido en la manera pensar y tratar los problemas que surgen en estos dominios de convivencia e interacción ciudadana. Se trata del paradigma de los incentivos para ejecutar y llevar a cabo la resolución de problemas en el área de la ciencia económica aplicada.

No se trata solamente de una forma de representar la agencia racional, a diferencia de lo señalado anteriormente, sino del desarrollo de un enfoque conductual en el área de diseño institucional que ha servido para legitimar ciertas medidas y mecanismos de intervención sobre los medios más eficaces para modificar las preferencias e intereses de las personas en el plano social que dejan de lado completamente la participación prosocial y la responsabilidad cívica (Bowles 2016). Bajo esta perspectiva, el análisis de los problemas de cooperación

social y organizacional se ha centrado en resolver el problema del oportunismo (*free riding*): evitar que alguien disfrute de los beneficios de la cooperación sin haber asumido los costos de producirlos. Esto es lo que constituye la segunda fase del proceso de diseño económico que hemos planteado. Las fallas en la cooperación se interpretan como problemas de oportunismo racional en las cuales se representa a las personas como potenciales transgresores de la norma que requieren los apropiados incentivos que sirvan para disuadirlos y mantengan conformidad a la misma. Se parte de que la deserción racional —i.e. aprovechar la oportunidad de mejorar transgrediendo la norma— proporciona la imagen apropiada para entender, no solamente qué se requiere corregir o vigilar para lograr un resultado eficiente, sino cómo son los seres humanos si tienen la oportunidad de sacar ventaja propia en una situación dada, y por qué la función del diseño es eliminar los incentivos de esos potenciales bribones de la norma.

Si las medidas de intervención están diseñadas para el problema del oportunismo racional —lo cual, como hemos señalado está en la base de la modelación estratégica de la teoría de juegos— quizás estemos distorsionando cuál es exactamente la cuestión vinculada con el supuesto de que las personas buscan solamente su bienestar material. Hay reportes que demuestran que los incentivos materiales tienen un efecto contraproducente sobre motivaciones altruistas y prosociales que las personas exhibían para la realización de una tarea o en la participación de una actividad social (Gneezy y Rustichini 2000b; Fehr y Falk 2002). Se ha constatado, por ejemplo, que la aplicación de incentivos a través de multas, sanciones, y también en determinadas formas de retribuciones y compensaciones, terminan socavando la motivación no-económica previamente mostrada por las personas, de manera que las transgresiones a una norma se vuelven más frecuentes (Gneezy y Rustichini 2000a; Bowles y Polanía–Reyes 2012). A este fenómeno se le ha denominado *desplazamiento de incentivos*. El *desplazamiento* se ha identificado a partir de una serie de condiciones en que intervenciones basadas en mecanismos de motivación extrínseca —sea a través de sanciones, multas, o incentivos materiales— llegan a socavar la motivación inicial que las personas tenían para realizar una acción o una determinada tarea y que, en consecuencia, repercute en el desempeño que habían mostrado previamente (Bowles y Polanía-Reyes 2012; Besley y Ghatak 2018). Estos resultados experimentales apuntan a las limitaciones que tiene el paradigma de ciencia aplicada de la teoría económica al estar centrado en el desarrollo e implementación de mecanismos de motivación extrínseca mediante incentivos y soluciones de mercado.

A partir de esta evidencia experimental, se puede constatar que hay factores conductuales y estratégicos que tienen un papel crucial en procesos de interacción e intercambio que terminan siendo omitidos bajo esta estrecha visión y práctica de diseño económico. Hoy sabemos que la autoconfianza, la aversión

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


a la inequidad, el altruismo, la reciprocidad, entre muchos otros, son factores conductuales que determinan cómo las personas responden a incentivos (Frey y Jegen 2001; Bowles y Polanía–Reyes 2012). Hemos señalado que los modelos de agencia facilitan una comprensión de las condiciones estratégicas y los procesos de toma de decisión de una situación real al *simplificar* los factores críticos que requieren ser representados sobre la base de un problema de estudio. Pero esta simplificación puede resultar equívoca una vez que pasamos al área de diseño económico. Aquí hay una transición importante de la forma en que utilizamos los modelos para representar el comportamiento y la toma de decisiones y pasar a emplearlos como herramientas para llevar a cabo intervenciones conductuales. Y tal transición implica muchos riesgos y vicisitudes de las que comúnmente se ha hecho caso omiso. No se trata de un solo paso avanzar del carácter representacional hacia el plano de la ciencia aplicada. Se trata de dos dominios completamente diferentes. Aquí nos interesa el desarrollo y aplicación de instrumentos para modificar la conducta de las personas, no solo que nos permitan representar aspectos específicos del comportamiento humano. Las simplificaciones realizadas en los modelos no pueden ser equiparadas de la misma forma para comprender las dimensiones prácticas del diseño y sobre la manera en que se emplean para justificar la aplicación de un mecanismo de mercado como aparentemente funciona al nivel teórico. La función representacional de los supuestos de agencia, y de los modelos en los cuales los empleamos, no es equivalente a la dimensión normativa que tienen en la parte del diseño, lo cual requiere considerar muchos otros aspectos que usualmente se dan por descontado. Ciertamente, dentro del modelo podemos considerar que cambios en los mecanismos de precios, i.e., costos y beneficios asociados con determinadas opciones de acción, tendrán un efecto sobre la conducta de las personas, pero tal implicación teórica no aplica de la misma forma cuando se desarrolla una medida en el mundo real en el cual no tenemos garantía de que solo llevará al cambio conductual esperado. La aplicación de multas o sanciones, por ejemplo, imponen costos materiales para disuadir a potenciales infractores de la norma. Bajo una visión canónica del diseño económico, la respuesta conductual de las personas se reduce al balance costo-beneficio de su interés material que está en juego, pero esto deja de lado los efectos colaterales que puede tener en motivaciones no-económicas y en las normas sociales que enmarcan la orientación previa que las personas tienen *ex ante* de la intervención. Aunque los reportes experimentales sobre *desplazamiento* son incipientes, nos permite tener buenas razones para establecer que tales omisiones en el diseño económico pueden conllevar consecuencias perniciosas a mediano y largo plazo sobre las formas de convivencia e interacción que emergen en una sociedad (Sandel 2012; Bowles 2016). El análisis convencional de esta situación sería que,

si los incentivos funcionan, es porque las personas responden de manera adecuada, es decir, racionalmente, a los costos materiales de transgredir la norma. Si los incentivos fallan, tal resultado es porque no se encontró el monto apropiado para lograr el cambio conductual. En esta visión simplista de la naturaleza humana, cambios que no pueden ser cuantificados en términos de unidades de pérdidas o beneficios materiales son tratados como inexistentes, y esto representa un serio problema para la forma en que comprendemos, y pretendemos alcanzar, el bienestar social.

Quizás el mayor reto que hoy tenemos abierto en esta área es cómo la evidencia experimental sobre prosocialidad y sobre las complejas raíces de la motivación y del comportamiento humano nos conducen a desarrollar mejores herramientas e instrumentos de intervención, mejores maneras de cambiar la conducta de las personas en miras al bienestar social. Más que revolverse y atorarse en los debates teóricos sobre modelos, o tipos de explicaciones, la clave está en la forma en que se puede llevar a cabo una revolución sobre lo que significa el cambio de comportamiento integrando al paquete del encargado de diseño nuevas herramientas y concepciones de la agencia humana y la toma de decisiones que están surgiendo en las últimas décadas (Thaler y Sunstein 2008; Bicchieri 2017; Feldman 2018; Aguiar, Gaitán, y Viciana 2020). Muy seguramente, tales cambios traerían aparejado una forma distinta de entrenar a los practicantes y también implicaciones de carácter pedagógico y educativas que permitieran ampliar el panorama para un campo disciplinar unificado en las ciencias del comportamiento incluyendo, por supuesto, la economía.

## Conclusiones

Con el advenimiento de las ciencias del comportamiento, ha resultado crucial abordar cómo los resultados experimentales sobre toma de decisiones, percepción, y motivación humana han marcado un viraje sobre la manera en que se comprende la modelación y explicación de la agencia y la toma de decisiones en la ciencia económica. En este trabajo, hemos realizado una propuesta para comprender en qué partes de la práctica económica tales resultados no modifican de manera sustancial la teoría, y en qué otros se requieren llevar a cabo transformaciones que podrían tener un impacto de mayor alcance. Tomando como eje de estudio uno de los supuestos conductuales más discutidos, *el principio de interés propio*, hemos desarrollado un análisis sobre las diversas funciones que cumplen este tipo de supuestos, a saber, representacional en la modelación y normativa en el diseño, que sirva para clarificar algunos malentendidos y repercusiones que se han venido gestando en miras a alcanzar una visión emergente de síntesis y convergencia en el estudio del comportamiento humano.

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


Aquí hemos explorado una serie de implicaciones teóricas y metodológicas del *principio de interés propio* como un supuesto conductual fundamental de los modelos de agencia a partir del que se formulan hipótesis, explicaciones, y predicciones sobre el comportamiento humano. Como parte de cierre de este análisis, vamos a apuntar una serie de aserciones que sintetizan la línea de argumentación aquí presentada:

1. El *principio* cumple una función representacional y explicativa en la medida en que, a través de la modelación, se pueden trazar inferencias sobre la motivación y el comportamiento que pueden ser contrastados empíricamente. Aspectos concernientes al bienestar material resultan factores críticos para analizar problemas de cooperación humana que adquieren un tratamiento objetivo y replicable mediante los modelos de teoría de juegos. Asimismo, el supuesto de interés propio, junto con otros supuestos conductuales, sirven para establecer un marco de referencia en la investigación experimental que se emplea para identificar el papel y alcance de motivaciones no–económicas que amplían nuestra comprensión del comportamiento humano.
2. Ha sido comúnmente aceptado que la evidencia sobre prosocialidad pone en jaque la visión canónica de agencia racional, y en particular supuestos conductuales como el del interés propio. Se ha argumentado que tal aseveración es equivocada. Aunque, ciertamente, se requiere un trabajo ulterior de evaluación y complementariedad para determinar cuál es exactamente el impacto de esos resultados experimentales, un análisis más conspicuo nos permite esclarecer que las pretensiones de sustitución o remplazo están infundadas. La evidencia experimental se integra al análisis económico ampliando el repertorio de supuestos conductuales al interpretar y aplicar los modelos de agencia en el área de diseño y cambio de comportamiento.
3. La principal crítica aquí planteada a la perspectiva de agencia racional, y los supuestos conductuales que la conforman, se ha establecido trazando una diferencia crucial entre representar e *intervenir*: el proceso de diseño económico convencional simplifica enormemente los factores conductuales que explican cómo las personas responden a incentivos y otros mecanismos de cambio de conductual. La evidencia experimental, en particular sobre el fenómeno de *desplazamiento de incentivos*, ha mostrado una serie de repercusiones negativas que inducen los incentivos materiales con respecto a cambios de comportamiento esperado. Nuestra advertencia principal en relación con esta función normativa del *principio de interés propio* es que se ha ido desarrollando una visión

estrecha para legitimar ciertos medios de intervención basados en incentivos que requiere ser ampliada a partir de la evidencia experimental sobre preferencias sociales y motivaciones no-económicas recopilada en las ciencias del comportamiento. Tal reconsideración en el plano del diseño bien puede llegar a ser la nueva síntesis en la ciencia económica moderna.

## Referencias

Aguiar, Fernando, Antonio Gaitán, y Hugo Viciana. 2020. *Introducción a la ética experimental*. Madrid: Cátedra.

Angner, Eric. y George Loewenstein. 2012. Behavioral economics. En Uskali Mäki (ed.), *Handbook of the Philosophy of Science: Philosophy of Economics*. Amsterdam: Elsevier: 641–690.

Atkins, Paul, David S. Wilson y Steven Hayes. 2019. *Prosocial: using evolutionary science to build productive, equitable, and collaborative groups*. New Harbinger Publications.

Barrett, Scott. 2016. Coordination *vs.* voluntarism and enforcement in sustaining international environmental cooperation. *Proceedings of the National Academy of Sciences*, 113(51): 14515-14522.

Besley, Timothy, y Maitreesh Ghatak. 2018. Prosocial motivation and incentives. *Annual Review of Economics*, 10: 411-438.

Bicchieri, Cristina. 2017. *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bowles, Samuel. 2008. Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments. *Science*, 320 (5883): 1605-1609.

Bowles, Samuel. 2016. *The moral economy. Why good incentives are no substitute for good citizens*. New Haven, Londres: Yale UP.

Bowles, Samuel, y Sandra Polanía-Reyes. 2012. Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50 (2): 368-425.

Bowles, Samuel y Herbert Gintis. 2011. *A cooperative species: Human reciprocity and its evolution*. Princeton University Press.

Camerer, Colin. 1999. Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences*, 96(19): 10575-10577, 1999.

Camerer, Colin y Ernst Fehr. 2004. Measuring social norms and preferences using experimental games: a guide for social scientists. En Joseph Henrich *et al.* (eds.), *Foundations of human sociality*. Oxford: Oxford UP.

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973
Luis Enrique Segoviano Contreras, Mario Alberto Morales Sánchez

Chetty, Raj. 2015. Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5): 1-33.

Cropanzano, Russell, Barry Goldman y Barry Folger. 2005. Self-interest: Defining and understanding a human motive. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 26(8): 985-991.

Ferraro, Fabrizio, Jeffrey Pfeffer y Robert Sutton. 2005. Economics language and assumptions: How theories can become self-fulfilling. *Academy of Management Review*, 30(1): 8-24.

Ensminger, Jean y Joseph Henrich (eds). 2014. *Experimenting with social norms: Fairness and punishment in cross-cultural perspective*. Russell Sage Foundation.

Fehr, Ernst y Simon Gätcher. 2000. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4).

Fehr, Ernst y Armin Falk. 2002. Psychological foundations of incentives. *European Economic Review*. 46(4): 687-724.

Fehr, Ernst y Herbert Gintis. 2007. Human motivation and social cooperation: Experimental and analytical foundations. *Annu. Rev. Sociol*. 33: 43-64.

Feldman, Yuval. 2018. *The law of good people: Challenging states' ability to regulate human behavior*. Cambridge University Press.

Frey, Bruno y Reto Jegen. 2001. Motivation crowding theory. *Journal of Economic Surveys*, 15 (5): 589-611.

Gintis, Herbert. 2000. Beyond *homo economicus*: evidence from experimental economics. *Ecological economics*, 35(3): 311-322.

Gintis, Herbert., Samuel Bowles, Robert Boyd y Ernst Fehr (eds.). 2005. Moral sentiments and material interests. *The foundations of cooperation in economic life*. Cambridge MA: MIT Press.

Gneezy, Ury y Aldo Rustichini. 2000a. A Fine is a price. *Journal Legal Studies* 29 (1): 1-17.

Gneezy, Ury y Aldo Rustichini. 2000b. Pay enough or don't pay at all. *Quarterly Journal of Economics*. 791-810.

Grüne-Yanoff, Till y Aki Lehtinen. 2012. Philosophy of game theory. En Uskali Mäki (ed.), *Handbook of the Philosophy of Economics*. Oxford: 531-576.

Grüne-Yanoff, Till y Paul Schweinzer. 2008. The roles of stories in applying game theory. *Journal of Economic Methodology*, 15(2): 131-146.

Gowdy, John y Raluca Polimeni. 2005. The death of *homo economicus*: is there life after welfare economics? *International Journal of Social Economics*, 32 (11): 924-938.

Hardin, Garrett. 1968. The tragedy of commons. *Science*, 162 (3859).

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert

Gintis y Richard McElreath. 2001. In search of *homo economicus*: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2): 73-78.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr y Herbert Gintis. *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford UP.

Henrich, Joseph y Natalie Henrich. 2006. Culture, evolution and the puzzle of human cooperation. *Cognitive systems research*, 7(2-3): 220-245.

Henrich, Joseph, y Natalie Henrich. 2007. *Why humans cooperate. A cultural and evolutionary explanation*. Oxford: Oxford UP.

Jolls, Christine, Cass Sunstein y Richard Thaler. 1998. A behavioral approach to law and economics. *Stanford Law Review*, 1471-1550.

Kirchgässner, Gebhard. 2008. *Homo Oeconomics. The economic model of behavior and its applications in economics and other social sciences*. Springer.

Kirchgässner, Gebhard. 2014. On self-interest and greed. *Journal of Business Economics*, 84(9): 1191-1209.

Kollock, Peter. 1998. Social dilemmas: the anatomy of cooperation» *Annu. Rev. Sociol*. 24.

Mill, John Stuart. 1951 [1848]. *Principios de economía política*. Cd. de México: Fondo de Cultura Económica.

Morgan, Mary. 2006. Economic man as model man: ideal types, idealization and caricatures. *Journal of the History of Economic Thought*, 28(1): 1-27.

Morgan, Mary y Tarja Knuuttila. 2012. Models and modelling in economics. En Uskali Mäki (ed.), *Handbook of the Philosophy of Economics*. Amsterdam: Elsevier: 49-87.

Mullainathan, Sandhil, Richard Thaler. 2000. Behavioral economics (No. w7948). *National Bureau of Economic Research*.

Olson, Mancour. 1965. *The logic of collective action. Public goods and the theory of groups*. Harvard University Press.

Ostrom, Elinor. 1998. A behavioral approach to the rational choice theory of collective action: Presidential address. American Political Science Association, 1997». *American Political Science Review*, 92(1): 1-22.

Ostrom, Elinor. 2000. Crowding out citizenship. *Scandinavian Political Studies*, 23 (1): 3-16.

Ostrom, Elinor. 2005. *Understanding institutional diversity*. Princeton University Press.

Ostrom, Elinor. 2010. Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change*, 20(4): 550-557.

Rodrik, Dani. 2015. *Economics rules: The rights and wrongs of the dismal science*. WW Norton & Company.

Volumen 9, número 25, (185-208), septiembre–diciembre 2021
DOI: https://doi.org/10.22201/ceiich.24485705e.2021.25.79973


Samson, Alain. 2014. An introduction to behavioral economics. En Alain Samson (ed.), *Behavioral economics guide*, con prólogo de George Loewenstein y Rory Sutherland.

Sandel, Michael. 2012. *What money can't buy: the moral limits of markets*. Macmillan.

Schroeder David y William Graziano. 2015. The field of prosocial behavior: and introduction and overview. En David Schroeder y William Graziano (eds.), *The Oxford handbook of prosocial behavior*. NY: Oxford UP.

Smith, Adam. 1994 [1776]. *La riqueza de las naciones*. Madrid: Alianza Editorial.

Tyler, Tom. 2011. *Why people cooperate. The role of social motivations*. Princeton University Press.

Thaler, Richard. 2000. From *homo economicus* to *homo sapiens*. *Journal of economic perspectives*, 14(1):133-141.

Thaler, Richard. 2016. Behavioral economics: Past, present, and future. *American Economic Review*, 106(7): 1577-1600.

Thaler, Richard y Cass Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tittenbrun, Jacek. 2013. The death of the economic man. *International Letters of Social and Humanistic Sciences*, (11): 10-34.

Van Dijk, Eric. 2015. The economics of prosocial behavior. En David Schroeder y William Graziano (eds.), *The Oxford handbook of prosocial behavior*. NY: Oxford UP.

Van Lange, Paul, David Cremer, Erik Van Dijky Mark Van Vugt. 2007. Self-interest and beyond. Basic principles of interaction. En A. W. Kruglanski y E. T. Higgins (eds.), *Social psychology: Handbook of basic principles*. The Guilford Press. 540-561.

Varian, Hal. 2010. *Microeconomía intermedia*. Barcelona: Antoni Bosch.

Viciana, Hugo. 2014. *¿Por qué somos morales? Una introducción a la ética naturalista*. Amazon, Kindle Publishing.

Weber, Max. 1997 [1913]. *The theory of social and economic organization*. Nueva York: Free Press.