

# Una simulación multi-agente del mecanismo de generalización de una norma social

*A Multi-Agent Simulation of a Social Norm Spreading Mechanism*

**Francisco Linares**

## Palabras clave

Expectativas • Normas • Sanciones • Simulación

## Key words

Expectations • Norms • Sanctions • Simulation

## Resumen

El presente artículo trata de abordar dos cuestiones interrelacionadas: ¿qué es una norma social? y ¿cómo se generaliza? Se argumenta que una norma social no es ni mero comportamiento sancionador ni tampoco comportamiento que tan solo se ajusta a las expectativas de los demás; sino que ambos requisitos deben cumplirse para que esta exista. A través de un modelo de simulación multi-agente se muestra cómo una dinámica de refuerzo mutuo entre la ejecución de sanciones y la generalización de expectativas explica la generalización de una norma social, dada la existencia de una masa crítica de sancionadores incondicionales. Además se muestra que a mayor densidad social, menor el tamaño de la masa crítica requerida en la población.

## Abstract

This article attempts to answer two related questions: what is a social norm? And how does it generalize? It is argued that a social norm is neither mere sanctioning behaviour, nor behaviour which just satisfies other's expectations; but both conditions are simultaneously required for a social norm to exist. By means of an agent-based simulation, it is shown how a reinforcing dynamic between the execution of sanctions and the generalization of expectations explains the generalization of a social norm, provided a critical mass of unconditional sanctioners does exist. Furthermore, it is shown that the larger the social density, the smaller the critical mass required in the population.

## INTRODUCCIÓN<sup>1</sup>

Este trabajo pretende abordar las preguntas: ¿qué es una norma social? y ¿cómo llega a

generalizarse una norma social? A pesar de su aparente sencillez, ninguna de estas cuestiones tiene una respuesta evidente. En la literatura especializada es frecuente encontrar voces muy autorizadas que aportan definiciones tan distintas como las que subrayan la aplicación de castigos cuando el comportamiento se desvía del usual entre los individuos (Axelrod, 1986: 1097), el grado de generalidad de las creencias sobre las expectativas de comportamiento (Bicchieri, 2006: 2) o las prescripciones internas de los individuos sobre lo que se debe o no hacer (Ostrom, 2005:

<sup>1</sup> El presente trabajo se ha desarrollado en el marco de un proyecto I+D+I con referencia CSO2009-09890 y de un proyecto CONSOLIDER-INGENIO con referencia CSD 2010-00034, ambos financiados por el Ministerio de Ciencia e Innovación (MICINN). Quiero agradecer a los miembros del grupo de investigación GSADI y a los dos evaluadores anónimos de la REIS su estímulo y sus detalladas sugerencias que han permitido mejorar notablemente una versión anterior de este trabajo.

121). Con todo, como apunta Karl D. Opp, la mayoría de los estudiosos manejan definiciones que suelen tener en común una serie de criterios que es posible aunar en una definición combinada (Opp, 2001: 103): una norma implica una regularidad de la conducta, implica la ejecución de sanciones y unas expectativas sobre el comportamiento compartidas por una colectividad de individuos.

La tesis que sostenemos en este artículo es que el grado de generalización de la conducta prescrita normativamente en la colectividad depende de un proceso no planificado en el que las expectativas y las sanciones de los individuos van moldeando su comportamiento, conduciendo a un estado en el que la conducta normativa se hace regular, al tiempo que las expectativas sobre el comportamiento de los demás se vuelven más homogéneas y las sanciones sobre el comportamiento se hacen menos frecuentes. Como se verá este proceso se ve facilitado o inhibido por el grado de densidad social del sistema en el que emerge la norma.

Para dar cuenta de este proceso se ha construido un modelo de sociedad artificial en el que los agentes deben consumir un determinado recurso (e.g. agua, leña, etc.) para sobrevivir, pero un consumo superior al necesario para la supervivencia produce la desaparición del recurso y, por ende, de la propia población de agentes, generándose así un dilema del tipo «tragedia de los comunes» (Hardin, 1968). Los agentes de nuestro modelo no actúan con el fin de evitar tal tragedia; no obstante el modelo muestra como, en el transcurso de las interacciones, puede que adopten un tipo de comportamiento que les permite sobrevivir. Un comportamiento que, de acuerdo con los rasgos señalados anteriormente, se corresponde con una norma social.

El uso de la técnica de la simulación es adecuada para explorar qué condiciones, y qué mecanismos, son responsables de la generalización de una norma social porque en los estudios de campo no es posible respon-

der a preguntas del tipo: ¿qué ocurriría si los sujetos no tuvieran expectativas sobre el comportamiento? La técnica de la simulación, sin embargo, permite establecer artificialmente condiciones experimentales de otra forma imposibles y obtener así conclusiones rigurosas y precisas sobre la dinámica de sistemas complejos. No obstante, la validez de sus resultados depende crucialmente del tipo de asunciones realizadas para modelar el comportamiento y la interacción entre los agentes, considerándose una práctica deseable que tales asunciones estén respaldadas por trabajo empírico, especialmente de tipo experimental. Así mismo, si los resultados obtenidos son muy sensibles a la variación de parámetros establecidos más o menos arbitrariamente, su validez queda también en entredicho.

Desde los años ochenta del pasado siglo, la simulación multi-agente ha conocido un auge creciente en las ciencias sociales (González Bailón, 2004; Gilbert y Troitzsch, 2006; García-Valdecasas, 2010). A partir del trabajo pionero de Robert Axelrod (1984, 1986) se ha publicado un importante número de artículos que bien están expresamente dedicados al análisis de las normas sociales (e.g. Coleman, 1990b; Heckathorn, 1990; Kitts, 2006a, 2006b) o bien a algún aspecto estrechamente vinculado con las mismas. Por razones que quedarán claras más adelante, nos interesan especialmente tres tipos de aportaciones: modelos de «influencia social» (e.g. Gould, 1993; Axelrod, 1997; Groeber *et al.*, 2009; Fischbacher y Gächter, 2010), modelos de «sanciones descentralizadas» (e.g. Axelrod, 1986; Heckathorn, 1989; Macy, 1993; Kitts, 2006a; Carpenter, 2007) y modelos de «masa crítica» (e.g. Marwell *et al.*, 1988; Kim y Bearmann, 1997; Chew, 1999).

Así, por ejemplo, el reciente trabajo de Giangiacomo Bravo (2010) simula agentes que hacen uso de un recurso común en función de sus «creencias» sobre el estado del recurso (es decir, cuánto puede explotarse); estas creencias guían el comportamiento de

los agentes, ajustándose en función de las consecuencias de dicho comportamiento. En un segundo escenario, Bravo asume que existe una institución que regula el grado de explotación permitido. La regla establecida en cada momento depende de las creencias de los sujetos, y sufre un proceso de ajuste cuando desfavorece a una mayoría amplia de la población. En este escenario las decisiones de los agentes conducen a la institución hacia una regla de gestión de los recursos que evita la tragedia de los comunes.

Nuestro trabajo difiere del de Bravo en varios aspectos. En primer lugar, nuestro modelo no se centra en las creencias de los agentes sobre los recursos, sino en las creencias de los agentes sobre las expectativas que los otros agentes pueden tener sobre el uso de los recursos. En este sentido incorpora un elemento de influencia social, cuya ausencia el propio Bravo reconoce como una deficiencia en su trabajo (Bravo, 2010: 147). En segundo lugar, Bravo toma por resueltos problemas de diseño institucional que, en la práctica (y en la teoría), no tienen una solución simple. Entre ellos, el problema de la vigilancia y sanción de los potenciales desertores, ya que asume que ninguno de sus agentes se desvía de la regla establecida institucionalmente (Bravo, 2010: 137). Nuestro modelo no trata de simular la emergencia de una institución, sino de una norma social. Una norma social no presupone necesariamente la existencia de un conjunto de reglas formales para la toma de decisión colectiva, pero sí presupone la posibilidad de ejecutar sanciones sociales de forma descentralizada y no planificada. Por tanto, nuestro modelo explícitamente formaliza tal posibilidad.

En este sentido el modelo propuesto en este artículo es más similar al de James Kitts (2006a), en el que 17 agentes que constituyen una red neuronal están sometidos tanto a una «influencia mimética» (*mimetic influence*) como a una «influencia inductiva» (*inductive influence*), términos con los que Kitts

pretende referirse a los procesos por los que un agente ajusta su comportamiento en función del «group standard» (Kitts, 2006a: 411), y en función de las sanciones, positivas o negativas, recibidas de sus iguales. A diferencia del trabajo de Kitts, sin embargo, nuestros agentes ni están incrustados en una red dada de relaciones, ni tienen en consideración las valoraciones y las sanciones que su comportamiento merece para todos y cada uno de los agentes restantes en el grupo, sino que vagan por el espacio social interactuando con aquellos agentes que, casualmente, encuentran en su camino. Además, por razones que se expondrán más adelante, en nuestro modelo no todos los agentes tienen capacidad para sancionar, lo que hace el cómputo necesario para la toma de decisiones menos complejo y menos artificioso, especialmente para grupos no reducidos.

Hasta donde alcanza nuestro conocimiento de la materia, el análisis presentado en este artículo es novedoso en dos sentidos: primero, un importante número de trabajos previos están dirigidos a explicar el sostenimiento de la cooperación entre agentes *racionales*; esto puede ser resultado de mecanismos distintos al de la norma social, que es el que específicamente nos preocupa en este trabajo. Segundo, aunque existen estudios que abundan en el conocimiento de cómo operan sanciones y creencias, no conocemos ninguno que explícitamente estudie la relación entre ambos conceptos.

En adelante, el desarrollo del artículo sigue el siguiente orden: en primer lugar se diseccionan los componentes de una norma social y se pasa revista a la forma en que estos componentes han sido tratados en la literatura de simulación social. En segundo lugar se exponen los detalles del modelo. Posteriormente se extraen los resultados más significativos y se examina la robustez de los mismos. Finalmente se discuten estos resultados con el fin de echar luz sobre los interrogantes formulados al inicio de esta introducción.

## EL PROBLEMA TEÓRICO: ¿QUÉ ES UNA NORMA SOCIAL Y CÓMO FORMALIZARLA?

### El concepto de norma social

En las últimas dos décadas, especialmente en los años recientes, se han venido realizando aportaciones significativas al estudio de las «normas sociales». Tales aportaciones provienen de campos tan dispares como la sociología empírica (Hechter y Opp, 2001), la economía experimental (Fehr *et al.*, 2002), la sociología experimental<sup>2</sup> (Horne, 2007), la economía institucional (Ostrom, 2005), la filosofía analítica (Bicchieri, 2006), la teoría social analítica (Elster, 2007, 2009) y la simulación social (Groeber *et al.*, 2009).

Si hay un rasgo en común entre estas aproximaciones es que todas ellas asumen la premisa de que una norma social es un mecanismo que intentan resolver un *dilema social*, entendiéndose como tal una situación de interdependencia en la que existe una tensión entre el interés individual a corto plazo y el interés colectivo<sup>3</sup>; insertándose así en una tradición teórica ya afianzada en la sociología (Ullman-Margalit, 1977; Coleman, 1990a). Por supuesto, esto no implica que la norma emergente siempre acabe resolviendo satisfactoriamente la tensión entre estos intereses (Kitts, 2006b), ni que sea consecuencia de un proceso intencionalmente dirigido.

No obstante, más allá de este punto de partida, parece haber poco más entendimiento común de la cuestión. Donde para algunos las «emociones» como la indignación y la culpa son la fuerza que dirige el comportamiento propio de una norma social (Elster, 2009; Francisco, 2010), para otros lo

son las «expectativas» (Bicchieri, 2006) y aun para otros los «mecanismos de sanción», que pueden estar asociados a disposiciones motivacionales específicas de la especie humana, como la denominada «reciprocidad fuerte» (*strong reciprocity*) (Gintis, 2000), o a consideraciones estrictamente utilitarias (Horne, 2007). Sin menoscabo de ninguna de estas aportaciones, lo que se antoja a todas luces evidente es que una norma social difícilmente puede reducirse a uno solo de estos elementos.

Aunque frecuentemente el concepto de sanción se considera un componente indispensable de la definición de una norma social, lo cierto es que pensar una norma social como un tipo de conducta (cumplidora) que responde a otro (sancionadora) impediría, como señala Coleman (1990a), distinguir una norma social de meros gestos de aprobación mutua, como los que puedan intercambiarse dos colegas en el trabajo. Sin duda alguna, una norma no puede verse circunscrita a intercambios bilaterales, sino que responde, como señala Bicchieri (2006), a una *expectativa generalizada*. No obstante, por otra parte, si tal expectativa no está fundamentada en mecanismos de sanción, no podría distinguirse una norma social de otros tipos de conducta que también son fruto del ajuste mutuo de creencias, como las rutinas sociales.

Tanto el comportamiento de los agentes que se ajustan a la norma como los de los agentes que sancionan pueden estar fundamentados en diversas fuerzas motivacionales más allá de las consideraciones utilitarias, como los compromisos morales, las emociones o las motivaciones pro-sociales; aunque no necesariamente en las mismas. Esto puede hacer difícil distinguir en la práctica las normas sociales de otro tipo de normas (Tena y Güell, 2011). Estas motivaciones en sí mismas, sin embargo, no constituyen normas sociales, puesto que la norma no es una propiedad interna del sujeto sino una propiedad del sistema social. Así, el abstenerse de sus-

<sup>2</sup> Entendemos por «sociología experimental» la investigación sociológica empírica que emplea el método de experimentación.

<sup>3</sup> Se entiende por interés colectivo la producción de un «bien público», según la definición estándar de los textos de economía.

traer objetos ajenos no sería una norma social si tal comportamiento respondiera a una convicción moral más que a la esperable reacción de los otros sujetos.

Con todo, dada la necesidad de operar con una definición concreta, en este trabajo daremos por buena la aportada por C. Horne: «I define norms as rules, about which there is at least some degree of consensus, that are enforced through social sanctions. Understanding the emergence of norms, then requires explanations of norm content, enforcement, and distribution» (2001: 5). Esta definición, entendemos, es compatible con todas las aportaciones señaladas; tiene la ventaja de resaltar tanto que una norma requiere un cierto grado de consenso (es decir, unas expectativas compartidas sobre el comportamiento requerido) como que se hace cumplir a través de sanciones sociales (es decir, sanciones que son proporcionadas por los propios individuos, no por una agencia centralizada); y, a su vez, deja abierta a la investigación el estudio de los posibles mecanismos concretos por los que una norma recibe su contenido, grado de sanción y distribución.

En otro trabajo (Linares, 2007) hemos tratado de avanzar hipótesis sobre el contenido de las normas sociales. Aquí se mostrará que las cuestiones de cómo se hace cumplir y de cuál sea su distribución o grado de generalización se hallan analíticamente vinculadas. Para ello se ha procedido a la construcción de un modelo de simulación multi-agente, empleando el *software* Netlogo, en el que los agentes se enfrentan a un dilema social típico. El diseño de este modelo pretende reflejar los elementos más relevantes de una norma social, tal y como se acaban de exponer. Este método de investigación es apropiado porque no existe un procedimiento alternativo que nos permita diseccionar una norma y estudiar el impacto diferencial de cada uno de sus componentes en su proceso de generalización. En concreto el modelo nos permitirá echar luz sobre tres cuestiones interrelacionadas: ¿es posible una solución des-

centralizada de los dilemas sociales?, ¿en qué condiciones un subconjunto minoritario (una «masa crítica») de individuos cooperadores conseguirán que se expanda la cooperación entre el resto de la población?, ¿cuál es el mecanismo que, provista tal masa crítica, permite la generalización de una norma social?

### **La simulación de una norma social en una sociedad artificial**

Tradicionalmente las simulaciones de *dilemas sociales* han adoptado el del Prisionero como estructura básica de interacción (Axelrod, 1986; Heckathorn, 1988; Coleman, 1990b). En nuestro caso se ha optado por crear un entorno tipo «tragedia de los comunes»<sup>4</sup> que, sin embargo, no está formalizado como un problema típico de teoría de juegos, en el que los actores conocen los pagos de cada estrategia y buscan aquella que les proporciona un mayor beneficio. Alternativamente, los agentes, que tienen solo un conocimiento local de su entorno, sufren las consecuencias de sus decisiones, pero no tienen poder de anticipación y de cálculo. Dicho de una manera sencilla, el modelo es respetuoso con el hecho de que dos extraños que, por ejemplo, talaran árboles en extremos opuestos del bosque alterarían recíprocamente sus posibilidades de supervivencia, aunque no tuvieran conocimiento de ello y, por tanto, no consideraran este hecho en sus decisiones.

Por otra parte, el concepto de *expectativa*, que implica que un individuo cambia su comportamiento para ajustarlo a lo que él cree (o espera) que los individuos de su entorno esperan de él, se representa en nuestro modelo a través de un mecanismo análogo al que en otros se ha denominado «influencia social». Así, en un ejemplo significativo (Gould, 1993)<sup>5</sup>, los agentes ajustan sus con-

<sup>4</sup> Véase Janssen y Ostrom (2006) para una revisión de la literatura de simulación en este campo

<sup>5</sup> Véase también Kitts (2006a).

tribuciones a un bien público en función de las contribuciones de los demás, según un parámetro que refleja la fuerza de la influencia que los agentes ejercen los unos sobre los otros. El modo en que opera este proceso en nuestro modelo sigue una lógica similar, aunque es más parecido al de Axelrod (1997), en el que cada agente solo tiene en cuenta las acciones de los agentes en su entorno inmediato, y no de toda la población.

Un tercer elemento del modelo es el mecanismo de *sanción*. Desde un punto de vista teórico, la explicación de un proceso de sanción descentralizado es problemático porque, si el acto sancionador en sí mismo es costoso, los agentes pueden verse atrapados en un dilema social de segundo orden. Una familia de modelos evolutivos (Axelrod, 1986; Heckathorn, 1989, 1996; Macy, 1993) exploró las posibilidades de emergencia de un sistema de sanciones autosostenible; y un resultado robusto de los mismos subraya la importancia de la conducta hipócrita (sancionar a quien no contribuye, pero no contribuir uno mismo) como desencadenante de un efecto de bola de nieve que acaba produciendo una explosión de cooperación. No obstante, los experimentos que valoran la contribución voluntaria a la producción de bienes públicos indican que, cuando existe la posibilidad de sancionar a los incumplidores, una importante proporción de sujetos muestran una predisposición a ejecutar sanciones con independencia de su coste (Fehr y Gächter, 2000); así, en la práctica, el posible dilema de segundo orden suele resolverse con facilidad (Gächter y Thöni, 2011). En línea con estos resultados, hemos asumido que las sanciones son ejecutadas incondicionalmente por un determinado tipo de agentes. De esta forma nuestro modelo es distinto tanto de los recién señalados, en los que todos los agentes tienen capacidad sancionadora, como de otros (e.g. Bravo, 2010) que asumen un mecanismo de sanción centralizado.

Finalmente, al restringir la ejecución de sanciones a un subconjunto de la población

surge naturalmente la pregunta sobre cuál es el tamaño de ese subconjunto que permite el sostenimiento de la conducta guiada por una norma social. En otras palabras, ¿cuál es el tamaño mínimo de la «masa crítica» de sancionadores incondicionales necesaria para la generalización de la norma social? Los modelos que han explorado el problema de la masa crítica (Marwell *et al.*, 1988; Kim y Bearmann, 1997; Chew, 1999) coinciden en señalar que determinados rasgos de la topología del grupo (especialmente alta centralización y alta densidad de las redes sociales) favorecen la expansión de la cooperación a partir de un reducido número de individuos. En nuestro caso, sin embargo, no se establece una topología fija a la que estén sujetos los agentes sino que, como se señaló, estos se desplazan aleatoriamente e interactúan unos con otros cuando accidentalmente coinciden en el curso de su desplazamiento. No obstante, el modelo permite someter al entorno a una mayor/menor presión, al aumentar o disminuir la densidad de población, lo que a su vez afecta la frecuencia (o densidad) de las interacciones. Como se verá, este mecanismo tiene un efecto consistente con los hallados en los estudios precedentes.

## EL MODELO

### El dilema social

La «tragedia» a la que se enfrentan nuestros agentes es la siguiente: las parcelas por las que se desplazan los agentes tienen una reserva de un determinado recurso consumible que, a título ilustrativo, en adelante consideraremos que es agua. Dicha reserva es asignada de forma aleatoria al inicio de la simulación y, con independencia de la acción humana, aumenta o disminuye a consecuencia de la caída (o no) de la lluvia. El modelo se ha programado para que, en ausencia de agentes, las parcelas mantengan en término medio su reserva de agua estable a lo largo del tiempo.

Por su parte, los agentes se desplazan aleatoriamente por las parcelas. En su desplazamiento pierden energía (una propiedad común a todos los agentes, asignada por igual al comienzo de la simulación) que solo pueden reponer tomando agua de las parcelas. Si el nivel de energía llega a cero (es decir, si el agente no es capaz de encontrar agua con la que reponer la energía gastada), el agente muere y la población, por tanto, queda mermada.

Al encontrar una parcela con suficiente agua almacenada<sup>6</sup> los agentes pueden tomar una cantidad «grande» o tomar una «pequeña», lo que aumenta proporcionalmente su energía pero disminuye, también de forma proporcional, el recurso disponible no solo para este agente sino también para los demás. Esto crea un entorno característico de los dilemas sociales. El modelo se ha diseñado de forma que cuando todos los agentes toman una cantidad pequeña (cooperan) el cien por cien de la población sobrevive. En el otro extremo, cuando todos los agentes toman una cantidad grande (desertan), el efecto de la lluvia no es capaz de reponer el agua consumida por los agentes y el cien por cien de la población muere. Entre los dos casos extremos señalados hay un número de posibilidades que se exploran manipulando los parámetros indicados después. Dado que el modelo no asume que la población pueda reproducirse, la ejecución de la simulación producirá necesariamente uno de estos tres resultados: toda la población inicial sobrevivirá, toda la población inicial perecerá o la población quedará estable en algún punto entre esos dos extremos.

### **Tipos de agentes y reglas de decisión**

Es un hecho experimentalmente asentado que en una población de individuos enfrenta-

dos a un dilema social es posible encontrar una pluralidad de motivaciones entre las que destacan cuantitativamente la reciprocidad fuerte, la reciprocidad débil y el egoísmo incondicional (Fischbacher *et al.*, 2001). Es decir, aunque pocos individuos cooperan siempre incondicionalmente, algunos lo hacen si otros también cooperan, otros además muestran una predisposición a sancionar incondicionalmente y, finalmente, otros no cooperan nunca. Simplificando esta heterogeneidad motivacional, se han diseñado dos tipos de agentes: «cooperadores incondicionales» y «cooperadores condicionales»<sup>7</sup>. El modelo siempre genera 100 agentes, permitiendo manipular el número de «cooperadores incondicionales» con el que se inicia cada simulación. La regla de comportamiento que sigue este tipo de agente es que, cuando quiera que se encuentren en una parcela con agua, siempre toman una cantidad pequeña. Este comportamiento es «asocial» en el sentido weberiano de que no tienen en consideración a otros agentes para actuar: su cooperación es incondicional.

El número de cooperadores condicionales en cada simulación es 100 menos el número de cooperadores incondicionales establecido. Por defecto, la regla de comportamiento que siguen estos agentes es la de cooperar o no en cada iteración con igual probabilidad. Es decir, estos agentes están programados para no seguir una pauta fija y, por tanto, su comportamiento es también «asocial» en el sentido de que no se ve influido por el comportamiento de otros agentes (no obstante, el modelo permite que su comportamiento devenga social a través de dos mecanismos que pueden activarse conjunta o independientemente). El hecho de que estos agentes estén programados para cooperar o desertar aleatoriamente en cada itera-

<sup>6</sup> Se asume que, por debajo de un umbral, la parcela no dispone de agua suficiente para ser usada.

<sup>7</sup> Para evitar confusión es preciso aclarar que el término «cooperador condicional» en nuestro caso no hace referencia al agente que sigue estrategias del tipo «tit-for-tat».

ción con igual probabilidad implica que, en el conjunto de la población, estadísticamente siempre habrá una proporción de «free riders». Esta proporción no aumenta o disminuye como fruto del aprendizaje o del cálculo racional de los agentes (puesto que nuestro modelo no simula estos procesos), pero sí como fruto de los mecanismos de «expectativas sociales» y «sanciones descentralizadas» que se detallan a continuación.

### Expectativas sociales

Cuando el mecanismo «expectativas sociales» está activado, los cooperadores condicionales prestan atención a la ratio de agentes en su entorno<sup>8</sup> que muestra un comportamiento cooperativo. Con esta ratio, que denotamos  $e_i$ , se pretende operacionalizar la creencia del agente de que la población de su entorno sigue (o no) una norma de cooperación y, por tanto, la creencia de que se espera de él un comportamiento de este tipo. Este valor es tanto mayor cuanto más alto sea el número de agentes que cooperan en su entorno en relación al número total de agentes en el mismo. Como los agentes se desplazan por el mundo virtual, para cada uno de ellos el valor de  $e_i$  no solo será distinto al de otros agentes, sino que también variará en las sucesivas iteraciones.

Además, los cooperadores condicionales tienen una propiedad denominada «umbral de expectativa-sanción»,  $u_i$ , cuyo valor para cada agente varía entre 0 y 1, que es asignado aleatoriamente al inicio de la simulación y permanece fijo. Dicho umbral es una característica exclusivamente de este tipo de agentes y trata de operacionalizar la sensibilidad de cada individuo al hecho de que otros de su entorno sean cooperadores y puedan tener, por tanto, la expectativa de que él siga la misma regla de comportamiento. Así,

cuanto más bajo sea el valor de  $u_i$  menor proporción de cooperadores se requerirá en el entorno del agente para que este cambie su conducta. De esta manera, con el mecanismo «expectativas sociales», un agente cooperador condicional cooperará siempre que  $e_i > u_i$ . Esto permite reflejar la heterogeneidad entre los individuos y dar más realismo al modelo, ya que en la sociedad no-virtual no todos los individuos son igualmente sensibles a las expectativas de los demás: cuantos más cooperadores haya en su entorno más probable será que un cooperador condicional cualquiera coopere, pero no todos lo harán igualmente.

### Sanciones descentralizadas

Los modelos más tempranos (e.g. Axelrod, 1986) tratan el problema de cómo es posible incentivar la cooperación a través de actos de sanción que, a su vez, están sometidos a un posible dilema social. Este tipo de sanción «racional», sin embargo, probablemente no sea la más frecuente desde el punto de vista empírico ya que, como se señaló anteriormente, lo que los experimentos sobre la cooperación en bienes públicos rotundamente demuestran es que existe un conjunto de individuos que sanciona con independencia del coste. Es por ello por lo que en nuestro modelo se opta por un mecanismo de sanción diferente, asumiendo que los agentes «cooperadores incondicionales» tienen una predisposición a sancionar también incondicional.

Los agentes cooperadores condicionales, por su parte, no sancionan nunca pero son sensibles en diversa medida a las posibles sanciones de los anteriores. Así, ante la presencia de un cooperador incondicional en una parcela adyacente el cooperador condicional cooperará si un número aleatorio entre 0 y 1 es superior al «umbral de expectativa-sanción». De esta forma,  $u_i$  cumple también el papel de indicador de la sensibilidad de cada individuo bien al reproche moral (si no

<sup>8</sup> El entorno del agente se ha operacionalizado en Netlogo como los agentes dentro de un radio 4. Véase, más adelante, la discusión sobre este parámetro.



coopera) o al gesto de aprobación (si coopera) procedentes de un cooperador incondicional<sup>9</sup>. Lo que a su vez implica que el contenido, prescriptivo o proscriptivo (León, 2011), de la norma no queda especificado en la simulación sino que esta es formalmente compatible con ambos. Así, manteniendo la coherencia con el mecanismo de «expectativas sociales», cuanto más alto sea  $u_i$ , no solo harán falta más cooperadores en su entorno para inducir la cooperación sino que, además, menos probable será que se vea influido por un acto de sanción.

### **Variables poblacionales y condiciones experimentales**

En resumen, los agentes del modelo no actúan racionalmente en el sentido de anticipar las consecuencias futuras de las acciones y valorar las distintas alternativas. Tampoco actúan según el canon conductista de valorar las consecuencias de las acciones pasadas. Más bien toman decisiones en cada momento discreto de tiempo teniendo en consideración, para el caso de los cooperadores incondicionales, exclusivamente la existencia o no de agua en cada parcela y, para el caso de los cooperadores condicionales, además, el comportamiento de otros agentes en su entorno y/o la posibilidad de ser sancionados. Al vagar por el espacio social, los agentes se encontrarán de forma aleatoria (como los vecinos que salen a pasear por el barrio). Esto permite que el problema de la monitorización del comportamiento se resuelva de forma descentralizada (es decir, sin recurrir a una agencia central de inspección). Partiendo de estas premisas el modelo muestra cómo, con el correr del tiempo, la población

sobrevive o no a consecuencia de la imitación, o no, de la conducta cooperadora mostrada inicialmente por un subconjunto de los agentes. Tras realizar 8.000 iteraciones el modelo deja de ejecutarse, si antes no se ha extinguido toda la población. El estado resultante del sistema puede caracterizarse con las siguientes variables:

- Población superviviente (P): número total de agentes en la última iteración.
- Cooperadores incondicionales finales (I): número de cooperadores incondicionales en la última iteración.
- Cooperadores finales (C): número de cooperadores condicionales que cooperaron en la última iteración.
- Desertores finales (D): número de cooperadores condicionales que desertaron en la última iteración.
- Expectativa generalizada (E): media de las  $e_i$  de los individuos en la última iteración.
- Sanciones acumuladas (S): suma de sanciones ejecutadas al finalizar la simulación.

Con el objeto de desentrañar los mecanismos que permiten la expansión de la conducta cooperadora, se establecieron cuatro condiciones experimentales distintas:

- 1) Condición Experimental 1 (CE1) (condición asocial): en este escenario los cooperadores condicionales no tienen en consideración la presencia de otros agentes y, por tanto, actúan siempre usando la regla por defecto de cooperar y desertar con igual probabilidad. Se trata del modelo base con respecto al cual comparar los resultados de las otras condiciones. Las variables poblacionales a medir en este caso son: P, I, C y D.
- 2) Condición Experimental 2 (CE2) (condición de expectativas sociales): en este escenario los cooperadores condicionales actúan teniendo en consideración la ratio cooperadores en su entorno/total de agentes en su entorno, según la explicación ofrecida anteriormente.

<sup>9</sup> El uso de un mismo parámetro,  $u_i$ , para establecer el umbral de dos características distintas de los agentes es obviamente una estrategia de simplificación. Nótese, no obstante, que el programa realiza dos operaciones independientes (una relacionada con las expectativas y otra relacionada con las sanciones), comparando dicho umbral en cada caso con un número distinto.

Las variables poblacionales a medir en esta condición son: P, I, C, D y E.

- 3) Condición Experimental 3 (CE3) (condición de sanciones descentralizadas): en este escenario los cooperadores condicionales actúan teniendo en consideración las sanciones de los cooperadores incondicionales en su vecindario inmediato, según la explicación ofrecida anteriormente, pero no el porcentaje de cooperadores en su entorno. Las variables poblacionales a medir en esta condición son: P, I, C, D y S.
- 4) Condición Experimental 4 (CE4) (condición de expectativas sociales y sanciones descentralizadas): en este escenario los cooperadores condicionales actúan teniendo en consideración tanto la ratio de cooperadores como la posibilidad de ser sancionado. Las va-

riables poblacionales a medir en esta condición son: P, I, C, D, E y S.

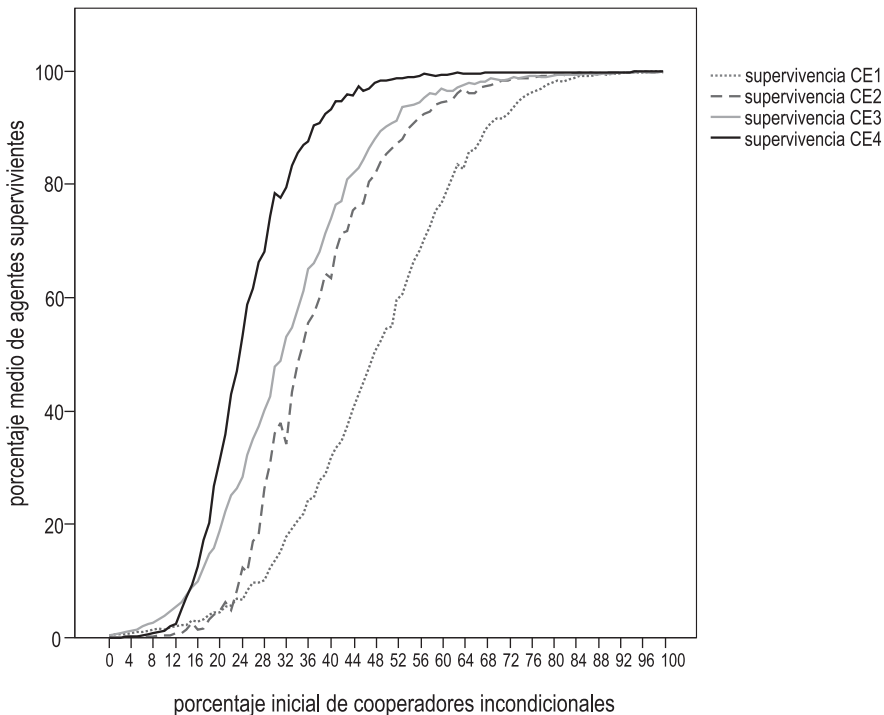
Para medir el efecto del número de cooperadores incondicionales introducidos en la población en cada condición experimental, se ejecutaron en cada condición cien repeticiones por cada nuevo cooperador incondicional añadido a la población inicial (desde 0 hasta 100); lo que en total suman 40.400 simulaciones de 8.000 iteraciones en el conjunto de las cuatro condiciones experimentales establecidas.

## RESULTADOS

### Población superviviente

Los resultados más evidentes aparecen en el gráfico 1, que muestra la media de supervivencia de la población en cien repeticiones,

**GRÁFICO 1.** Población superviviente



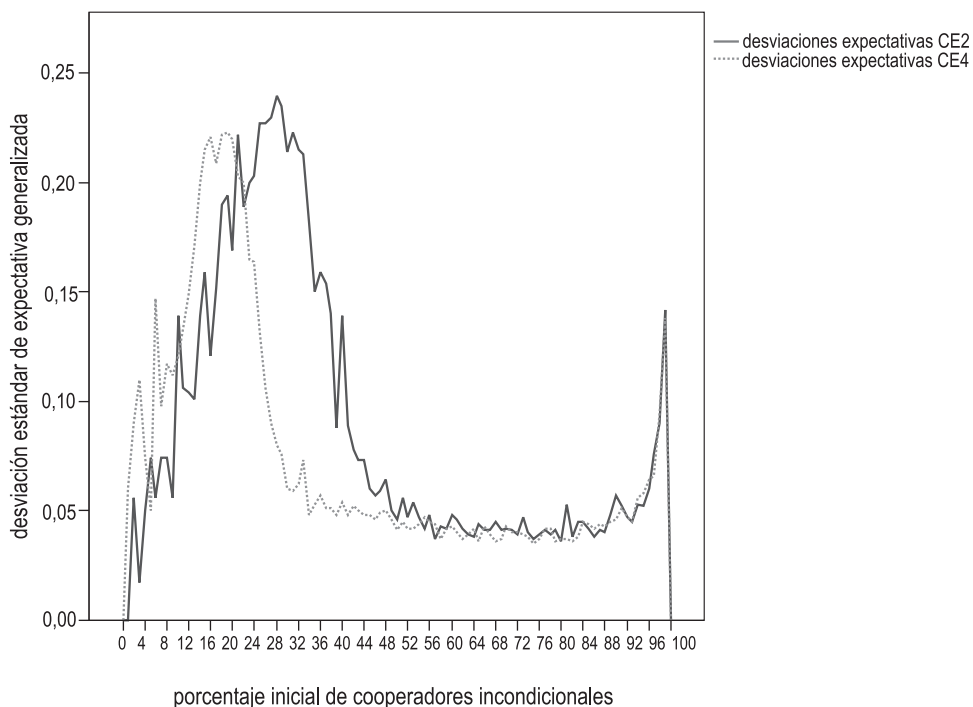
para cada condición experimental, en función del porcentaje de cooperadores incondicionales introducidos sucesivamente en la población. Las cuatro líneas dibujan una característica S que se inicia en el origen de coordenadas y finaliza en la esquina superior derecha, indicando que, como se señaló anteriormente, el cien por cien de cooperación implica el cien por cien de supervivencia en todas las condiciones experimentales. El hecho de que las cuatro líneas corran paralelas sencillamente indica que el modelo funciona según se espera, de forma que con la exposición a las expectativas sociales (CE2) o a las sanciones (CE3) se producen mayores niveles de supervivencia que en el estado asociado (CE1); y que ambas condiciones juntas (CE4) producen mayor supervivencia que cada una por separado. A título ilustrativo, la introducción de 45 cooperadores incondicionales eleva la supervivencia media en la CE1

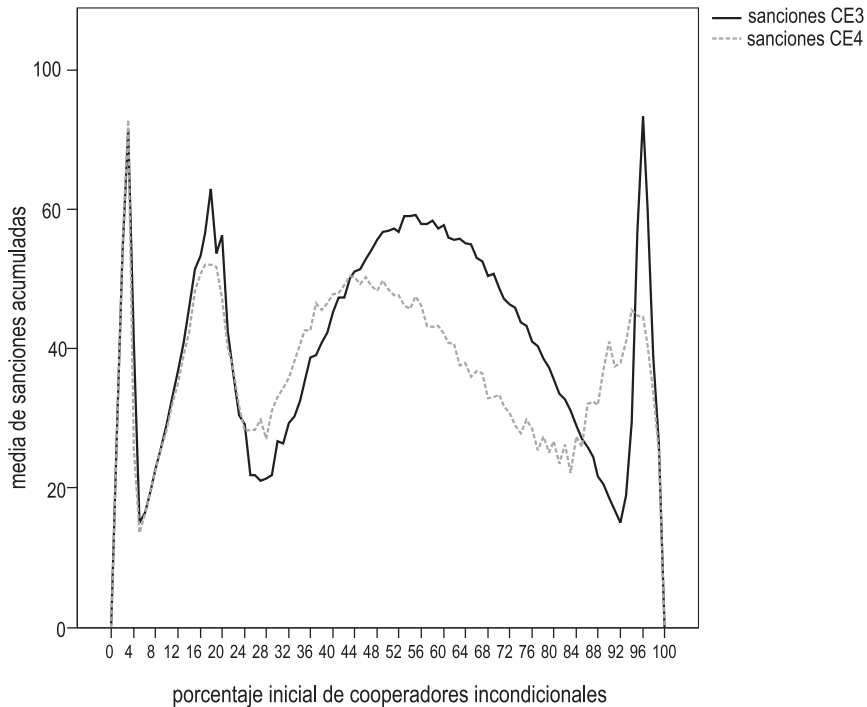
de 0,36 agentes (cuando no hay ningún cooperador incondicional) a 43,03 agentes, pero la eleva de 0,00 a 97,30 agentes en la CE4.

### Expectativas generalizadas y sanciones acumuladas

En el gráfico 2 se representa el promedio en 100 repeticiones de las desviaciones típicas de la variable «expectativa-generalizada», en función del número de cooperadores incondicionales, en CE2 y CE4. Puede observarse cómo los valores de E en ambas condiciones son más altos en la mitad izquierda del gráfico y caen progresivamente hasta mantenerse más o menos constantes a lo largo de la mitad derecha del eje de abscisas, donde el comportamiento de esta variable es muy similar en ambas condiciones experimentales. Esto responde a la intuición de sentido común de que cuando hay más cooperadores

**GRÁFICO 2.** *Expectativas generalizadas*



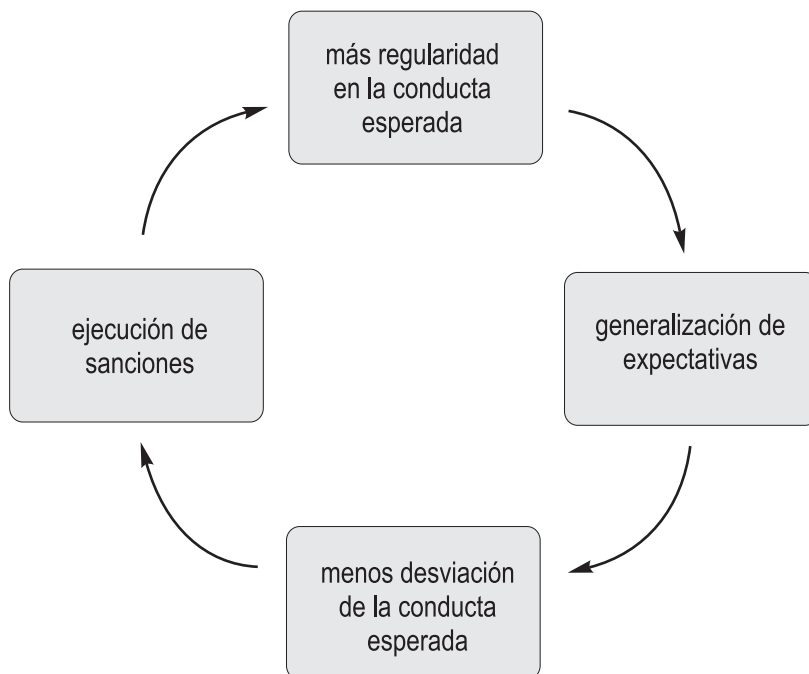
**GRÁFICO 3.** Sanciones acumuladas

incondicionales las expectativas de los individuos varían menos puesto que hay más regularidad en la conducta.

La información más importante que proporciona este gráfico es, empero, que la línea que representa la variación de E en la CE4 comienza su caída antes que la que la representa en la CE2. Así, en la condición de expectativas más sanciones, E empieza a caer a partir del valor 19 del eje de abscisas mientras que, en la condición de expectativas, sigue ascendiendo hasta alcanzar su máximo en el valor 28. La brecha entre ambas líneas alcanza su mayor amplitud entre los valores 28 y 32. Lo que esto significa es que, en las poblaciones en las que los cooperadores incondicionales representan poco más del 25 por ciento, la exposición a sanciones tiene el efecto de reducir la variabilidad de las expectativas de los cooperadores condicionales.

Esto es debido a que el cambio de conducta inducido en algunos agentes por la exposición a sanciones incrementa la ratio de cooperadores en el entorno de otros agentes que, a su vez, cambian su comportamiento. La existencia de sanciones permite así hacer más homogéneas las expectativas que los agentes tienen sobre el comportamiento cooperativo de los demás.

El gráfico 3 muestra la evolución de la media de las sanciones acumuladas en 100 repeticiones, en CE3 y CE4, en función del porcentaje de cooperadores incondicionales. Esta variable parece tener un comportamiento caprichoso tanto en los valores más bajos como en los más altos del eje de abscisas. No obstante, las dos líneas corren paralelas hasta el valor 43 del mismo, en el que la línea que representa el comportamiento en la CE4 muestra un máximo local de 50,47 sanciones

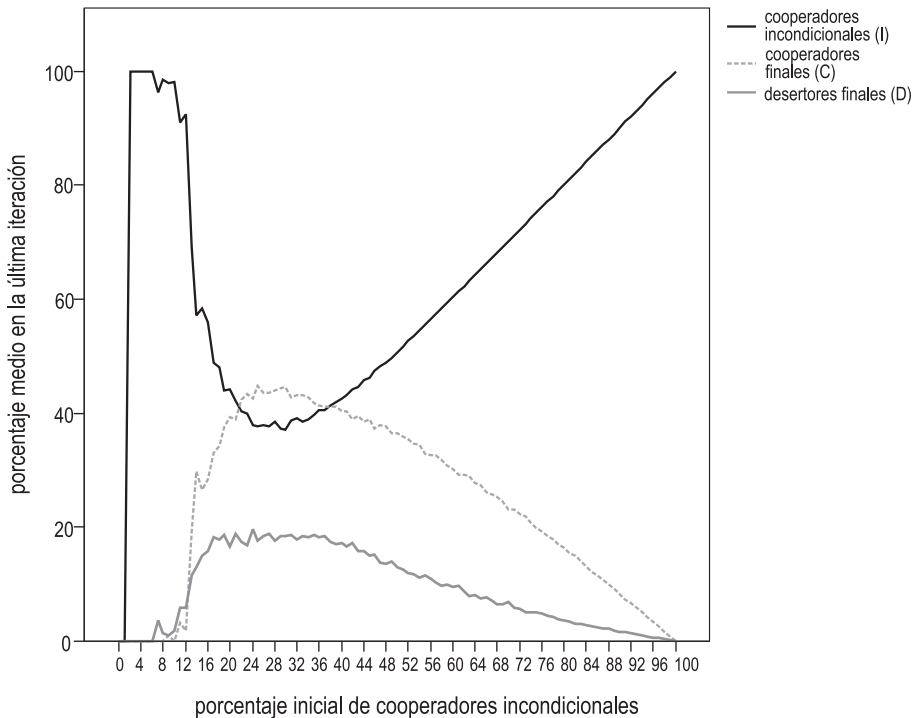
**DIAGRAMA 1.** Mecanismo de retroalimentación entre sanciones y expectativas

acumuladas (6,3 sanciones cada mil iteraciones) mientras que la línea de la CE3 sigue ascendiendo hasta alcanzar su máximo local de 59,20 sanciones acumuladas (7,4 sanciones por cada mil iteraciones) en el valor 55 del eje de abscisas. Ambas líneas vuelven a cruzarse posteriormente en el valor 86. Esto indica que, en el espectro de cooperadores incondicionales que va desde los 43 hasta los 85, la activación del comportamiento guiado por expectativas sociales tiene el efecto de reducir la tasa de sanción. O, lo que es lo mismo, los individuos son menos sancionados porque tienen en consideración el comportamiento de los agentes de su entorno. Estos resultados son complementarios de los inferidos del gráfico 2. Ambos se representan en el diagrama 1, que muestra el proceso de retroalimentación entre sanciones y expectativas que constituye el mecanismo principal de expansión del comporta-

miento cooperativo entre los cooperadores condicionales.

### **Evolución de subpoblaciones**

Los datos de la evolución de las subpoblaciones en la condición experimental de expectativas más sanciones, presentados en el gráfico 4, ilustran el efecto del proceso descrito. Este gráfico indica el porcentaje *final* de C y D en la población, en función del porcentaje *inicial* de cooperadores incondicionales, y muestra cómo, salvo en los valores más bajos del eje de abscisas, se produce una explosión de la cooperación. Más aún, la existencia de expectativas y sanciones conjuntamente permite que en un amplio espectro de valores del eje de abscisas (entre 13 y 40) se genere un número de cooperadores finales mayor que el número de cooperadores incondicionales iniciales, teniendo su máximo en el valor 25 de este eje. En este

**GRÁFICO 4.** *Porcentajes finales de tipos de agente en CE4*

caso en concreto, la población evoluciona (en una simulación típica) desde una composición promedio de un 25 por ciento de cooperadores incondicionales y un 75 por ciento de cooperadores condicionales que cooperan y desertan con igual probabilidad, al inicio de la simulación, hasta un 17,53 por ciento de desertores, un 37,73 por ciento de cooperadores incondicionales y un 44,74 por ciento de cooperadores en promedio al final de la misma. Este efecto no ocurre en ninguna otra condición experimental y, por tanto, puede legítimamente considerarse como ilustración de la emergencia y generalización de una norma social.

#### **Análisis de robustez: el efecto de la densidad social**

Los resultados mostrados quedarían en entredicho si estos dependieran de características secundarias del modelo, lo que vendría a sig-

nificar que el proceso propuesto en el diagrama 1 es más el reflejo de una selección particular de parámetros que de un mecanismo con más amplio alcance. En concreto los resultados podrían ser dependientes de tres rasgos: la distribución de los recursos, la definición del «entorno» de los agentes y el número de agentes introducidos en la simulación.

Con relación a la distribución de recursos, debe recordarse que el modelo fue deliberadamente creado para producir un entorno tipo «tragedia de los comunes» (otro tipo de entorno, en el que los agentes no tuvieran la necesidad de cooperar mutuamente para sobrevivir, carecería de interés analítico). Este entorno puede simularse de muchas formas distintas, habiéndose optado por un mecanismo que genera una distribución de recursos por parcela distinta en cada simulación,

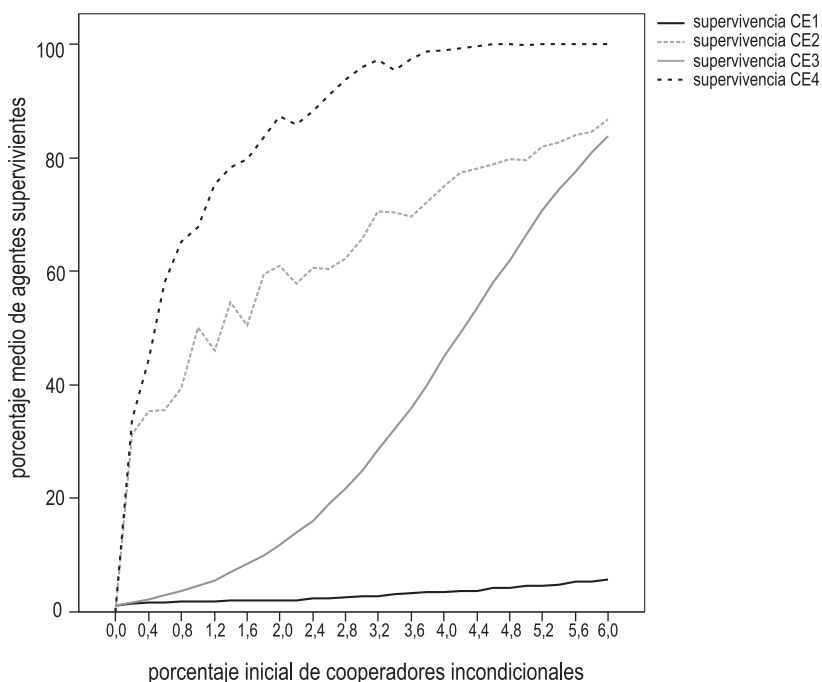
de manera que los resultados globales no dependan de ninguna distribución espacial concreta.

Los resultados, sin embargo, sí dependen de forma bastante obvia de la definición del entorno de cada agente como todos los agentes comprendidos en un determinado radio (véase la nota 8). Esto es así porque un radio mayor que el establecido implicaría que prácticamente todos los agentes tuvieran el mismo entorno (es decir la población total de agentes) y sus expectativas por tanto no variarían. Por otra parte, un radio menor implicaría que, en una población de 100 agentes desplazándose aleatoriamente, algunos de ellos tendrían un entorno muy limitado, de forma que su conducta se vería poco afectada por el comportamiento de otros agentes. Así, una reducción sensible de este parámetro implicaría, de manera bastante obvia, que la reo-

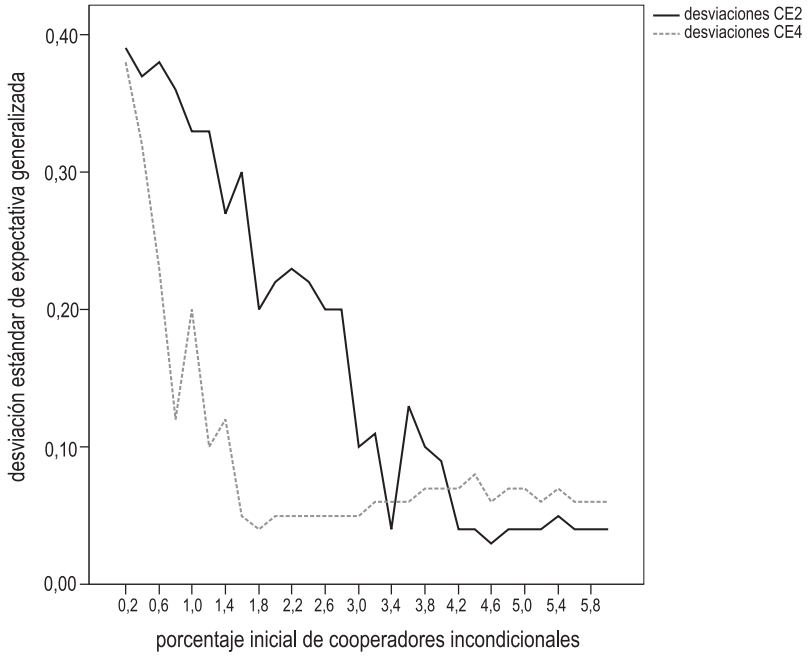
troalimentación entre sanciones y expectativas no tendría lugar. Además, en la medida en la que el comportamiento de los agentes respondiera poco a las características de su entorno, podría igualmente dudarse de que el proceso simulado es el de la generalización de una norma social.

El efecto menos obvio, sin embargo, es el del número de agentes que constituyen la población. Parece evidente, por una parte, que una reducción del número de agentes implicaría que la generalización de la conducta cooperativa se hiciera más difícil, dado que los agentes se encontrarían solo ocasionalmente. Por otra parte, el incremento del número de agentes debería hacer igualmente difícil la expansión de la cooperación, aunque por una razón distinta: a mayor número de cooperadores condicionales, los cooperadores incondicionales habrán de aumen-

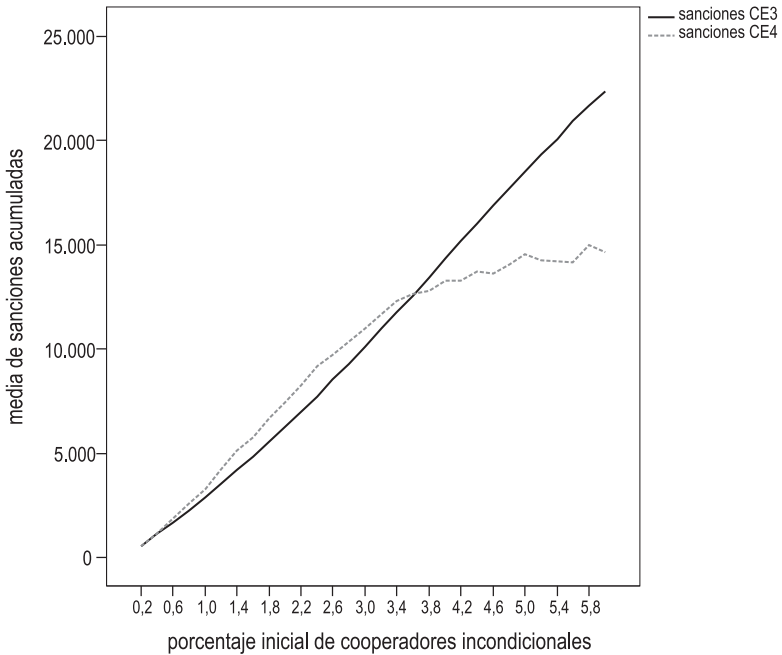
**GRÁFICO 5.** Población superviviente (500 agentes)



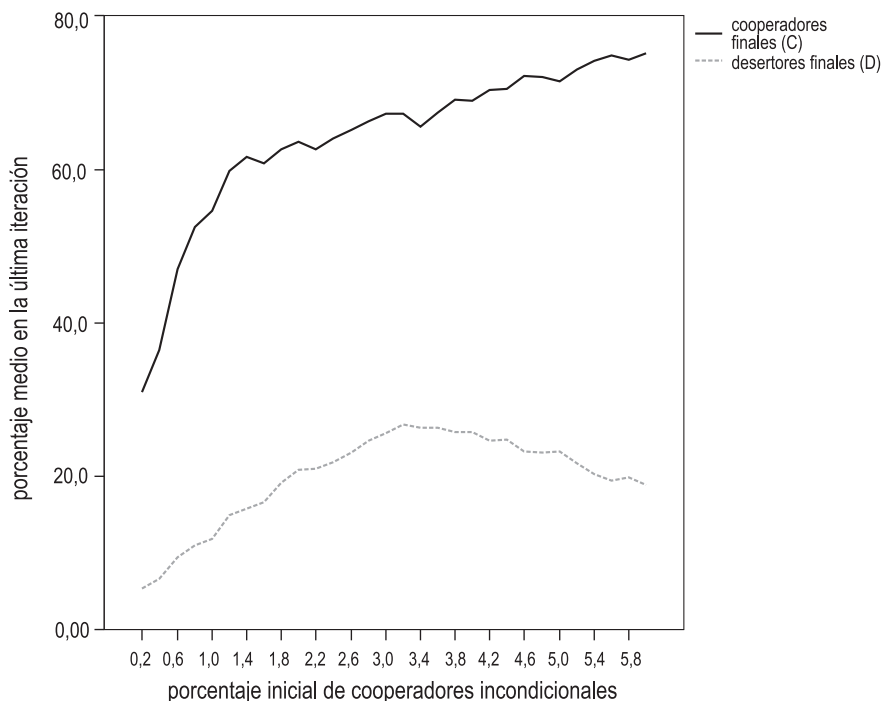
**GRÁFICO 6.** Expectativa generalizada (500 agentes)



**GRÁFICO 7.** Sanciones acumuladas (500 agentes)





**GRÁFICO 8.** *Porcentajes finales de tipos de agentes en CE4 (500 agentes)*

tarse proporcionalmente para controlar la conducta de los primeros y evitar una rápida sobreexplotación de los recursos.

Para testar esta hipótesis se replicaron los experimentos en las cuatro condiciones experimentales aumentando la población de 100 a 500 agentes. Esto implica un incremento notable de la densidad de población, dado que el espacio en el que se desplazan e interactúan los agentes sigue siendo el mismo. Obviamente también implica un incremento de la presión sobre el entorno, dado que los recursos disponibles también se mantienen constantes. Simultáneamente, se redujo el entorno de los sujetos del radio 4 al radio 1, con el objeto de generar unas condiciones especialmente difíciles para la expansión de la cooperación. Los resultados que se muestran en los gráficos 5, 6, 7 y 8 muestran, sin embargo, un efecto contrario a la intuición, y demuestran que las conclusiones

obtenidas del modelo original son bastante robustas con respecto a estos cambios en los parámetros de iniciación del modelo.

Con los nuevos parámetros se verifica que la tasa de supervivencia es mayor en la CE4 que en la CE3 y en la CE2, y en estas últimas mayor que en CE1 (gráfico 5). También se verifica que la variación de las expectativas en CE4 es menor que en CE2 (y tanto menor cuanto mayor es el número de cooperadores incondicionales introducidos en la simulación); e igualmente que el total de sanciones acumuladas es menor en CE4 que en CE3, a partir de un determinado número de cooperadores incondicionales introducidos en la simulación (gráficos 6 y 7). Finalmente se verifica que, en CE4, el número de cooperadores condicionales que cooperaron en la última iteración es mayor que el número de los que desertaron en la última iteración (gráfico 8).

Existe, no obstante, una diferencia muy significativa entre las simulaciones realizadas con 100 agentes y las realizadas con 500 agentes. En contra de lo que podría conjeturarse, en el segundo caso no son necesarios más, sino menos cooperadores incondicionales para garantizar la expansión de la cooperación y, por ende, la supervivencia de la población. A título de ejemplo, solo 18 agentes de este tipo son suficientes para poner en marcha el proceso de reotroalimentación entre sanciones y expectativas, y garantizar la supervivencia del 90 por ciento de la población. En términos relativos, esta cantidad de cooperadores incondicionales apenas supone el 4 por ciento de la población total.

### Discusión teórica

Tanto la investigación experimental como la investigación empírica de estudios de caso ofrecen evidencia concluyente de que los individuos son frecuentemente capaces de cooperar de forma sostenida en dilemas sociales (Fehr y Gächter, 2000; Ostrom, 2005). Uno de los requisitos que suele considerarse necesario para que esto ocurra es la posibilidad de sancionar. La evidencia sobre instituciones de gestión de recursos comunes muestra que un rasgo característico de estas, cuando es posible un mecanismo de vigilancia descentralizado, es que habitualmente las tasas de sanción son bajas (Ostrom, 1990). Esto, lejos de ser obvio, requiere una explicación. La tesis que sugieren los resultados de este artículo es que, cuando los individuos tienen incertidumbre sobre qué se espera de ellos, y la conducta es por tanto variable, un sistema sancionador confirma la existencia de una expectativa generalizada, propiciando una mayor regularidad en la cooperación. Esto, a su vez, disminuye la frecuencia con la que se ejecutan sanciones.

El acto de sancionar es, no obstante, un acto costoso; y la evidencia experimental muestra también que a mayor coste, menor

tasa de sanción (Horne y Cutlip, 2002). Sin embargo, a través de los experimentos sobre bienes públicos, también es conocido que hay una importante proporción de sujetos que está dispuesta a sancionar incondicionalmente (Fischbacher *et al.*, 2001). Una cuestión relevante es, por tanto, cuántos sujetos con este tipo de disposición motivacional son necesarios para desencadenar un efecto de bola de nieve que expanda la cooperación. La pregunta es análoga a la planteada por Marwell y Oliver (1993) en su conocida «teoría de la masa crítica». Los resultados de nuestra simulación muestran que, en un mundo en el que los cooperadores condicionales se abstuvieran de ejecutar sanciones, un porcentaje minoritario de cooperadores incondicionales dispuestos a sancionar podría expandir la cooperación. El volumen concreto de esta masa crítica obtenido en la simulación de 100 agentes, entre el 22 y el 37 por ciento de la población total, aunque dependiente de los parámetros de iniciación, resulta compatible con la evidencia acumulada experimentalmente, que suele cifrarse en un porcentaje que oscila entre un 25 y un 40 por ciento de la población (Fischbacher *et al.*, 2001).

El aumento de la densidad de población tiene el efecto de reducir sustancialmente el tamaño de tal masa crítica hasta el 4 por ciento. Este resultado no es evidente, dado que cabría esperar que el incremento de los cooperadores condicionales requiriera un aumento proporcional (y no una disminución) de cooperadores incondicionales para contrarrestar este efecto. No obstante, como sugería Coleman (1990a), en condiciones en las que existe una alta densidad de relaciones sociales, el comportamiento sancionador tiene un claro efecto multiplicador en la tasa de cooperación, debido a la posibilidad de compartir el coste de las sanciones. Nuestro modelo aporta un mecanismo complementario para este efecto multiplicador de la densidad de relaciones sociales: el contagio del comportamiento cooperativo que se hace

factible cuando cada individuo está rodeado de muchos otros y los sujetos son sensibles a las expectativas de los demás.

Así, un solo cooperador incondicional puede alterar significativamente el entorno de varios cooperadores condicionales (incluso cuando el «entorno» se define de una manera restrictiva) y estos, a su vez, los de otros cooperadores condicionales, en un proceso que adopta una lógica exponencial. En un trabajo anterior, Jeffrey Carpenter (2007) también muestra que el tamaño del grupo no es un obstáculo para la provisión eficiente de un bien público. Sus resultados, no obstante, difieren de los nuestros en su conclusión de que un menor grado de monitorización tiene el efecto de reducir la provisión del bien público. Sin embargo, en el modelo de Carpenter los agentes tipo «free riders» no están sujetos a la influencia de otros en su entorno (aunque podrían ser eliminados de la simulación, a través de una dinámica evolutiva) y, por tanto, nunca cambian de comportamiento.

Cabe señalar que, desde el punto de vista teórico, la cooperación en un dilema social podría explicarse a través de distintos mecanismos, que van desde la conducta guiada por la regla de la reciprocidad (Axelrod, 1984) hasta los mecanismos de sanción establecidos de forma exógena (Bravo, 2011). Pero el problema del que se ocupa este artículo es explicar cómo funciona el mecanismo que denominamos «norma social», que tiene características muy específicas: la regularidad en la conducta debe ser fruto de un proceso de sanción descentralizado y, simultáneamente, de un proceso de homogeneización de expectativas. En nuestra simulación se muestra cómo la existencia de expectativas sobre la conducta de los demás explica una baja tasa de sanciones, y la existencia de sanciones explica un alto grado de generalización de las expectativas. Sanciones y expectativas, actuando recíprocamente una sobre la otra, son así el haz y el envés de una misma realidad: las normas sociales.

Para finalizar, pueden sugerirse posibles extensiones del modelo presentado en este artículo que permitirían explorar más ampliamente el grado de robustez de sus conclusiones. Estas extensiones tienen que ver con el diseño de los agentes y con el diseño del entorno. En primer lugar, con relación a la arquitectura de los agentes, cabría establecer reglas de actuación más complejas. Así, por ejemplo, en un reciente trabajo, Fischbacher y Gächter (2010) modelan la formación de creencias de un agente sobre las contribuciones de los otros agentes como un proceso que se ajusta a las observaciones del comportamiento de los demás en rondas anteriores. Esto da a las creencias de sus agentes una estabilidad de la que carecen en nuestro modelo, en el que, en cada iteración, cada agente es sensible exclusivamente al comportamiento de los agentes en su entorno en esa misma iteración. Por otra parte, también podría dotarse a los agentes de capacidad para reconocer, y recordar, con qué agentes están interactuando en su entorno y, de acuerdo con reciente evidencia experimental (Gächter y Thöni, 2011), ajustar así su comportamiento.

En segundo lugar, con relación al entorno, dado que los experimentos de laboratorio ofrecen una tasa más o menos constante de «free riders» de aproximadamente un 25 por ciento de la población, una extensión inmediata del modelo es permitir que una parte de los agentes deserte incondicionalmente, es decir, con independencia del comportamiento de otros agentes en su entorno. Obviamente estos agentes nunca estarían sometidos a una norma social pero, dado que constituirían parte del entorno de otros agentes potencialmente sometidos, podrían impedir la generalización de la misma. Finalmente, también sería de interés comparar los resultados de un entorno en el que los agentes deambulan e interactúan libremente unos con otros, con otro tipo de entorno en el que los agentes estén incrustados en redes sociales con características topológicas espe-

cificas. Todas estas variaciones abren un rico campo para la investigación futura.

## BIBLIOGRAFÍA

- Axelrod, Robert (1984): *The Evolution of Cooperation*, Nueva York: Basic Books.
- (1986): «An Evolutionary Approach to Norms», *American Political Science Review*, 80: 1095-1111.
- (1997): «The Dissemination of Culture: A Model with Local Convergence and Global Polarization», *Journal of Conflict Resolution*, 41: 203-226.
- Bicchieri, Cristina (2006): *The Grammar of Society*, Cambridge: Cambridge University Press.
- Bravo, Giangiaco (2011): «Agent's Beliefs and the Evolution of Institutions for Common-Pool Resource Management», *Rationality and Society*, 23: 117-152.
- Carpenter, Jeffrey P. (2007): «Punishing Free-riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods», *Games and Economic Behavior*, 60: 31-51.
- Chew, M. (1999): «Structure and Strategy in Collective Action», *American Journal of Sociology*, 105: 128-156.
- Coleman, James S. (1990a): *Foundations of Social Theory*, Harvard, MA: Harvard University Press.
- (1990b): «Norm-generating Structures», en K. S. Cook y M. Levi (eds.), *The Limits of Rationality*, Chicago: The Chicago University Press.
- Elster, Jon (2007): *Explaining Social Behavior*, Cambridge: Cambridge University Press.
- (2009): «Norms», en P. Hedström y P. Bearman (eds.), *The Oxford Handbook of Analytical Sociology*, Oxford: Oxford University Press.
- Fehr, Ernst, Urs Fischbacher y Simon Gächter (2002): «Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms», *Human Nature*, 13: 1-25.
- y Simon Gächter (2000): «Cooperation and Punishment in Public Goods Experiments», *American Economic Review*, 90: 980-994.
- Fischbacher, Urs y Simon Gächter (2010): «Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments», *American Economic Review*, 100: 541-556.
- , — y Ernst Fehr (2001): «Are People Conditionally Cooperative? Evidence from Public Good Experiments», *Economics Letters*, 71: 397-404.
- Francisco, A. de (2010): «Feeling Cooperation», III Conference on Analytical Sociology and Social Mechanisms, Barcelona, 7-8 junio.
- Gächter, Simon y Christian Thöni (2011): «Micromotives, Microstructure, and Macrobehavior: The Case of Voluntary Cooperation», *Journal of Mathematical Sociology*, 35: 26-65.
- García-Valdecasas, José I. (2010): «La simulación basada en agentes: una nueva forma de explorar los fenómenos sociales», *Revista Española de Investigaciones Sociológicas*, 136: 91-110.
- Gilbert, Nigel y Klaus Troitzsch (2006): *Simulación para las Ciencias Sociales*, Madrid: McGraw Hill.
- Gintis, Herbert (2000): «Strong Reciprocity and Human Sociality», *Journal of Theoretical Biology*, 206: 169-179.
- González Bailón, Sandra (2004): «¿Sociedades Artificiales? Una Introducción a la Simulación Social», *Revista Internacional de Sociología*, 39: 199-222.
- Gould, R. (1993): «Collective Action and Network Structure», *American Sociological Review*, 58: 182-196.
- Groeber, Patrick, F. Schweitzer y K. Press (2009): «How Groups Can Foster Consensus: The Case of Local Cultures», *Journal of Artificial Societies and Social Simulation*, 12 (2), (en línea). <http://jasss.soc.surrey.ac.uk/12/2/4.html>, acceso 1 de diciembre de 2011.
- Hardin, Garrett (1968): «The Tragedy of Commons», *Science*, 162: 1243-1248.
- Hechter, Michael y Karl-Dieter Opp (eds.) (2001): *Social Norms*, Nueva York: Russell Sage Foundation.
- Heckathorn, Douglas (1988): «Collective Sanctions and the Emergence of Prisoners' Dilemma Norms», *American Journal of Sociology*, 94: 535-562.
- (1989): «Collective Action and the Second Order Free Rider Problem», *Rationality and Society*, 1: 78-100.
- (1990): «Collective Sanctions and Compliance Norms: A Formal Theory of Group-mediated Social Control», *American Sociological Review*, 55: 366-384.
- (1996): «The Dynamics and Dilemmas of Collective Action», *American Sociological Review*, 61: 250-277.

- Horne, Cristine (2001): «Sociological Perspectives on the Emergence of Social Norms», en M. Hechter y K. D. Opp (eds.), *Social Norms*, Nueva York: Russell Sage Foundation.
- (2007): «Explaining Norm Enforcement», *Rationality and Society*, 19: 139-170.
- y A. Cutlip (2002): «Sanctioning Costs and Norm Enforcement», *Rationality and Society*, 14: 285-307.
- Jansen, Marco A. y E. Ostrom (2006): «Governing Social-Ecological Systems», en L. Tesfatsion y K. L. Judd (eds.), *Handbook of Computational Economics*, vol. 2, Amsterdam: North Holland.
- Kim, Hyojoung y Peter S. Bearman (1997): «The Structure and Dynamics of Movement Participation», *American Sociological Review*, 62: 70-93.
- Kitts, James A. (2006a): «Social Influence and the Emergence of Norms Amid Ties of Amity and Enmity», *Simulation Modelling Practice and Theory*, 14: 407-422.
- (2006b): «Collective Action, Rival Incentives and the Emergence of Antisocial Norms», *American Sociological Review*, 71: 235-259.
- León, Francisco J. (2011): «Peer Loyalty and Quota Restriction as Social Norms: A Case Study of Their Emergence», *Rationality and Society*, 23: 75-115.
- Linares Martínez, Francisco (2007): «El problema de la emergencia de las normas sociales en la acción colectiva», *Revista Internacional de Sociología*, 46: 131-160.
- Macy, Michael W. (1993): «Backward Looking Social Control», *American Sociological Review*, 58: 819-836.
- Marwell, Gerald, Pamela Oliver y R. Prael (1988): «Social Networks and Collective Action: A Theory of Critical Mass III», *American Journal of Sociology*, 94: 502-534.
- y Pamela Oliver (1993): *The Critical Mass in Collective Action*, Cambridge: Cambridge University Press.
- Opp, Karl-Dieter (2001): «How Do Norms Emerge? An Outline of a Theory», *Mind and Society*, 3: 101-128.
- Ostrom, Elinor (1990): *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
- (2005): *Understanding Institutional Diversity*, Princeton, NY: Princeton University Press.
- Tena, Jordi y Ariadna Güell (2011): «¿Qué es una norma social? Una discusión de tres aproximaciones analíticas», *Revista Internacional de Sociología*, 69: 561-583.
- Ullmann-Margalit, Edna (1977): *The Emergence of Norms*, Oxford: Oxford University Press.

**RECEPCIÓN:** 15/12/2010

**REVISIÓN:** 19/04/2011

**APROBACIÓN:** 1/12/2011

