TEORÍAS DE LA EVOLUCIÓN DEL COMPORTAMIENTO COOPERATIVO: UNA REVISIÓN CRÍTICA

RICARDO A. GUZMÁN[*]
ROBERT E. ROWTHORN[**]
CARLOS RODRÍGUEZ-SICKERT[***]

ABSTRACT

In this paper we survey the theories of gratuitous cooperation, i.e., in favour of non-relatives and without repeated interaction. We also describe our work on the area, whose objective is to integrate the various theories of gratuitous cooperation into a self-contained framework. Our conclusions are as follows. First: altruistic punishment, conformism, and gratuitous cooperation coevolve, and group selection is a necessary for the coevolution to take place. Second: people do not cooperate by mistake, as most theories imply. On the contrary, people knowingly sacrifice themselves for others. Third: in cooperative dilemmas conformism is an expression of preference, not a learning rule. Fourth: group mutations are necessary to sustain cooperation in the long-run.

Keywords: Gratuitous cooperation, cooperative dilemma, altruistic punishment, conformism, group selection, group mutations.
JEL classification: H41, D64, D83, D03

RESUMEN

En este artículo revisamos las teorías de la cooperación gratuita, esto es, entre personas no emparentadas entre sí, y en ausencia de interacción repetida. También describimos nuestro propio trabajo en el área, cuyo objetivo es integrar diversas teorías de la cooperación gratuita en un marco autocontenido. Nuestras conclusiones son las siguientes. Primero: el castigo altruista, el conformismo, y la cooperación gratuita coevolucionan; y la selección grupal es necesaria para que la coevolución tenga lugar. Segundo: las personas no cooperan por error, como supone implícitamente la mayor parte de las teorías. Por el contrario, las personas se sacrifican conscientemente por los demás. Tercero: en los dilemas de cooperación el conformismo es expresión de una preferencia, no es una regla de aprendizaje. Cuarto: mutaciones grupales son necesarias para sostener la cooperación en el largo plazo.

Palabras clave: Cooperación gratuita, dilema de la cooperación, castigo altruista, conformismo, selección grupal, mutaciones grupales.

[*]Escuela de Administración, Pontificia Universidad Católica de Chile. Autor corresponsal
[**] Faculty of Economics, University of Cambridge and King's College, Cambridge, UK
[***] Escuela de Administración, Pontificia Universidad Católica de Chile

IN PRINT

Una característica del animal humano es su propensión a sacrificarse por personas que no son sus parientes y que tampoco devolverán el favor (Gintis, Bowles, Boyd y Fehr. 2003). Este fenómeno, al que denominamos *cooperación gratuita*, escapa a explicaciones evolutivas tradicionales. La literatura reciente intenta explicar la cooperación gratuita usando herramientas de varias ciencias de la conducta: biología, antropología cultural, psicología evolucionista y economía. En este artículo revisamos las diversas explicaciones que se postulan en la literatura, y describimos nuestro propio trabajo, cuyo objetivo es integrar dichas explicaciones en un marco coherente y autocontenido.

Las principales conclusiones de nuestro trabajo son las siguientes.

Primero: el *conformismo* (la tendencia a imitar conductas frecuentes), el *castigo altruista* (la tendencia a castigar a los *free riders* aunque hacerlo sea costoso para quien castiga) y la cooperación gratuita coevolucionan. La *selección grupal* (esto es, la selección natural de grupos que compiten entre sí) es una condición necesaria para que la coevolución tenga lugar.

Segundo: la cooperación gratuita no es resultado de un impedimento cognitivo; esto es, las personas no cooperan por error. El supuesto de impedimento cognitivo, predominante en la literatura de la cooperación gratuita, no sólo es inconsistente con las teorías modernas de la evolución de la inteligencia: es además un supuesto innecesario para explicar la cooperación gratuita.

Tercero: en los dilemas de cooperación, el conformismo debe entenderse como una preferencia aprendida, inculcada por los padres a sus hijos. En palabras sencillas: las personas son inteligentes; si se conforman, lo hacen porque les gusta seguir a la mayoría.

Cuarto: al interior de los grupos, la cooperación sigue un ciclo de auge repentino, prosperidad prolongada y decadencia abrupta. Para que la cooperación sea estable, *mutaciones grupales* deben ocurrir esporádicamente, por ejemplo el surgimiento de un líder carismático que trae orden al grupo. Nuestras simulaciones muestran que incluso si las mutaciones grupales son muy infrecuentes (cada centenares de años), una cultura de conformismo, castigo altruista y cooperación gratuita puede persistir en el largo plazo.

I. EL DILEMA DE CHERNOBYL

En esta sección presentamos un ejemplo real de cooperación gratuita a gran escala: el caso de los liquidadores de Chernobyl. El propósito de este ejemplo es persuadir al lector de que la cooperación es en efecto gratuita y deliberada; no, como muchos autores argumentan, el resultado de una confusión. Es también nuestro propósito argumentar la relevancia empírica y social de la cooperación gratuita.

En 1986, 600 mil liquidadores —bomberos, soldados, trabajadores, personal médico y muchos voluntarios— ingresaron a la central nuclear de Chernobyl, pocos días después de que la central explotara. Exponiéndose a la radiación mortal, los liquidadores extinguieron el incendio, limpiaron el área de los residuos radioactivos y construyeron el sarcófago de concreto que hasta el presente sella el reactor dañado. De los 40 bomberos que fueron los primeros en lidiar con el desastre, la mayoría murió durante los tres primeros meses. De los restantes, ninguno sobrevive hoy. Alrededor de 60 mil liquidadores han muerto desde 1986. Entre los sobrevivientes, unos 165 mil sufren de discapacidad y enfermedades crónicas.

IN PRINT

Cuadro I
El dilema de Chernobyl

El cuadro muestra los pagos materiales del liquidador en cada posible escenario. Si todos los demás liquidadores se quedan a ayudar, la familia del liquidador se salva: éste obtiene un beneficio de $B > 0$. El costo para el liquidador de quedarse a ayudar es $C > 0$.

	Todos los demás liquidadores se quedan a ayudar	Todos los demás liquidadores huyen
El liquidador se queda a ayudar	B – C	–C
El liquidador huye	B	0

El Cuadro I muestra el dilema que el liquidador enfrenta. Si todos los demás liquidadores se quedan a ayudar en la contención del desastre, la familia del liquidador se salvará: es mejor para él huir del lugar con su familia que exponerse a la radiación ($B > B - C$). Si los demás liquidadores huyen del lugar, la familia del liquidador de todas maneras se verá afectada por la radiación; no hay nada que el liquidador pueda hacer para evitarlo. Y si huye junto con los demás liquidadores, el liquidador y su familia tendrán una mayor probabilidad de sobrevivir ($0 > -C$). Por lo tanto, un liquidador egoísta concluirá que es mejor para él huir con su familia independiente de lo que hagan los demás. Entonces, si todos los liquidadores son egoístas, huirán del lugar y nadie ayudará a contener el desastre. Y, sin embargo, se quedaron, previniendo una catástrofe de consecuencias inimaginables.

El heroísmo de los liquidadores es difícil de explicar desde un punto de vista evolutivo. ¿Por qué querría alguien suicidarse, esto es, removerse a sí mismo del *pool* genético, por el bien de la humanidad? Un gen que nos predispusiese a un acto de esa naturaleza debiera extinguirse rápidamente. La *selección de parentesco* (en inglés, *kin selection*) no puede ser la solución del acertijo.[1] Los liquidadores habrían servido mejor a sus familias escapando con ellas. El *altruismo recíproco*,[2] esto es, el intercambio de favores a lo largo del tiempo, tampoco es la solución. Cuando se ha de morir antes de tres meses, no se cuenta con mucho tiempo para intercambiar favores. La explicación debe ser otra.

El dilema de Chernobyl es un ejemplo de lo que se conoce como *dilema de cooperación*, una situación en la que una persona debe decidir si subordinar o no su propio interés al interés del grupo. Y el comportamiento de los liquidadores representa un caso extremo de cooperación gratuita: el sacrificio de una persona por otras con las que no tiene parentesco genético, y que además tendrán poca o ninguna oportunidad de devolver el favor.

Tres temas surgen una y otra vez en la literatura que se ocupa de la cooperación gratuita: selección grupal, castigo altruista, y conformismo.

[1] La selección de parentesco y el principio relacionado de *adecuación inclusiva* (en inglés, *inclusive fitness*) fueron propuestos por Hamilton (1964).

[2] El altruismo recíproco no es altruismo realmente, sino egoísmo de largo plazo. La noción se debe a Trivers (1971).

IN PRINT

II. LA SELECCIÓN GRUPAL FAVORECE A LAS BANDAS COOPERATIVAS

La hipótesis de selección grupal como justificación del comportamiento cooperativo fue formulada por el propio Charles Darwin. En el *Origen del Hombre*, Darwin especula acerca del desarrollo de facultades morales en los humanos:

> "No debe olvidarse que, a pesar de que un alto estándar de moralidad confiere ligera o ninguna ventaja a cada hombre en particular y a sus hijos sobre otros hombres de su misma tribu, aún así un aumento del número de hombres prósperos y el avance de la moralidad conferirá ciertamente una inmensa ventaja a una tribu sobre otra. Una tribu que incluye muchos miembros que, al poseer en alto grado el espíritu del patriotismo, la fidelidad, la obediencia y la compasión, estarán siempre dispuestos a ayudarse unos a otros y a sacrificarse a sí mismos por el bien común, triunfará sobre la mayoría de las otras tribus; y esto sería selección natural." (Darwin ,1871, p. 166; la traducción es de los autores.)

En resumen, las tribus compiten unas con otras por la supervivencia, y aquellas conformadas por personas cooperativas llevan a la extinción a las tribus conformadas por personas egoístas.

Los humanos modernos aparecieron sobre la tierra hace unos 200 mil años. Durante los primeros 190 mil vivieron de la caza y la recolección. Cada mañana, los hombres se separaban en pequeñas partidas y salían a cazar. Al retornar al campamento, compartían la carne entre todos los miembros de la banda. Dado que el éxito en la caza dependía en gran medida de la suerte, compartir el alimento era una forma de seguro cruzado que garantizaba la disponibilidad del alimento (casi) todos los días (Gurven, 2004).

El comportamiento de nuestros ancestros sólo puede ser descrito como cooperación gratuita (Bohem, 1999). La caza es costosa en términos reproductivos: consume miles de calorías que son preciosas cuando se vive al borde de la inanición. Un cazador egoísta siempre puede culpar a la suerte si retorna al campamento con las manos vacías. Pelear también es costoso porque involucra riesgo de muerte o lesiones severas. No obstante, nuestros ancestros cumplían con su parte en la caza y arriesgaban sus vidas peleando por la banda. La hipótesis de selección grupal identifica a la intensa competencia entre bandas como explicación al comportamiento de nuestros ancestros (Bowles, Choi &y Hopfensitz, 2003, Boyd y Richerson, 1985; Cavalli-Sforza y Feldman, 1981; Sober y Wilson, 1994).

La historia anterior suena razonable excepto por una cosa. No es claro qué debe predominar: la selección individual en favor de los free riders (esto es, en favor de aquéllos que no cooperan con la banda), o la selección grupal en favor de las bandas cooperativas. La dirección en la que esta tensión se resuelve depende, entre otras cosas, de la frecuencia de los conflictos inter-grupales y de la tasa de migración entre bandas. Se ha argumentado que la tasa migratoria entre los humanos tempranos era demasiado alta y la frecuencia del conflicto demasiado baja para que la selección grupal pudiera operar (Williams, 1966). Si suponemos que esto es verdad, nos vemos obligados a descartar la selección grupal como una explicación suficiente para la existencia de cooperación gratuita.

III. LOS CASTIGADORES ALTRUISTAS PUEDEN FORZAR LA COOPERACIÓN, PERO SÓLO TEMPORALMENTE

Una segunda explicación para la cooperación gratuita se basa en el llamado castigo altruista (Fehr y Gächter, 2000, 2002; Ostrom, Walker y Gardner, 1992). En la mayoría de las sociedades

IN PRINT

humanas, ser un free rider no es realmente gratis. El free rider se expone a la sanción social: la desaprobación, el ostracismo e incluso la agresión física. Si el número de castigadores es lo suficientemente alto, es posible que ser un free rider resulte más costoso que cooperar. Si eso es así, los free riders desaparecerán de la población o serán muy escasos.

Castigar a los free riders también es costoso, siendo el costo principal el riesgo de que los free riders se defiendan o tomen venganza. Por esta razón, el castigo a los free riders se denomina altruista. Si los free riders son pocos, el costo de ser un castigador altruista es pequeño, ya que rara vez será necesario castigar. Aun así, ese pequeño costo pondrá a los cooperadores en una desventaja reproductiva respecto de quienes cooperan pero no castigan: los denominados *cooperadores puros*. El resultado será que la proporción de cooperadores puros aumentará a expensas de los castigadores altruistas.

A medida que el número de castigadores disminuye, la intensidad del castigo se reduce y ser un free rider resulta cada vez menos costoso. Llegará un momento en el que el número de castigadores será tan bajo que el costo de ser castigado será menor que el beneficio de no cooperar. Cuando eso ocurra, los free riders obtendrán un beneficio material mayor que los castigadores y los cooperadores puros. Por lo tanto, el número de free riders en el grupo aumentará de manera progresiva, hasta desplazar por completo a los demás tipos de personas.

La Figura 1 ilustra la secuencia recién descrita.

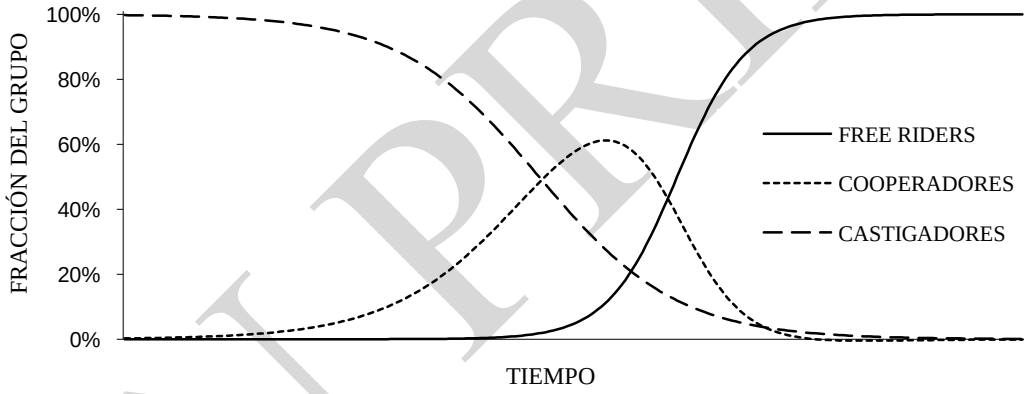

Figura 1. Los free riders se apoderan del grupo. Los primeros cooperadores puros aparecen en el instante T0. A partir del instante T1, el número de castigadores es tan bajo que el costo de castigar es menor que el beneficio de no cooperar. Desde ese momento en adelante, los free riders aumentan en número hasta desplazar a todos los demás tipos de personas.

IV. EL CASTIGO ALTRUISTA SUMADO A LA SELECCIÓN GRUPAL PUEDE EXPLICAR LA COOPERACIÓN GRATUITA EN GRUPOS PEQUEÑOS

Boyd, Gintis, Bowles y Richerson (2003) desarrollaron un modelo que combina castigo altruista con selección grupal. En dicho modelo, bandas de cazadores recolectores compiten unas con otras, mientras que al interior de cada banda tiene lugar una competencia entre tres tipos de personas: free riders, quienes nunca cooperan en la caza; cooperadores puros, quienes cooperan en la caza pero no castigan a los free riders; y castigadores, quienes cooperan en la caza y además castigan a los free riders. La competencia inter-grupal penaliza a las bandas que tienen muchos free riders. También

UNPRINT

confiere una ventaja a los grupos que tienen muchos castigadores, ya que estos grupos tenderán a tener pocos free riders. Usando parámetros calibrados con información paleontológica y etnográfica, el modelo es capaz de explicar la persistencia de la cooperación en grupos del tamaño de una banda (30 a 50 personas) hasta el tamaño de una tribu (unos pocos cientos).

Una falencia del modelo de Boyd et al. (2003) es que no logra explicar la cooperación en sociedades de gran tamaño. Esto se debe a la migración y a la Ley de los Grandes Números. La migración es como el flujo en un vaso comunicante que elimina las diferencias entre grupos. Y, mientras más homogéneos son los grupos más débil es el efecto de la selección grupal. Si dos grupos casi idénticos entran en conflicto, la selección grupal no tiene, en la práctica, nada que seleccionar.

En grupos pequeños, innovaciones aleatorias pueden contrarrestar el efecto homogeneizador de la migración. Por ejemplo, en un grupo formado por cuatro free riders, si uno de ellos se transforma en cooperador, la cooperación salta de 0% a 25%. En un mundo de grupos pequeños, entonces, la innovación genera diversidad constantemente, lo que asegura que la selección grupal siga operando.

En grupos más numerosos, la Ley de los Grandes Números atenúa la variabilidad inter-grupal generada por la innovación aleatoria. Supongamos que la probabilidad de que una persona se transforme en castigador es 1%. La Ley de los Grandes Números predice que en un grupo de 100.000 free riders aproximadamente 1.000 se transformarán en castigadores. Por lo tanto, la innovación generará muy poca diversidad entre grupos y ésta no será suficiente para contrarrestar el efecto homogeneizador de la migración. Con el tiempo, los grupos se volverán muy similares entre sí e igualmente fuertes. Por lo tanto, la selección grupal dejará de operar.

V. LOS CONFORMISTAS EVITAN QUE EL *FREE RIDING* SE DIFUNDA EN EL GRUPO, PERO, ¿DE DÓNDE VIENE EL CONFORMISMO?

La difusión del free riding es un proceso de evolución cultural. Buena parte del repertorio conductual humano se transmite a través de ideas llamadas *memes* (Dawkins, 1976). Los memes son el análogo cultural de los genes. Al igual que los genes, los memes mutan y están sujetos a un proceso de selección que elimina a aquellos memes que inducen conductas no adaptativas.

En un dilema de cooperación hay tres memes en competencia: no cooperar, cooperar sin castigar, y cooperar y castigar a los free riders. Supongamos que originalmente todo el mundo es castigador. Repentinamente, unas pocas personas innovan (la innovación es el equivalente memético de la mutación genética). Algunos innovadores se transforman en cooperadores puros y otros se transforman en free riders. Como los castigadores predominan en la población, el castigo a los free riders es feroz. Estos últimos rápidamente se dan cuenta de que les conviene imitar una conducta más exitosa, la que puede ser cooperar solamente, o bien cooperar y castigar (la imitación es el equivalente memético de la reproducción).

Explicaciones de la cooperación gratuita basadas exclusivamente en la selección grupal y/o el castigo altruista suponen que las personas tienden a maximizar sus pagos materiales. Tal supuesto no es en absoluto realista. Por el contrario, la mayor parte del tiempo las personas se limitan a adoptar el comportamiento de la mayoría. A dicha regla de aprendizaje se le denomina *aprendizaje conformista* o, simplemente, *conformismo* (Boyd y Richerson, 1985, Henrich y Boyd, 1998).

En un dilema de cooperación, el conformismo se traduce en la estrategia "cooperar si la mayoría coopera, no cooperar si la mayoría no coopera". Si la cooperación es la conducta mayoritaria, el equilibrio cooperativo es extremadamente estable. Un pequeño número de innovadores nunca inducirá a los conformistas a cambiar de estrategia.

IN PRINT

Más aún, mientras más grande sea un grupo de conformistas, más estable será el equilibrio cooperativo, justo lo contrario de lo que ocurre cuando la gente imita comportamientos exitosos. Los conformistas insistirán en cooperar a menos que más de la mitad del grupo no coopere. Y si el número de conformistas es grande, la Ley de los Grandes Números hace extremadamente improbable que una fracción sustancial de los conformistas simultánea e independientemente innove o deje de cooperar por accidente. Un incremento en el tamaño de los grupos ayuda a estabilizar la cooperación cuando las personas se conforman a las conductas mayoritarias.

Esta historia, al igual que las anteriores, no es del todo satisfactoria. Elimina la pregunta "¿cómo se explica la cooperación?", sólo para levantar otra: "¿cómo se explica el conformismo?"

VI. LA HIPÓTESIS DEL GEN POLIZÓN COMO EXPLICACIÓN DEL CONFORMISMO

Hay quienes argumentan que el conformismo evolucionó en el contexto de problemas de decisión que nada que tenían que ver con el comportamiento cooperativo. Desde ahí, el conformismo se coló, a la manera de un polizón, en los dilemas de cooperación. Esta explicación se debe a Simon (1990), y ha sido explorada en detalle por Gintis (2003) y Henrich (2004). Por ejemplo, en lugar de diseñar su propio arco y flecha, un cazador podría decidir imitar el diseño predominante en su tribu. Esta conducta aumenta el *fitness* del cazador porque reduce sus gastos en "investigación y desarrollo". Más tarde, el mismo cazador podría decidir cooperar en la caza, simplemente porque todos los demás cazadores cooperan. Esta conducta disminuye el fitness del cazador. Los partidarios de la hipótesis del gen polizón argumentan que todo lo que se necesita para que el conformismo evolucione es que aumente el fitness individual *en promedio*.

El conformismo economiza recursos al aprovechar el conocimiento acumulado por el grupo en las prácticas comunes y en las normas sociales, y por esta vía podría aumentar el fitness (Feldman, Aoki y Kumm, 1996; Henrich y Boyd, 1998). Para que el conformismo sea eficiente son necesarias tres condiciones. Primero, optimizar debe demandar gran cantidad de tiempo y recursos. Segundo, al menos una parte de la población debe buscar por sí misma la conducta óptima. De otro modo, el grupo no generará conocimiento nuevo y los conformistas no tendrán ideas de las cuales apropiarse. Tercero, el ambiente debe ser relativamente estable. Si no lo es, el conocimiento acumulado por el grupo se vuelve obsoleto rápidamente, en cuyo caso optimizar puede ser la opción más razonable.

En su exposición canónica de la hipótesis del gen polizón, Gintis (2003) postula la existencia de un "gen internalizador", el cual predispone a la gente a conformarse a las normas sociales. La probabilidad de que una persona portadora del gen internalizador se conforme a cierta norma depende del nivel de exposición de la persona a la norma durante algún período formativo. Por ejemplo, si ambos padres se conforman a determinada norma, entonces un niño que porta el gen internalizador también se conformará a esa norma. La probabilidad de que se conforme será 50% si sólo uno de los padres se conforma a la norma. Un niño que porta el gen internalizador será cazador en la edad adulta si es criado entre cazadores. Será altruista en la edad adulta si es criado entre altruistas. Será un cazador altruista si es criado entre cazadores altruistas, etc. El fitness de cada uno de estos tipos de persona será distinto. Para que el gen internalizador sobreviva, basta que las personas que lo portan alcancen, *en promedio*, a menos el mismo fitness que las personas que no lo portan. La pérdida en fitness que un portador experimenta por comportarse de manera altruista debe ser menor que la ganancia que el portador obtiene al seguir normas que aumentan el fitness. Gintis

IN PRINT

(2003) demuestra que, si el gen internalizador sobrevive, existirá (para ciertos parámetros) un equilibrio estable en el que la cooperación predomine.

El argumento anterior no es suficiente para explicar el conformismo a normas altruistas costosas. Se requiere algo más: o bien el dilema de cooperación es lo suficientemente complejo para justificar el uso del conformismo en lugar del aprendizaje individual, o bien los humanos no pueden distinguir las situaciones en las que el conformismo aumenta el fitness de las situaciones en las que lo disminuye. Ambos requisitos son poco plausibles. Los dilemas de cooperación suelen ser simples, la gente es consciente de su simplicidad, y cálculos sencillos llevan a la gente a concluir que la cooperación es una forma de autosacrificio. El que las personas elijan conformarse a la norma altruista nos fuerza a rechazar la hipótesis del gen polizón como explicación del conformismo. Debemos concluir que el conformismo en los dilemas de cooperación no es una regla de aprendizaje sino la expresión de una preferencia: asimilarse a la mayoría genera placer, y ser distinto es desagradable. A continuación elaboramos estas ideas.

Primero, los experimentos en los que la gente juega alguna versión estilizada del dilema de cooperación (tal como el dilema del prisionero o el juego de bien público) sugieren que la gente no coopera por error. La evidencia experimental revela que las personas comprenden muy bien la lógica del dilema y entienden que el free riding es la conducta óptima desde una perspectiva egoísta. Sin embargo, la mayoría de los sujetos elige conformarse a la mayoría: cooperan si otros cooperan, y no cooperan si los demás no cooperan. Dicho patrón de conducta ha sido denominado cooperación condicional (Fischbacher, Fehr y Gächter 2000 y 2002).

El dilema de cooperación es igualmente fácil de entender en la vida real que en el laboratorio. Consideremos el caso de Chernobyl. Sarah Wallace (2007) entrevistó a muchos hombres que se ofrecieron como voluntarios para ser liquidadores. A juzgar por sus respuestas, la conducta de esos hombres difícilmente puede atribuirse a una confusión.

Uno de los liquidadores entrevistados por Wallace era profesor física en Zhytomyr Oblast, y ayudó a evacuar a los residentes cercanos a la planta nuclear. Wallace reporta: "Él no cambiaría su decisión de ser un liquidador. Piensa que las personas deben ayudarse unas a otras. No considera que lo que hizo haya sido nada especial." Desde entonces, el profesor ha sufrido enfermedades cardiovasculares y neurológicas, las que atribuye a la radiación.

Otro liquidador era un obrero de la construcción que en 1986 trabajaba en Kiev. Dice Wallace: "A él y a sus colegas se les pidió que acudieran a Chernobyl después del desastre. Su trabajo fue sellar el piso del reactor para que la materia irradiada no escurriera hasta el agua subterránea. Accedió a ir a Chernobyl por patriotismo. En sus propias palabras, si él no lo hacía, ¿entonces quién?" Éste hombre ha sufrido problemas de salud desde 1986, probablemente debido a la dosis de 28.2 roentgens que recibió mientras trabajó en Chernobyl. Durante la entrevista, declaró que cuando se ofreció como voluntario estaba plenamente consciente de los riesgos y que, si pudiera volver a 1986, se ofrecería nuevamente.[3]

Podría conjeturarse que las personas entienden el dilema de cooperación, pero aun así deciden conformarse dado que no confían en su propio juicio. Esto es, fallan en distinguir las situaciones en la que el conformismo aumenta el fitness de las situaciones en las que lo disminuye. Obviamente, tal conjetura hace que uno se pregunte por qué la habilidad de distinguir ambos tipos de situaciones no evolucionó. El propio Gintis reconoce este agujero en la hipótesis del gen polizón (Gintis 2003, p. 416).

[3] Es posible que el profesor de física y el obrero justifiquen su conducta con el fin de reducir la disonancia cognitiva (Festinger 1957).

IN PRINT

Supongamos que se produce una mutación genética que permite a una persona distinguir las normas altruistas que aumentan el fitness de aquéllas que lo reducen. Esa mutación desplazará el gen del conformismo ciego. El individuo seguirá la norma social sólo cuando lo beneficia, y de otro modo la ignorará. En respuesta a esta objeción, Gintis (2003)sostiene que los humanos tienen una capacidad muy limitada para discriminar las normas que reducen fitness; e incluso si pudieran discriminarlas, la degradación de esa facultad es muchísimo más probable que su evolución.

Antes de evaluar la afirmación de Gintis, debemos notar que a los humanos rara vez les importan las consecuencias de sus acciones en términos de fitness. En cambio, y al igual que otros animales, los humanos basan sus decisiones en aproximaciones del fitness, tales como recompensas materiales, cantidad de esfuerzo y riesgo de muerte (en la prehistoria, cuando la naturaleza humana evolucionó, dichas variables estaban altamente correlacionadas con el fitness). Los humanos suelen tener un conocimiento bastante preciso de los probables costos y beneficios de sus acciones, y cuando los evalúan calculan implícitamente consecuencias de fitness. Una vez que reconocemos esto, resulta difícil aceptar que el conformismo en dilemas de cooperación sea resultado de un impedimento cognitivo. La gente entiende muy bien que la cooperación es una forma de autosacrificio. Aun así, la gente elije conformarse a normas altruistas. El único modo de hacer sentido de los hechos es concluir que el conformismo en dilemas de cooperación es expresión de una preferencia.

Una razón final para dudar de la hipótesis del gen polizón es que ésta es inconsistente con las teorías modernas de la evolución de la inteligencia. Según estas teorías, la inteligencia evolucionó precisamente para lidiar con la complejidad social (Dunbar, 1993). Dado que el cerebro nos asiste en la interacción social, ¿por qué debiéramos desactivarlo cuando participamos en dilemas de cooperación? Somos, de hecho, sorprendentemente hábiles en tareas relacionadas a la interacción social; por ejemplo, detectando la violación de normas sociales o llevando la cuenta de favores adeudados (Cosmides y Tooby, 1992). El conformismo como preferencia es enteramente consistente con un nivel elevado de inteligencia social.

VII. CONFORMISMO, COOPERACIÓN Y CASTIGO ALTRUISTA COEVOLUCIONAN; LA MEZCLA PUEDE EXPLICAR LA COOPERACIÓN GRATUITA EN GRANDES GRUPOS

Para explicar la evolución del conformismo en los dilemas de cooperación, desarrollamos un modelo en el cual las reglas de aprendizaje y las conductas evolucionan en paralelo (Guzmán, Rodríguez-Sickert y Rowthorn, 2007). Nuestro modelo (el cual es una extensión del modelo de Boyd et al. 2003), incluye dos reglas del aprendizaje que compiten entre sí: conformismo a la conducta más común e imitación dependiente de la ganancia (la imitación de conductas exitosas; en inglés, *payoff-dependent imitation*). Cada persona sigue una y sólo una regla de aprendizaje durante su vida, y usa esa regla de aprendizaje para elegir entre tres conductas: no cooperar, cooperar sin castigar, y cooperar y castigar a los free riders. Los imitadores dependientes de la ganancia tienen una ventaja reproductiva sobre los conformistas, ya que los imitadores dependientes de la ganancia buscan activamente conductas que maximizan fitness, mientras que a los conformistas no les importa en absoluto el fitness.

En nuestro modelo, las reglas de aprendizaje son genéticas y se heredan de padres a hijos. La herencia de reglas de aprendizaje no es perfecta: con una probabilidad muy pequeña, el hijo de un conformista se convertirá en un imitador dependiente de la ganancia, y el hijo de un imitador dependiente de la ganancia se transformará en un conformista.

IN PRINT

Siguiendo a Boyd et al. (2003), simulamos nuestro modelo para condiciones que se aproximan a aquéllas en las cuales vivieron los primeros humanos, y encontramos que la mezcla formada por el castigo altruista, el conformismo y la selección grupal es capaz de sostener la cooperación gratuita en grupos de miles de personas. Este resultado se obtiene incluso cuando la tasa de migración entre grupos es alta o cuando el conflicto entre grupos es infrecuente. En nuestro modelo, el castigo altruista, el conformismo y la cooperación gratuita coevolucionan: ninguno de estos fenómenos puede desarrollarse en ausencia de los otros.

¿Qué mecanismo subyace a estos resultados? Primero, la selección grupal favorece a los grupos cooperadores. Segundo, el conformismo combinado con el castigo altruista inmuniza al grupo contra la invasión de free riders, sea desde fuera (inmigración) o desde dentro (innovación). Si un inmigrante o innovador free rider es también un conformista, automáticamente cambiará de estrategia y se transformará en un castigador, ya que el castigo altruista es la conducta más frecuente en su grupo. Si el free rider es un imitador dependiente de la ganancia, imitará a la persona más exitosa que encuentre, la que probablemente sea un castigador en un grupo dominado por castigadores (son más exitosos que el free rider porque nunca son castigados y casi nunca necesitan castigar). Además, si los castigadores locales son conformistas, resistirán la tentación de dejar de castigar y convertirse en cooperadores puros cuando aparecen los primeros free riders.

VIII. ¿POR QUÉ IMITAN LOS HUMANOS?

La imitación es una forma de aprendizaje social ya que involucra el aprender de otros. La alternativa al aprendizaje social es el aprendizaje individual, el cual consiste en una evaluación racional de las opciones disponibles y requiere de un nivel más elevado de inteligencia. La inteligencia se relaciona con el tamaño del cerebro (Williams, 2002) y el cerebro es un órgano caro de desarrollar y de mantener. Los humanos dependen de sus padres durante más tiempo que cualquier otro animal terrestre, ya que sus cerebros son muy grandes en relación a su masa corporal. Tener que apoyar a sus crías durante tanto tiempo impone un enorme costo de fitness a los progenitores humanos. El cerebro es también caro en términos metabólicos; es, por lejos, el mayor consumidor de energía en el organismo humano (Aiello y Wheeler, 1995). Por consiguiente, la evolución del cerebro involucra un *trade-off* entre los beneficios de procesar mejor la información, el elevado costo metabólico del tejido cerebral y la enorme inversión parental en crías con cerebros grandes (Niven, 2005). El aprendizaje por imitación podría tener una ventaja selectiva sobre el aprendizaje individual, ya que la imitación reduce los costos de desarrollar y mantener el cerebro.

Si bien el ahorro en costos cerebrales es una explicación plausible para la evolución del aprendizaje por imitación, existe una explicación alternativa basada en la selección grupal. Para explorar esta posibilidad, agregamos a nuestro modelo una regla de aprendizaje adicional: *mejor respuesta* (Rodríguez-Sickert, Rowthorn y Guzmán, 2008). Los *best responders* (en español: quienes aplican la regla de mejor respuesta) son tanto egoístas como hiper-racionales, lo que significa que siempre emiten la conducta que genera la mayor ganancia en términos de fitness. El modelo supone que la regla de mejor respuesta no impone mayores costos de fitness que el conformismo. Por consiguiente, el tamaño del cerebro no juega un rol en el modelo. Cada individuo debe elegir una entre tres estrategias: no cooperar, cooperar sin castigar y cooperar y castigar a los free riders. Estas estrategias se difunden al grupo mediante las reglas de aprendizaje presentes. Como en el modelo anterior, las presiones selectivas intra-grupos e inter-grupales determinan el curso de la evolución.

IN PRINT

Para explorar las implicancias del modelo, lo simulamos usando los mismos parámetros que en Guzmán et al. (2007). Un patrón híbrido emergió de las simulaciones: en promedio, observamos un porcentaje elevado de conformistas y un porcentaje bajo de best responders. ¿Por qué? Los best responders aceleran la invasión del free riding, dado que los best responders nunca castigan a los free riders, y se convierten automáticamente en free riders cuando los castigadores son lo suficientemente escasos. En consecuencia, los best responders debilitan al grupo, por lo que la selección grupal tiende a eliminar a los grupos en los que predominan los best responders. El conformismo y la imitación dependiente de la ganancia prosperan ya que promueven el auto-sacrificio, clave del éxito en la competencia inter-grupal. Los resultados del modelo no cambian si aumenta la tasa de migración o disminuye la frecuencia de los conflictos inter-grupales.

IX. TRABAJO EN PROGRESO: LÍDERES CARISMÁTICOS Y LA EVOLUCIÓN DEL CONFORMISMO COMO PREFERENCIA

Los modelos de selección grupal que hemos descrito comparten una característica importante con la hipótesis del gen polizón: todos suponen que ciertos individuos (los imitadores dependientes de ganancia y los conformistas) tienen un impedimento cognitivo que en ocasiones los lleva a actuar en contra de su propio interés. En nuestra crítica a la hipótesis del gen polizón argumentamos que el supuesto de impedimento cognitivo no es realista. La conformidad a normas sociales costosas no es (en general) un subproducto de un impedimento cognitivo, sino el producto de una decisión deliberada. En los dilemas de cooperación, la conformidad a normas sociales debiera ser modelada como una preferencia y no como una regla de aprendizaje.

Nuestro trabajo actual explora las implicancias de este segundo enfoque mediante un modelo de selección grupal, en el cual los impedimentos cognitivos no juegan ningún rol. En el modelo nuevo hay dos tipos de individuos: *conformistas* y *egoístas*. Estos tipos son idénticos en términos de habilidad cognitiva pero difieren en sus preferencias. El objetivo del conformista es seguir la norma social predominante como un fin en sí mismo, sin contemplar la ganancia material. El objetivo del egoísta es maximizar su ganancia material. Ambos tipos tienen conocimiento completo y preciso acerca del comportamiento actual de los otros miembros del grupo, conocimiento que es adquirido sin costo. Las personas eligen su conducta para el período siguiente de acuerdo con sus propias preferencias: el conformista elige el comportamiento más frecuente; el egoísta elige el comportamiento que le reportaría la mayor ganancia dada la distribución actual de conductas entre los miembros de su grupo.[4] Para simplificar, suponemos que individuos son conformistas puros o egoístas puros, si bien los seres humanos tenemos motivaciones mixtas.

En el modelo nuevo, los niños heredan las preferencias de sus padres. En esto el modelo es similar a los modelos de selección grupal antes descritos. La diferencia es que en el modelo nuevo el mecanismo de transmisión de padres a hijos es cultural y no genético. Los individuos son genéticamente idénticos y sus preferencias les son inculcadas por sus padres durante la infancia. Suponer que la transmisión es cultural tiene dos implicancias claves. En primer lugar, el error de

[4] Los egoístas en el modelo nuevo son idénticos, en términos de habilidad cognitiva y preferencias, a los best responders en Rodriguez-Sickert et al. (2008). El objetivo del cambio en terminología es dejar en claro que en el modelo nuevo lo central son las preferencias. El término "*best responder*" insinúa que maximizar ganancias individuales es "mejor" que conformarse a la norma social. Lo cierto es que tanto conformistas como egoístas eligen, a la luz de sus propias preferencias, la mejor respuesta disponible.

UNPRINT

copia es más alto en la transmisión cultural que en la transmisión genética, por lo que el contagio del egoísmo será más rápido que el de una regla de aprendizaje. Por lo tanto, la cooperación será más frágil en este nuevo modelo que en los anteriores. En segundo lugar, las preferencias transmitidas culturalmente podrían ser alteradas por experiencias que se tienen durante la vida adulta. Esto abre la puerta a eventos de nivel grupal que simultáneamente transformen las preferencias y las conductas de un gran número de individuos del mismo grupo.

Para explorar las propiedades del modelo anterior hemos realizado algunas simulaciones exploratorias. Supusimos que inicialmente hay un único grupo cooperador cuyos miembros son en su mayoría conformistas-castigadores, y que el resto de los grupos se componen egoístas-free riders. Al entrar en conflicto con un grupo no cooperador, lo más probable es que el grupo de conformistas-castigadores se imponga y reemplace al otro grupo con un clon de sí mismo. En el transcurso del tiempo, mediante sucesivas conquistas, el número de grupos de conformistas-castigadores aumentará hasta que eventualmente todos los grupos sean de este tipo.

Cuando todos los grupos son igualmente cooperadores, la selección grupal se debilita y las dinámicas de selección intra-grupales predominan (recordemos que la selección grupal requiere de heterogeneidad inter-grupal para operar). Dentro de cada grupo, errores aleatorios producirán unos pocos egoístas quienes inicialmente cooperarán porque son minoría y desean evitar el castigo. Los egoístas cuentan con una leve ventaja de fitness sobre los conformistas, ya que los egoístas eluden el costo de castigar a los pocos free riders egoístas. Eventualmente, los free riders egoístas se convertirán en la mayoría, lo cual significa que el free riding será la nueva norma del grupo. En este punto, los conformistas cambiarán su conducta del castigo al free riding, y la cooperación se extinguirá por completo. Este proceso ocurrirá a un ritmo similar en todos los grupos, por lo que la cooperación desaparecerá en todos los grupos más o menos al mismo tiempo.

En este punto el sistema queda atrapado en un equilibrio no-cooperativo. La probabilidad de escapar de dicho equilibrio es un cero absoluto desde el punto de vista práctico: eones tendrían que pasar antes de que, por puro azar, suficientes miembros de un grupo no cooperador se transformen simultánea e independientemente en conformistas-castigadores. La ruta de escape alternativa es una *mutación grupal* que convierta a la mayoría de los miembros de algún grupo en conformistas-castigadores. Un ejemplo de mutación grupal es la emergencia de un líder carismático que impone el orden entre su gente y los moviliza a conquistar otros grupos. Si tal cosa sucede, el sistema nuevamente seguirá la secuencia antes descrita. El conformismo se propagará mediante la conquista hasta que la cooperación sea un fenómeno universal. A pesar de que dicha situación es muy estable, la aparición lenta pero progresiva de los egoístas amenazará con quebrar el equilibrio cooperativo. Pero antes de que eso ocurra un nuevo líder carismático emergerá en algún lugar. Y dado que en cada grupo la cooperación persiste (según nuestro modelo) centenares de años, el sistema se mantendrá por siempre en el equilibrio cooperativo incluso si la emergencia de un líder carismático es un evento muy poco probable. Es suficiente que cada grupo produzca un Genghis Kahn cada 500 años más o menos para prevenir la difusión del free riding y mantener las ruedas de la cooperación girando.

Entonces, el porcentaje de egoístas aumentará gradualmente dentro de cada grupo hasta que los castigadores se vuelvan tan escasos que el free riding se convierta en la conducta más rentable. En ese preciso instante los egoístas optaran por el free riding y la cooperación se reducirá drásticamente. Si los conformistas siguen siendo la mayoría, continuarán castigando por un tiempo, pero gradualmente serán reemplazados por free riders accidentales.

IN PRINT

REFERENCIAS

Aiello, L. C., and Wheeler, P. (1995). The expensive tissue hypothesis: The brain and the digestive system in human and primitive evolution. *Current Antropology, 36*(2), 199-221.

Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge: Harvard University Press.

Bowles, S., Choi, J. K., and Hopfensitz, A. (2003). The coevolution of individual behaviors and social institutions. *Journal of Theoretical Biology, 223*(2), 135-147.

Boyd, R., Gintis, H., Bowles, P., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100*(6), 3531-3355.

Boyd, R., and Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Cavalli-Sforza, L. L., and Feldman, M. W. (1981). *Cultural transmission and evolution*. Princeton: Princeton University Press.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The adpated mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York: Oxford University Press.

Darwin, C. D. (1981). *The descent of man*. Princeton: Princeton University Press.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dunbar, R. (1993). Coevolution of neocortex size, group size and language in humans. *Behavioral and Brain Science, 16*(4), 681-735.

Fehr, E., and Gächter, S. (2000). Cooperation and punishment in public good games. *American Economic Review, 90*(4), 980-994.

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*((6868)), 137-140.

Feldman, M., Aoki, K., and Kumm, J. (1996). Individual versus social learning: Evolutionary analysis in a fluctuating environment. *Anthropological Science, 104*(3), 209-231.

Festinger, L. (1957). *A theory of cognitive dissonance*. Standford: Standford University Press.

Fishbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economic Letters, 71*(3), 397-404.

Gintis, H. (2003). The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology, 220*(4), 407-418.

Gintis, H., Bowles, R., Boyd, R., and Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 71*(3), 397-404.

Gurven, M. (2004). To give and give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences, 27*(4), 543-583.

IN PRINT

Guzmán, R. A., Rodríguez-Sickert, C., and Rowthorn, R. E. (2007). When in Rome, do as the Romans: The coevolution of altruistic punishment, conformism and cooperation. *Evolution and Human Behavior, 28*(2), 112-117.

Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology, 7*(1), 1-52.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization, 53*(1), 3-35.

Henrich, J., and Boyd, R. (1998). The evolution of conformist transmission and between-group differences. *Evolution and Human Behavior, 19*(4), 215-242.

Niven, J. (2005). Brain evolution: Getting better all the time? *Current Biology, 15*(16), R624-R626.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review, 86*(2), 404-417.

Rodríguez-Stickert, C., Rowthorn, R. E., and Guzmán, R. A. (2008). *The social benefit of slow learners.* Manuscrito no publicado.

Simon, H. (1990). A mechanism for social selection and successful altruism. *Science, 250*(4988), 1665-1668.

Sober, E., and Wilson, D. (1994). Reintroducing group selection to the human behavioral sciences. *Brain and Behavioral Sciences, 17*(4), 585-654.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quaterly Rewiew of Biology, 46*(1), 35-57.

Wallace, S. *Notes from Україна: A blue devil's ukrainian experience.*, 2007, from http://chernobyl-summer.blogspot.com/2007/07/chernobyl-liquidators-incredible-men.html.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton: Princeton University Press.

Williams, M. (2002). Primate encephalization and intelligence. *Medical Hypotheses, 58*(4), 284-290.

IN PRINT