
EL DILEMA DEL PRISIONERO: UNA LECTURA ÁCRATA¹

*Martín REMPEL**

Fecha de recepción: 29 de julio de 2016

Fecha de aprobación: 6 de septiembre de 2016

Resumen

El primer objetivo de este trabajo es investigar algunas de las ideas que Piotr Alekseiévich KROPOTKIN (1977) presenta en su artículo “La moral anarquista” (p. 177) y, a partir de la crítica de esos postulados, mostrar que se puede construir una “solución” al dilema del prisionero que permita coordinar la racionalidad individual con la racionalidad colectiva.

Palabras clave

Racionalidad – teoría de juegos – crítica

THE PRISONER’S DILEMMA: AN ANARCHIST READING

Abstract

This work’s main objective is to analyze some of the ideas presented by Piotr Alekseiévich KROPOTKIN (1977) in his article “Anarchist Morality” (p. 177) and, upon the critic of those postulates, to show that it is possible to construct a solution to the prisoner’s dilemma that enables the coordination of individual and collective rationality.

¹ Una versión preliminar de este trabajo —mucho menos desarrollada y que abarcaba solo las primeras dos partes y un fragmento de la tercera— se presentó bajo otro título en la “VI Conferencia Latinoamericana de Crítica Jurídica” organizada conjuntamente por la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata y por el Colegio de Abogados de La Plata, que se realizó los días 13 y 14 de octubre de 2011 en la facultad mencionada.

* Abogado graduado de la Universidad de Buenos Aires (UBA) (Argentina). Profesor adjunto e investigador, Facultad de Derecho, Universidad de Buenos Aires. Correo electrónico de contacto: martin.art.in@gmail.com

Keywords

Rationality – game theory – critique

“El hombre, siempre y en todas partes, sea quien sea, quiere hacer lo que quiere y no lo que le dictan la razón y el beneficio”.

— Fiodor Mijáilovich DOSTOIEVSKI

Introducción

La propuesta que presentamos en este trabajo consiste en investigar algunas de las ideas que Piotr Alekseiévich KROPOTKIN (1977) presenta en su artículo “La moral anarquista” (p. 177) y, a partir de una crítica de esos postulados, mostrar que es posible concebir la posición ética resultante como una respuesta a los problemas presentados por el dilema del prisionero en el marco de la teoría de juegos. En cierto modo, lo que se pretende es mostrar que, a partir de la posición ácrata postulada por el autor ruso, se puede construir una “solución” al dilema del prisionero que permita coordinar la racionalidad individual con la racionalidad colectiva.² La hipótesis que se intentará desarrollar es que la definición que KROPOTKIN brinda de individuo egoísta —diferente, por supuesto, de aquella que es propia de la tradición economicista y hasta de la del utilitarismo clásico— permite realizar un abordaje superador del planteo tradicional del dilema del prisionero, posición desde la cual se puede alcanzar el mejor resultado posible tanto para la sociedad como también para cada uno de los integrantes de la sociedad considerados individualmente.

Para ello, en primer lugar se hará un recorrido por algunas de las ideas más importantes que el maestro ruso expone relativas a la cuestión de la solidaridad —que no es otro que el problema de la colaboración— y de la toma de decisiones, y se expondrá la antropología resultante de tomar esas ideas, en la medida en que resulta posible, de modo sistemático (Sección I). A continuación expondremos el dilema del prisionero en su versión tradicional y explicaremos los resultados a los que —necesariamente— llegan dos egoístas racionales tradicionales (Sección II). Luego mostraremos algunos abordajes del

² Respecto de las posibles objeciones a la expresión “solución al dilema del prisionero” que *a priori* podrían alzarse, señalando que se trata de un sinsentido, se remite a la Sección V, más abajo, donde se brinda una teoría de la solución que consideramos adecuada a los fines que nos proponemos.

dilema del prisionero que revelan que un posicionamiento racional egoísta no es necesariamente excluyente de comportamientos altruistas, introduciendo para afirmar esta hipótesis el escenario teórico de la reiteración de encuentros entre los mismos participantes del intercambio (Sección III). Con este panorama ya delineado, en lo siguiente revisaremos un modelo teórico que hace un abordaje de estas herramientas acorde con nuestra propuesta, a saber: el modelo del altruismo recíproco de Richard DAWKINS. En este punto haremos una breve alusión a las consecuencias de la reiteración de intercambios entre los mismos individuos e intentaremos aclarar un poco la noción de "solución" que aquí se propone (Sección IV). Finalmente, intentaremos desarrollar un breve modelo de aplicación de las ideas éticas de KROPOTKIN al dilema del prisionero y sostendremos que, mediante esta metaética, es posible alcanzar una "solución" racional —tanto en sentido colectivo como en sentido individual— del dilema del prisionero (Sección V).

I.

En el ensayo utilizado como fuente principal de este trabajo,³ KROPOTKIN reflexiona acerca del problema de la solidaridad y de la toma de decisiones. En ese recorrido, luego de formular la pregunta metodológica fundamental que abre la reflexión, a saber, "¿por qué he de ser moral?" —aunque sin dejar de señalar que en su interés también está analizar si la pregunta está bien formulada—, pasa a analizar los motivos de las acciones humanas, esto es, los motivos de *toda* acción humana. Escribe KROPOTKIN (1977: 182):

[s]e comprende fácilmente la admiración de nuestros abuelos cuando los filósofos ingleses, y más adelante los enciclopedistas, afirmaron, frente a estas ideas tan primitivas, que ni diablos ni ángeles tienen que ver con las acciones humanas, sino que todos los actos del hombre, buenos o malos, útiles o perjudiciales, derivan de un solo motivo: el deseo de placer. Todo el clero dejó escapar la palabra "inmoralidad". [...] Y cuando, más adelante, en nuestro siglo, las mismas ideas fueron predicadas por Bentham, Stuart Mill, Tchenichevski y tantos otros, y cuando estos pensadores probaron que el egoísmo es el verdadero móvil de

³ Cabe señalar que —en las tres versiones en español a las que pudimos acceder— el título del *Panfleto revolucionario* que estudiamos es "La moral anarquista".

todas nuestras acciones, las maldiciones fueron en aumento. Y, sin embargo, ¿hay verdad más grande que aquella?

La *verdad* de la que habla KROPOTKIN es, más adelante, explicada con el ejemplo del hombre que le da un pan a un niño: un hombre le da un pan a un niño porque siente placer en el acto de dárselo.⁴ Para KROPOTKIN, todo individuo actúa en obediencia de una necesidad natural, todo acto—sea repugnante, indiferente o atractivo para otros— estará siempre motivado por la búsqueda del placer de quien lo realiza. Según él, esta búsqueda y la búsqueda de la ausencia de dolor son una y la misma búsqueda; explícitamente dice que toda acción humana encuentra fundamento en el hecho de que quien la realiza siente en ella algún placer o en que con ella evita (o cree evitar) algún dolor. Con lo hasta aquí desarrollado podría considerarse que, en este esquema, las motivaciones para toda acción pueden ser dos, a saber: obtener placer o evitar dolor. Sin embargo, como intentaremos mostrar inmediatamente, se trata de una única motivación.

Más adelante, en una nueva exposición de esta idea, KROPOTKIN (1977: 185) dirá: “[h]emos visto que las acciones —conscientes— del ser humano, se consideren virtuosas o viciosas, sean atractivas o repulsivas para los demás, tienen todas el mismo origen: una necesidad, natural en el individuo, de buscar el placer y evitar el dolor”.⁵ De esta manera queda claro que todas las elaboraciones brindadas hasta aquí, que podemos resumir en la especulación de que buscar el placer, por un lado, y, evitar el dolor, por el otro, “viene a ser lo mismo” (p. 183) se explican en el hecho de que para el autor ruso la motivación de toda acción humana es *siempre la misma*.

Ahora bien, en relación con las acciones humanas, se presenta el problema de la posibilidad de distinguir lo bueno de lo malo. Para KROPOTKIN (1977), cualquier individuo puede distinguir perfectamente lo bueno de lo malo sin necesidad de recurrir a la Biblia (o a cualquier libro) o, incluso, a la filosofía: en las necesidades de su propia naturaleza —de la naturaleza del mundo animal— que impulsa a actuar a todo individuo, estarían ya

4 La fórmula utilizada por KROPOTKIN (1977) es “buscar el placer, o bien evitar el dolor, que viene a ser lo mismo” (p. 183).

5 Resulta aquí relevante la utilización de la conjunción puesto que el razonamiento de KROPOTKIN, de aquí en más, tomará ambas acciones (buscar el placer, evitar el dolor) como equivalentes. Nótese que en la definición dada anteriormente se había utilizado una disyunción: “toda acción humana encuentra fundamento en el hecho de que quien la realiza siempre siente en ella algún placer o en que con ella evita, o cree evitar, algún dolor” (la cursiva no está en el original).

inscriptos los lineamientos morales de lo bueno y de lo malo, de modo que todo lo que el hombre debe hacer es atender y responder honestamente a esas necesidades, que son dos: la preservación de la raza y la búsqueda del mayor desarrollo posible de la individualidad (p. 186).

En el capítulo IV de su panfleto revolucionario, en plan de mostrar la capacidad natural de los seres vivos de distinguir entre lo "bueno" y lo "malo", y luego de presentar una serie de ejemplos en los que se analizan, sucesivamente, las acciones de pequeños insectos, de pájaros, de marmotas y hasta del "hombre primitivo", el naturalista ruso nos enfrenta con una idea central de su pensamiento, cual es la idea de *bien*. Con elocuencia, dice KROPOTKIN (1977: 188):

[]a hormiga, el pájaro, la marmota y el salvaje no leyeron a Kant, a los Santos Padres ni a Moisés. Y sin embargo, tienen la misma idea del bien y del mal. Y si reflexionáis un momento acerca de lo que hay en el fondo de esta idea, veréis que lo considerado *bueno* por las hormigas, las marmotas, los moralistas cristianos o ateos es todo lo *útil para la preservación de la raza*, y que lo reputado malo es lo perjudicial para ese fin. No lo bueno para el individuo, como decían Bentham y Mill, sino para la raza entera.

A partir de esta cita se puede comenzar a delinear el carácter del individuo que más tarde, según nuestro plan, se verá enfrentado al dilema del prisionero:⁶ es que tal como explícitamente aquí lo plantea KROPOTKIN, este individuo actuará de modo egoísta, pero será este un *egoísmo kropotkiniano*; esto es, una actitud que tendrá como base axiológica (como base metaética) la afirmación de que aquello que es bueno para la sociedad en la que el individuo vive será bueno para él. Sin lugar a dudas, si bien KROPOTKIN no lo expresa de forma explícita, el trasfondo filosófico de esta postura es que existe —o bien, por lo menos, que puede existir— un determinado tipo de coordinación, de coincidencia o de identidad entre el beneficio individual y el beneficio colectivo.

⁶ Puede aquí comprenderse, entonces, por qué hemos seleccionado la expresión "dilema del prisionero" frente al habitual "dilema de los prisioneros". En efecto, la denominación en singular resulta más adecuada, ya que da cuenta del dilema frente al cual se ubica *cada uno* de los prisioneros. Resultaría metodológicamente erróneo y formalmente falto de rigor considerar que el dilema es enfrentado por ambos prisioneros "en conjunto", y la denominación elegida satisface esta exigencia de precisión terminológica.

Pero nuestro autor anarquista va más allá. Al principio cristiano de no hacer a los demás lo que no queremos que los demás nos hagan a nosotros, KROPOTKIN (1977: 189) contrapone —con fundamento en la observación de “todo el conjunto del reino animal”— el principio de hacer a los demás *lo que quisieras que te hicieran* en similares circunstancias. Y añade (p. 190):

[t]en en cuenta que esto no es más que un consejo; pero este consejo es fruto de la larga experiencia de la vida de los animales en sociedad. Y que en la inmensa mayoría de las sociedades animales, incluida la humana, obrar con arreglo a este principio ha pasado a ser una *costumbre*. Sin esto, por otra parte, ninguna sociedad podría existir, ninguna raza podría haber vencido los obstáculos naturales contra los que ha tenido que luchar.

Esto significa que, para el maestro ruso, la ayuda mutua es necesaria para la superación de los obstáculos naturales contra los que el hombre debe luchar en el desarrollo de la individualidad en este mundo. A partir de estas reflexiones, reformula la definición de lo *bueno*, que previamente expuso como “útil para la preservación de la raza”, y lo presenta ahora como lo “útil a la sociedad en la que [el hombre] vive”. Malo será cuanto resulte perjudicial para esa sociedad (KROPOTKIN, 1977: 190). Nótese que en ambas instancias de la elaboración teórica que analizamos, el criterio empleado para señalar lo característico de lo bueno es la utilidad. Esta idea de lo útil como criterio para identificación de lo óptimo se encuentra desarrollada con claridad por quien mejor ha trabajado el concepto de lo útil: el filósofo John Stuart MILL. En efecto, en el tercer capítulo de su extraordinario *On Liberty*, MILL (1997) expresa lo siguiente: “[s]ería absurdo pretender que la gente [viviera]⁷ como si nada se hubiera conocido en el mundo antes de su venida a él; como si la experiencia no hubiera hecho nada para mostrar que una manera de vivir es preferible a otra” (p. 129).

Es importante resaltar que en el pensamiento de MILL, este criterio utilitarista del “consejo” que puede cualquiera extraer para su propia experiencia de las elecciones de otros en el pasado, es presentado como constitutivo —nada menos— que de la educación,

⁷ Aquí, y en todo lo que sigue, los corchetes insertados en medio de una cita deben llamar la atención sobre algún elemento no incluido en la fuente; el contenido de todos los corchetes es nuestro.

de la formación de la persona de quien se espera que realice una lectura crítica de esas elecciones. Y es en esta línea que continúa MILL (1997: 129-30):

[n]adie niega que la juventud deba ser instruida y educada de manera que conozca y utilice los resultados obtenidos por la experiencia humana. Pero el privilegio y la propia madurez de sus facultades [consisten] en utilizar e interpretar la experiencia a su manera. A [la juventud] corresponde determinar lo que, de la experiencia recogida, es aplicable a sus circunstancias y carácter. Las tradiciones y hábitos de otras gentes son, en una cierta extensión, testimonio de lo que la experiencia les ha enseñado a *ellas*; testimonio y presunción que deben ser acogidos con deferencia por ellas [ya que] su experiencia puede haber sido demasiado escasa, o pueden ellos no haberla interpretado derechamente.

Y, luego, en línea con la metodología que proponemos para este trabajo, señala el filósofo inglés que de lo que se trata es de una elección individual que debe ser voluntaria: "[l]as facultades humanas de percepción, juicio, discernimiento, actividad mental y hasta preferencia moral, solo se ejercitan cuando se hace una elección. El que hace una cosa cualquiera porque esa es la costumbre, no hace elección ninguna" (MILL, 1997: 130). Y, por último, más adelante, afirma: "[l]os seres humanos se deben mutua ayuda para distinguir lo mejor de lo peor" (p. 154), lo que despeja toda duda acerca de que, tanto siguiendo la ética de KROPOTKIN como el criterio utilitarista de John Stuart MILL, se puede alcanzar en este asunto una misma conclusión: el único verdadero resultado que puede alcanzarse del análisis de la experiencia pasada es la importancia de la colaboración: el pasado enseña la ayuda mutua, y ésta debe ser asumida como de la voluntad propia de cada individuo.

Es aquí donde aparece el concepto de solidaridad. En el capítulo V de su panfleto, en forma explícita escribe el autor ruso —sin que esto resulte en absoluto contradictorio con las enseñanzas de Darwin⁸— que el sentimiento de solidaridad es el rasgo

⁸ KROPOTKIN es también reconocido como un importante naturalista y biólogo, y más adelante ahondaremos en este aspecto de su producción. Sin embargo, recordemos que *El origen de las especies...* fue publicado por primera vez en 1859, y que KROPOTKIN comenzó a dar a conocer sus escritos hacia 1873-1876.

predominante de la vida de todos los animales que viven en sociedad, incluido, por supuesto, el hombre. Sostiene que en toda sociedad animal, la solidaridad es una ley natural, infinitamente más importante que la de la lucha por la existencia “cuya virtud nos cantan los burgueses en sus refranes a fin de embrutecernos lo más posible” (KROPOTKIN, 1977: 192).

Quizá para reafirmar que esta idea no es de ninguna manera contradictoria con los postulados evolucionistas de Darwin, explica KROPOTKIN (1977: 190) que justamente, cuanto más desarrollados sean los principios de solidaridad y de igualdad en una sociedad animal, más probabilidades tendrá esa sociedad de salir triunfante en la lucha contra el medio físico y contra sus enemigos. Dicho de otro modo: la manera de vencer los obstáculos es mediante la instauración de la colaboración entre iguales. Esta cuestión de la colaboración entre iguales será retomado posteriormente, al momento de abocarnos al análisis del dilema del prisionero y de sus posibles resultados a partir de las ideas morales de KROPOTKIN.

Continúa el pensador ruso y afirma que la práctica de la solidaridad, en la que se resume el principio de hacer a los demás lo que quisieras que te hicieran en similares circunstancias, ha desempeñado un papel muchísimo más importante que todas las adaptaciones surgidas de la lucha entre individuos por la adquisición de ventajas personales. Y esta afirmación nos permite alcanzar una idea fundamental del ensayo sobre la ética anarquista que desarrolla KROPOTKIN, cual es que, en sus palabras (1977): “[e]l sentido moral es en nosotros una facultad natural, lo mismo que el del olfato y el del tacto” (p. 194). La importancia central de esta aseveración radica en el hecho de que, desde el punto de vista de este autor, definir al hombre como un egoísta racional, en la tradición de la escuela económica tradicional —o aún del utilitarismo clásico— resulta *desacertado*. El error estribaría en perder de vista que *la solidaridad forma parte de la propia naturaleza del hombre*, como lo hacen el sentido del olfato o el del tacto.⁹

En los siguientes capítulos de este trabajo vamos a analizar el dilema del prisionero, en principio, en su versión tradicional. También vamos a decir algo acerca de los resultados que se pueden alcanzar considerando para ese juego la participación de

⁹ Nótese el esfuerzo en “naturalizar” la capacidad de colaboración social solidaria, idea que hacemos propia. Al respecto, se remite a la *Sección IV*, más abajo, en todo lo relacionado a los modos de adopción de estrategias evolutivamente estables (sean estas altruistas o egoístas).

egoístas racionales. Luego intentaremos realizar una comparación de estos resultados con aquellos que alcanzarían jugadores egoístas kropotkinianos, esto es, egoístas solidarios. Sin embargo, antes de ello conviene completar la caracterización moral del hombre que surge de "La moral anarquista", porque esta ponderada solidaridad entre los hombres, esta colaboración mutua que se presenta, este mandato de tratar a los demás como uno quisiera ser tratado, no es otra cosa que el principio de igualdad: principio fundamental del anarquismo, como se observa enseguida (KROPOTKIN, 1977:195):

[l]a igualdad en las relaciones mutuas y la solidaridad que de ella resulta son las armas más poderosas del mundo animal en la lucha por la existencia. E igualdad significa equidad. Declarándonos anarquistas, proclamamos para empezar que renunciamos a tratar a los demás de la forma que no quisiéramos ser tratados por ellos; que no toleraremos en el futuro la desigualdad que ha permitido a algunos de nosotros emplear la fuerza, la astucia o la habilidad, de una manera que ha desagradado a los demás. La igualdad en todo, sinónimo de equidad, es la anarquía.

Más adelante, KROPOTKIN (1977) va a mostrar por qué este principio de igualdad así caracterizado no constituye una incursión iusnaturalista de su posición, sino que se apoya en fundamentos tan sólidos que su cuestionamiento resultaría trivial. Es que las afirmaciones que siguen, relacionadas con la vida inconsciente, pueden leerse en clave de estrategias evolutivamente estables,¹⁰ como se verá más adelante. Comienza KROPOTKIN (1977) el capítulo VII del ensayo que analizamos así (p. 195):

[h]asta ahora, hemos hablado de las acciones humanas conscientes, hechas de forma premeditada e intencionada. Pero junto a la vida consciente está la inconsciente, infinitamente más grande y desconocida [...]. Las tres cuartas partes de nuestras relaciones con los demás son hijas de esa vida inconsciente [...]. Nuestro modo de obrar respecto [de] los demás tiende a convertirse en costumbre [...]. Tratar a los demás como quisiera

¹⁰ Acerca del *concepto* de estrategia evolutivamente estable ver, más adelante, la *Sección IV* de este trabajo.

uno ser tratado llega a ser en el hombre y [en] los animales sociables una costumbre.

Debe tenerse presente esta idea: actuar *por costumbre*. En lo que sigue intentaremos mostrar que diversos comportamientos que vulgarmente pueden ser calificados como meras *costumbres* son, en realidad, estrategias¹¹ que se han afianzado como parte constitutiva del obrar humano. En esta línea, debe considerarse con especial atención cualquier situación en la que un individuo toma determinadas decisiones y *elige* cómo relacionarse con los demás, mientras su marco de acción es —esto es, mientras se conduce “dentro de”— una sociedad en la que las leyes son impuestas mediante la regla de mayoría o, si se quiere, desde otro punto de vista del análisis, por los más poderosos integrantes de la sociedad. En cierto sentido, esto desvirtúa la crítica que podría alzarse apuntando que el individuo que actúa por costumbre ha elegido eso para sí mismo: esa idea solo podría sostenerse si se supusiese la libertad absoluta del individuo al que se acusa de haber elegido la vida decadente.

Frente a este panorama, una última idea: la autoridad como justificación. Sin considerar aquí las implicaciones que el directo ejercicio de la fuerza por parte de cualquier autoridad —en un rango que podría ir desde el Estado hasta el macho dominante en una manada de leones— podría implicar en relación con las decisiones que adopta un individuo determinado, conviene subrayar el hecho de que, bajo determinadas circunstancias, la *mera referencia a la autoridad* —esto es: en completa ausencia de violencia *real*— opera como justificación ya-siempre de las acciones desplegadas. El ejemplo por antonomasia quizá esté constituido por la inevitable justificación romano-burguesa: “*C'est le loi!*”. Al considerar el plano de lo inconsciente, KROPOTKIN llama a reflexionar acerca de los motivos de las acciones, aún de aquellas acciones que él mismo no duda en considerar inconscientes. Esto nos ubica en un panorama de enorme riqueza teórica ya que, superada la instancia de la justificación por la autoridad —esto es: superado el caso de la autoridad *como* justificación—, todo individuo que forme parte de

¹¹ Como se podrá apreciar más adelante en el desarrollo, se trata de estrategias evolutivamente estables.

un colectivo se verá requerido de brindar buenas razones para las decisiones que adopta.¹²

La herramienta teórica que constituye el dilema del prisionero permite superar este tipo de argumentos y centrarse en el análisis de los motivos para la acción que cualquier actor pueda poseer bajo determinadas circunstancias.

II.

El dilema del prisionero es un juego que se estudia con especial interés en el marco de la teoría de los juegos. Vamos aquí a definir un juego en general como una situación en la que dos o más jugadores se ven obligados a realizar uno o más movimientos —esto es, se ven obligados a jugar: a tomar decisiones—, respetando una serie de reglas que determinan para cada movimiento si se produce o no un resultado, y, en caso de que se produzca, en qué consiste este resultado, dando forma así a las jugadas —es decir: secuencias de movimientos que dan lugar a resultados—.¹³ En este contexto, el azar es considerado un jugador más, por lo que pueden existir juegos en los que participen una persona y el azar.¹⁴

La particularidad del dilema del prisionero, al que GAUTHIER (1986) caracteriza como "la clásica situación de adversidad en la maximización de la utilidad" (p. 79), radica en que se trata de un juego en el que se puede predecir perfectamente cuál va a ser el resultado.¹⁵ Y la paradoja que presenta es que, aunque desde la perspectiva de cada jugador, las estrategias que escoge son las únicas alternativas racionales, el resultado al

¹² Podría aquí anotarse rápidamente el fenómeno de la representación, propio del Estado moderno: frecuentemente la invocación al instituto del representante político pareciera "autorizar" al representado a descansar —desprovisto de buenas razones— en la confianza depositada en quien actúa. La falacia radicaría en confundir la calidad del representante con la calidad de las acciones que realiza.

¹³ Una excelente definición de juego es la brindada en RESNIK (1998: 203). Seguimos esa fuente.

¹⁴ Esta importante diferencia entre jugador y persona, por un lado, y el dato de que el azar es tratado como un jugador más, por el otro, permiten explicar por qué en juegos en los que *compiten* dos personas, se pueda considerar que hay tres (o, inclusive, más) jugadores.

¹⁵ En mucha bibliografía sobre el tema suele atribuirse a un tal W. A. Tucker el haber expuesto este razonamiento por primera vez. El investigador Pablo ROBERTI (2015: comunicación personal) sostiene que existen razones para creer que el dilema del prisionero —o, cuanto menos, el modo en que se lo utiliza habitualmente, en particular en teoría económica— es una creación de la Corporación RAND.

que se llega no es el mejor para la sociedad. Y esto sugeriría, quizá, que existe una incompatibilidad entre la noción de racionalidad individual y la noción de racionalidad colectiva.¹⁶

Como última cuestión, previo al abordaje del dilema del prisionero, corresponde definir a qué nos referimos con “acción individual racional” a la que calificamos de *egoísta*. Vamos a definir, siguiendo a MILL (1993) en su “*Greatness Happiness Principle*”,¹⁷ a este tipo de acción de acuerdo al funcionamiento sistemático de los siguientes tres axiomas: (1) las acciones morales son aquellas que maximizan el placer y minimizan el dolor; (2) el mejor estado de cosas es aquel en el cual la suma de lo que resulta valioso es lo más alta posible; y, (3) lo que se debe hacer es lo que lleve al mejor estado de cosas conforme a los dos primeros axiomas. En adelante, consideraremos que los individuos que se enfrentan al dilema del prisionero son egoístas en el sentido de que su comportamiento responde a (y puede explicarse sobre la base de) la puesta en *práctica* de este sistema axiomático.

El dilema del prisionero¹⁸ se trata de la historia de *dos prisioneros* arrestados por vandalismo a los que se priva de la posibilidad de comunicarse entre sí. El funcionario que los arrestó posee pruebas suficientes como para conseguir que se los condene por el cargo de vandalismo —que es un delito menor— pero, al mismo tiempo, está convencido de que han cometido un delito mucho más grave —digamos, un robo a un banco— del que no tiene suficientes pruebas y quiere hacerlos confesar. Así que llama a cada uno de los prisioneros por separado y les propone lo siguiente: “voy a hacerle el mismo ofrecimiento a tu compañero y les voy a dar una hora para que lo piensen antes de darme una respuesta a la siguiente propuesta. Les propongo el siguiente pacto: si uno de ustedes confiesa el robo al banco y el otro no, voy a solicitarle al juez que el que confiese sea condenado solo a un año a prisión y que el otro sea condenado a veinticinco años de prisión. Si confiesan los dos, voy a solicitarle al juez que ambos sean condenados a diez años de prisión. Y si ninguno de los dos confiesa, la pena solo será de dos años de prisión para cada uno, puesto que solo podré probar el cargo de vandalismo, que es un delito mucho menos grave que el robo al banco. Pero estoy convencido de que al menos uno de ustedes va a confesar el robo

16 Además de las consideraciones desarrolladas más arriba en relación con la concepción del egoísmo como parte *naturalmente* constitutiva del individuo en el pensamiento de KROPOTKIN, ver RESNIK (1986: 245).

17 Sobre el “principio de la mayor felicidad del mayor número”, ver MILL (1993: 12-22).

18 Esta forma de exponer el dilema del prisionero está basada, con modificaciones, en RESNIK (1998: 246).

al banco". Por supuesto, una hora después, ambos prisioneros confiesan el robo al banco, y el funcionario le solicita al juez que ambos sean condenados a diez años de prisión.

Es fácil comprender que la única acción racional, para ambos prisioneros, es la de confesar el robo al banco. Pero con ello se ganan el pedido de que sean condenados a pasar diez años en la prisión, cuando uno de los mundos posibles¹⁹ consistía en que solo podrían ser condenados a dos años de prisión cada uno. Lo interesante del dilema del prisionero es que los prisioneros no podían impedir alcanzar el resultado que alcanzaron. Veamos. Supongamos que nuestros prisioneros se llaman Priuno y Pridos. A continuación vamos a explicitar el análisis que realiza *cada uno de ellos* y que los lleva, por ser racionales, a actuar como lo hacen.

El dilema les plantea la posibilidad de ubicarse en cuatro mundos posibles (en adelante, "MP"), cuales son: (MP 1) aquel en el cual ambos confiesan el robo al banco; (MP 2) aquel en el cual Priuno confiesa el robo al banco y Pridos no lo confiesa; (MP 3) aquel en el que Priuno no confiesa el robo al banco y Pridos sí lo confiesa; y, (MP 4) aquel en el que ni Priuno ni Pridos confiesan el robo al banco. Al mismo tiempo, *ambos* conocen de antemano cuál es el riesgo que corren, cada uno de ellos, tomados individualmente, al ubicarse —confesando o no el robo al banco— en cada uno de esos mundos posibles, a saber: en (MP 1) el riesgo es ir diez años a la prisión; en (MP 2) Priuno se arriesga a ir un año a la prisión y Pridos se arriesga a ir veinticinco años a la prisión; en (MP 3) Priuno se arriesga a ir veinticinco años a la prisión y Pridos se arriesga a ir un año a la prisión; y en (MP 4) ambos se arriesgan a ir dos años a la prisión. Es evidente que el mejor mundo posible para la sociedad que Priuno y Pridos constituyen es el MP 4 y, sin embargo, sus acciones egoístas (racionales) los llevan a ubicarse en el MP 1—esto es, aquel mundo posible en el que ambos serán condenados a pasar diez años en la prisión—.

¹⁹ Si bien sobre la noción de *mundo posible*, originalmente introducida por Gottfried Leibniz en 1710, no existe acuerdo, en este trabajo la empleamos bajo la siguiente definición: se trata de un universo (distinto al correspondiente a la realidad) que se estipula mediante descripciones.

Este juego se puede representar con el siguiente cuadro:

[Nota: el par ordenado debe leerse: (Priuno, Pridos)]	Pridos confiesa el robo al banco	Pridos no confiesa el robo al banco
Priuno confiesa el robo al banco	Mundo posible 1: (10, 10)	Mundo posible 3: (1, 25)
Priuno no confiesa el robo al banco	Mundo posible 2: (25, 1)	Mundo posible 4: (2, 2)

Este cuadro de doble entrada permite ver alineadas las posibles actitudes de Priuno, y encolumnadas las posibles actitudes de Pridos y los pagos (los resultados que se obtienen) para ambos por sus jugadas. En el interior se puede ver para cada uno de los cuatro mundos posibles los resultados alcanzados, es decir, la cantidad de años de prisión que sufriría cada uno de los prisioneros en caso de ubicarse en tal o cual mundo posible. Tal como aclara la referencia, los pares ordenados que se presentan entre paréntesis deben siempre leerse teniendo en cuenta que, siempre, el primer valor corresponde a Priuno y el segundo valor corresponde a Pridos. Así, por ejemplo, en el MP 3 —aquel en el cual Priuno confiesa el robo al banco y Pridos no confiesa el robo al banco—, el riesgo para Priuno es sufrir un año de prisión, mientras que para Pridos es sufrir veinticinco años de prisión.²⁰ Analizando el cuadro se puede ver en forma clara la contradicción que existe entre el mejor de los mundos posibles para la sociedad constituida por Priuno y por Pridos —esto es: aquel en el que ambos sufren solo dos años de prisión cada uno (MP 4)—, por una parte, y el mundo que ellos alcanzan mediante las jugadas que realizan, que consisten en confesar el robo al banco —esto es: aquel en el que ambos sufren diez años de prisión cada uno (MP1)—.

²⁰ Quizá sea importante aclarar —toda vez que no resulta lógicamente necesario— que para la lectura del precedente cuadro debe entenderse que los valores más “valiosos” son los menores, ya que se trata del potencial sufrimiento de años de prisión; y presuponemos, por hipótesis, que la racionalidad de nuestros prisioneros importa el deseo de sufrir la menor cantidad de años de prisión posibles.

Esta contradicción entre la *posibilidad* de alcanzar el mejor mundo posible y el hecho fáctico de, que luego de elegir *alcanzan* el peor mundo posible, se puede explicar a través del análisis de los riesgos que cada uno de ellos —tomados, ahora, individualmente— asume al actuar de uno u otro modo. En efecto, puede considerarse la posibilidad de que, consciente de los beneficios que el MP 4 conlleva, Priuno, por ejemplo, asuma seriamente la posibilidad de no confesar el robo al banco; pero con ello se ubicaría en una posición de sumo riesgo. Es que una "traición" por parte de Pridos —esto es, el hecho de que Pridos sí confiese el robo al banco cuando Priuno no lo hace— podría costarle veinticinco años de prisión. Pridos, por su parte, piensa lo mismo que Priuno y se da cuenta de que lo que más le conviene es, por estos mismos motivos, confesar el robo al banco. Dicho en otras palabras: tanto Priuno como Pridos se dan cuenta de que lo mejor que pueden hacer —puesto que no existe la certeza absoluta acerca de lo que hará el otro— es adoptar una actitud defensiva consistente en confesar el robo al banco.

Desde este punto de vista, la posición de Priuno, por ejemplo, podría exponerse así: "si Pridos confiesa y yo no confieso, él iría un año a la prisión mientras que yo iría veinticinco, y si él confiesa y yo también confieso, él iría diez años a la prisión y yo también. Por lo tanto, para el supuesto de que Pridos confiese, mis riesgos son: ir veinticinco años a la prisión (por no confesar) e ir diez años a la prisión (por confesar), y entre veinticinco y diez, naturalmente prefiero diez años de prisión, por lo tanto, para el caso de que Pridos confiese, me conviene confesar el robo al banco. Ahora bien, si Pridos no confiesa y yo no confieso, ambos iríamos dos años a la prisión mientras que si Pridos no confiesa y yo sí lo hago, él iría veinticinco años a la prisión mientras que yo iría solo uno. Por lo tanto, para el supuesto de que Pridos no confiese, mis riesgos son: ir dos años a la prisión (por no confesar) e ir un año a la prisión (por confesar), y entre dos y un año de prisión, naturalmente prefiero ir un año a la prisión, por lo tanto, para el caso de que Pridos no confiese, me conviene confesar el robo al banco. Total que, haga lo que hiciera Pridos, lo que me conviene es confesar el robo al banco". Este razonamiento, que es más sencillo de lo que aparenta, es el mismo que hace en su momento Pridos. Ambos prisioneros notan que, haga lo que haga el otro, lo que les conviene hacer, de acuerdo a lo que prefieren, es confesar. Y, llegado el momento, ambos confiesan el robo al banco y se ganan el pedido del funcionario estatal de que sean condenados a diez años de prisión cada uno.

III.

Corresponde que hagamos algunas precisiones. La actitud de nuestros hipotéticos prisioneros de confesar o de no confesar el robo al banco puede también ser interpretada en términos de colaboración y de traición. De esta manera, no confesar será sinónimo de colaborar o de no traicionar. Por el contrario, confesar será sinónimo de no colaborar o de traicionar. Como se observa, ambas ideas —la idea de colaboración y la idea de traición— se refieren al comportamiento de un actor (de un jugador) en relación con la sociedad de la que forma parte.²¹

La decisión de colaborar o de no colaborar con la sociedad a la que uno pertenece es una instancia de la vida social a la que todos los individuos que conforman una sociedad se ven continuamente llamados. Es la que se plantea, por ejemplo, cuando dos vehículos se encuentran —por vías que se intersectan— en una encrucijada en la que no hay semáforo o cuando al finalizar una cena llega el momento de juntar los platos sucios de la mesa. En el primer ejemplo, si bien el ordenamiento jurídico puede incluir una norma que resuelva la situación disponiendo quién tiene prioridad de paso en una encrucijada sin semáforo, lo cierto es que cualquiera de los dos automovilistas puede obtener una *ventaja* traicionando al otro, esto sería, realizando una maniobra temeraria para avanzar primero y no tener que *perder* tiempo esperando mientras el otro automovilista avanza. En la segunda situación puede verse claramente que la actitud de no colaborar tiene una repercusión individualmente positiva, ya que quien no colabora no *pierde* tiempo juntando los platos sucios.

Otros ejemplos en los que se ve con claridad un dilema del prisionero en su versión clásica son los referidos al cuidado del medio ambiente y a los recursos naturales. El ejemplo por antonomasia quizá lo constituya el caso de ahorro de energía. En efecto, puede en una comunidad darse el caso de que, por algún medio (coercitivo, a través de sugerencias, mediante la búsqueda de un consenso, etc.) se promueva el ahorro energético. Imaginemos un Estado hipotético en el que los gobernantes conminan a la población a colaborar con el ahorro de energía eléctrica, que los gobernantes fundan una

²¹ Como sugerimos, en nuestro caso hipotético la sociedad está constituida por dos individuos: Priuno y Pridos. Es claro que puede pensarse el caso de dilemas del prisionero en donde los “individuos” sean grupos de personas; por ejemplo, pueden concebirse dilemas del prisionero en los que cada prisionero esté representado por una ciudad o por un país. Al respecto, ver la nota 14 de este trabajo acerca de la diferencia entre jugador y persona.

intensa campaña propagandística en pos del ahorro de energía con consignas patrióticas, emotivas y, también, brindando información técnica, señalando, por caso, de qué modo conviene utilizar distintos electrodomésticos para obtener el mejor resultado posible tanto desde el punto de vista individual —esto es: del confort que representa la utilización de dichos electrodomésticos— como desde el punto de vista colectivo —esto es: señalando los beneficios que la sociedad obtendrá como producto de adoptar ese modo de utilización del artefacto de que se trate—. Cada ciudadano puede pensar de este modo: “si ahorro energía eléctrica y los demás no lo hacen, me veré afectado, puesto que la energía a la larga se va a agotar ya que cientos de miles de personas no ahorran y, en el ínterin, yo no habré utilizado, cuanto menos, lo poco que podría haber utilizado hasta el momento en que la energía se agotará finalmente. Por otro lado, si los demás ahorran energía eléctrica y yo no lo hago la energía no se va a agotar ya que cientos de miles de personas ahorran y, en el ínterin, me habré beneficiado del hecho de no haberme privado en absoluto durante la época de racionamiento. Por lo tanto, sea que la energía vaya a agotarse o no—esto es: sea que los demás cientos de miles de habitantes de mi ciudad vayan a ahorrar energía eléctrica o no— lo conveniente es no ahorrar energía eléctrica. Por tanto, no ahorraré electricidad, pase lo que pase”. Este razonamiento que, como es claro, lleva a traicionar (no colaborar), es realizado por todos y cada uno de los integrantes de la sociedad, puesto que todos ellos son igualmente racionales. Y, como resultado, obtenemos que la energía eléctrica no alcance a abastecer la demanda de esa sociedad.²²

Estas posturas que estamos revisando se identifican con las soluciones que brindan los representantes de la escuela del utilitarismo clásico, en el sentido de que los individuos que componen la sociedad (los jugadores de nuestro hipotético caso) se comportan como meros egoístas racionales que buscan la maximización de su beneficio en el caso que se analiza. Sin embargo, el dilema del prisionero es una herramienta que permite realizar un análisis más rico y complejo para el cual debemos considerar el denominado “dilema del prisionero reiterado”.

Este juego, como se puede prever, consiste en analizar los beneficios o los perjuicios para cada jugador, por sus jugadas, mediante la confección de una cadena de dilemas del prisionero que se da entre los mismos jugadores. Recordemos que, en nuestro

²² Si bien escapa a la propuesta de este trabajo, podemos recordar que, tal como se lo plantea en la teoría económica, una de las principales acciones tendientes a aumentar el consumo de cualquier producto consiste en generar la convicción general de que ese producto escasea.

planteo tradicional, la cuestión entre Priuno y Pridos se dirimía en una única jugada individual de cada uno de ellos; jugadas que eran realizadas simultáneamente. Luego de esa única jugada simultánea de cada uno de los jugadores, ya se contaba con toda la información necesaria para analizar el juego. Así concluíamos que se observaba una contradicción entre la racionalidad individual y la racionalidad colectiva: esto es, que si bien la racionalidad individual claramente llevaba a nuestros prisioneros a elegir sus acciones en un determinado sentido, los resultados de esa línea de elección *los alejaba* del mejor mundo posible que podían alcanzar como sociedad. En el caso del dilema del prisionero reiterado, la situación de que cada jugador deba elegir se repite una cantidad de veces. A partir de esta *cadena* de dilemas del prisionero, se pueden extraer conclusiones interesantes para continuar el análisis que proponemos.

Dijimos que se trataba de un juego en el que los individuos debían “competir” repetidamente una cantidad de veces pero no aclaramos un asunto muy relevante, a saber: cuántas veces van a encontrarse nuestros jugadores, uno frente al otro, en esa cadena de dilemas. Como puede fácilmente observarse, la verdadera pregunta no es cuántas veces habrán de encontrarse sino si se encontrarán una cantidad definida o indefinida de veces. En todo lo que sigue, nuestro análisis se va a ceñir al caso en que los jugadores se enfrentan repetidamente una cantidad indefinida de veces, lo que nos permite evitar hacer un abordaje de todos los casos en los cuales los jugadores se encuentran repetidamente una cantidad *predeterminada* de veces,²³ y solo centrarnos en los casos de la reiteración indefinida.

IV.

El año 1976, Richard DAWKINS publicó la primera edición de su extraordinario libro *The Selfish Gene*²⁴ en cuyo capítulo 10 desarrolla el análisis de un determinado tipo de comportamiento social que nos detendremos a analizar en detalle.

²³ Es fácil ver por qué: la si la serie de encuentros tiene, en efecto, un encuentro final, lo relevante es que ninguno de los jugadores sepa qué dilema de la serie es el último ya que, si lo sabe, su racionalidad individual lo llevará, indefectiblemente, a traicionar al otro desde un primer momento. Al respecto ver la extensa nota aclaratoria en REMPEL (2015: 573).

²⁴ DAWKINS (2006) se trata de una edición “aniversario” por haberse cumplido treinta años de la primera edición del libro. Se trata de la tercera edición.

Explica DAWKINS²⁵ que en numerosas especies animales (y el hombre es una de ellas) la evolución de la asociación para el beneficio mutuo es fácil de concebir si los favores son dados y recibidos simultáneamente (DAWKINS, 2006: 183). Sin embargo, los problemas aparecen cuando hay una demora entre el momento en el que se brinda ayuda a otro y el momento en el que ese favor es retribuido. Para trabajar este problema, DAWKINS emplea un ejemplo que constituye, como se verá a continuación, un dilema del prisionero reiterado una cantidad indefinida de veces.

Asumamos que una especie de pájaro es regularmente atacada por una garrapata que acarrea una peligrosa enfermedad. Como todo parásito, esta garrapata representa una amenaza constante de tal magnitud que se vuelve muy importante su eliminación inmediata. Cualquier pájaro puede eliminar sus propias garrapatas sin más ayuda que su propio pico, pero hay un lugar al que no puede alcanzar: la parte superior de la cabeza. La solución a este inconveniente se le ocurre a cualquier ser humano: nada es más fácil que un amigo se acerque y elimine la garrapata. Más tarde, cuando el amigo sea atacado por el parásito, el primero podrá devolverle el favor haciendo otro tanto.²⁶

Resulta intuitivamente razonable ingresar en el esquema de colaboración mutua. Ahora bien, la tesis sostenida por DAWKINS en su trabajo implica profundizar el análisis respecto de estas actitudes de *altruismo recíproco* allí donde se observa una *demora* entre la ayuda brindada y su retribución. La herramienta que poseemos para realizar este análisis profundo es, justamente, el dilema del prisionero en su versión reiterada. Diremos entonces que este comportamiento altruista recíproco, aparentemente estable, puede convertirse fácilmente en un dilema del prisionero reiterado, y ello está dado por el hecho de que esta ponderada estabilidad es, justamente, aparente.

A efectos de dar cuenta de la profundidad del alcance de este análisis, debemos desarrollar el concepto de *estrategia evolutivamente estable*. Este concepto proviene de la biología y su elaboración se atribuye a MAYNARD SMITH y PRICE (1973). En REMPEL (2015:

²⁵ De este punto en adelante se sigue, con alguna modificación menor, la exposición que se puede encontrar en DAWKINS (2006: 181-8).

²⁶ La actitud solidaria de acicalamiento común, de hecho, es muy común tanto entre pájaros como entre mamíferos; también es común el avistar aves sobre el lomo de grandes mamíferos, o mamíferos acicalándose mutuamente. Para un tratamiento de esto, ver "Bart Gets an Elephant" ["Bart gana un elefante"] en *The Simpsons*, temporada 5°, capítulo 98, escrito por John Swartzwelder y dirigido por Jim Reardon.

576) lo explicamos así: se trata de un modo de comportamiento que, una vez adoptado por una población, se impone como único modo de acción posible impidiendo que cualquier otro tipo de comportamiento prospere. Se la considera como un tipo de especialización del equilibrio de Nash y se la caracteriza como “evolutiva” porque la selección natural basta para excluir cualquier estrategia alternativa y porque se impone por vía evolutiva —es decir que se impone en tiempos evolutivos a lo largo de varias generaciones y no durante el tiempo de vida de un único individuo—.

Supongamos que Priuno tiene una garrapata en la parte superior de la cabeza y que Pridos la elimina por él. Más tarde llega el momento en que Pridos posee una garrapata sobre su cabeza por lo que naturalmente busca a Priuno para que le devuelva el favor, pero Priuno, simplemente da media vuelta y se retira del lugar. Priuno es un traidor: un individuo que recibe el beneficio producto del altruismo de otros individuos, pero que no “devuelve” los favores que recibe.²⁷ La clave de este comportamiento de Priuno, asumiendo que el costo de acicalar la cabeza de Pridos parece menor en relación con el beneficio de eliminar una peligrosa garrapata, está dada por la valiosa energía y, principalmente, por el muy valioso tiempo que perdería en caso de colaborar con Pridos.

Veamos ahora que sucede cuando los individuos adoptan estrategias que *determinan* su comportamiento. Supongamos que la población de una hipotética sociedad de pájaros adopta una de las dos siguientes estrategias, pero previamente aclaremos—de acuerdo a la definición que brindamos más arriba— que no se trata de estrategias conscientes, sino que se trata de comportamientos inconscientes que se ven afianzados en tiempo por el refuerzo que supone la obtención de resultados beneficiosos. Las dos estrategias que vamos a proponer para nuestro análisis son: (a) la estrategia que denominaremos *altruista* —consistente, como se puede prever, en que el individuo que adopte esta estrategia colaborará en toda interrelación social de la que participe—; y (b) la estrategia *egoísta*—consistente, como se puede prever, en que el individuo que adopte esta estrategia traicionará en toda interrelación social de la que participe—. Frente a un hipotético pedido de ayuda, los altruistas siempre colaboran—esto es, siempre dedican energía y tiempo a eliminar la garrapata del pájaro que requiere ayuda—, mientras que los egoístas nunca colaboran—esto es: siempre traicionan— y con eso ahorran la energía y el tiempo que dispensarían en eliminar la garrapata que amenaza al otro individuo.

²⁷ Siguiendo a DAWKINS (2006), consideraremos también traidor al individuo que devuelve *insuficientemente* los favores que recibe.

Si imaginamos una población en la que la mitad de los individuos adoptó la estrategia altruista y en la que la otra mitad de los individuos adoptó la estrategia egoísta, es fácil inferir qué sucedería con el paso del tiempo. Es que no hay modo de mejorar la posición de los egoístas ya que, hipotéticamente, cualquier egoísta podría obtener el beneficio de ser acicalado por algún altruista sin nunca tener que perder tiempo y energía en acicalar a nadie. Dicho en términos probabilísticos: si se toma en cuenta que la mitad de la población adoptó su misma estrategia, cualquier egoísta necesitaría "pedir ayuda" dos veces —asumiendo la potencial negativa en el primer encuentro con otro egoísta— para obtener la colaboración que requiere. De este modo, puede generalizarse el análisis y afirmar que en una sociedad en la que la mitad de los individuos están dispuestos a colaborar siempre, la otra mitad compuesta por quienes no están dispuestos a brindar esa colaboración va a recibir un beneficio neto sobre la suma de energía y tiempo que se "ahorra" por traicionar en cada intercambio social. Sin embargo, esta situación no es estable, ya que, sin que importe el *ratio* en la población, los egoístas siempre van a estar mejor, y por lo tanto va a perfeccionarse una tendencia a adoptar la estrategia egoísta en desmedro de la altruista.

Uno podría imaginar que esta población —inicialmente consistente en un 50 % de altruistas y en un 50 % de egoístas— paulatinamente va modificando su constitución en favor de la estrategia egoísta. Resulta claro que hasta en una sociedad que contenga un 90 % de egoístas —con independencia del hecho de que solo el 10 % de la población debe realizar el trabajo de acicalamiento del total de los individuos— cualquier egoísta sigue estando "mejor" que cualquier altruista, aun cuando desde un análisis probabilístico obtener la ayuda que le evita la muerte —esto es, lograr que otro individuo lo acicale— le va a "costar" nueve hipotéticos pedidos de ayuda. Aun cuando la totalidad de la población se vuelque a adoptar la estrategia egoísta²⁸ y, en consecuencia, la extinción de la especie esté asegurada —por la infección que genera la garrapata que ya nadie estará dispuesto a eliminar de la cabeza de otro individuo—, no va a haber ningún punto en el tiempo en que los egoístas no estén mejor que los altruistas. Esto es equivalente a afirmar que, para una población como la que hasta acá imaginamos, la estrategia egoísta constituye una estrategia evolutivamente estable, y "si una población adopta una estrategia evolutivamente estable que la conduce a la extinción, entonces se extingue y qué pena" (DAWKINS, 2006: 186).

²⁸ Lo que en efecto sucede porque, como intentamos mostrar, siempre se "está mejor" siendo egoísta que siendo altruista.

Ahora bien, supongamos que junto a la altruista y a la egoísta existe una tercera estrategia que denominaremos *intelectual*. Los intelectuales acicalan a extraños y a otros individuos (conocidos) que previamente los acicalaron a ellos, pero, si algún individuo los traiciona —esto es, si un conocido se niega a “devolver” el favor de un acicalamiento previo, o si un extraño se niega a acicalarlo la primera vez que se encuentran en un intercambio social—, los intelectuales recuerdan el incidente y se niegan a colaborar *con ese individuo* en el futuro. Podría decirse así: frente a una traición, el intelectual se vuelve “traidor” respecto del individuo que traicionó primero. Como consecuencia, los intelectuales desarrollan una protección estratégica, a saber: solo podrán ser traicionados una sola vez por cualquier otro integrante de la sociedad.

Una población enteramente constituida por altruistas y por intelectuales, sin importar qué porcentaje se corresponda con cada estrategia, importa la imposibilidad de determinar quién adoptó qué estrategia. Ambas estrategias determinan que los individuos se comporten, respecto de todos los demás, de modo igualmente altruista, a saber: todo intercambio social va a concluir con un acicalamiento. En este caso, todos los individuos, altruistas e intelectuales, estarán en un equilibrio total: nadie estará “mejor” que nadie, puesto que el gasto en energía y en tiempo que supone acicalar a otro individuo estará parsimoniosamente distribuido entre todos los integrantes de la sociedad.²⁹ En una población compuesta totalmente por egoístas con excepción de un único intelectual, este solitario individuo no va a tener mucho éxito: va a perder una enorme cantidad de energía y de tiempo acicalando a todos los demás individuos (desconocidos) de esa sociedad hasta poder generar un nuevo encuentro con cada uno de ellos en el cual podrá decidir, en lo sucesivo y para cada uno, *dejar de colaborar* con ese individuo. Además, como es evidente, no encontraría a nadie que colaborara con él por lo cual la muerte por la infección que genera la garrapata sería solo una cuestión de tiempo. En una sociedad en la que el número de intelectuales fuera muy menor al número de egoístas, los intelectuales perecerían.

Pero en una población en la que los intelectuales han logrado alcanzar un número interesante, sin perjuicio del esfuerzo que supondría el terminar de “conocer” a todos los demás para saber con quién se puede contar y con quién no, la estrategia daría resultados

²⁹ Acerca del concepto de parsimonia, a veces denominado principio de economía, quizá sea relevante recordar que la adopción de estrategias evolutivamente estables constituye un proceso evolutivo, y, por tanto, que estos modelos deben entenderse dinámicamente, manifestados en el tiempo.

muy distintos a la extinción final. Alcanzada esta proporción, los intelectuales comenzarían a recibir un beneficio por cada intercambio social que, en promedio, superaría siempre los beneficios alcanzados por los egoístas. Conformado este (pequeño) grupo de intelectuales cuya estrategia les permite confiar en la colaboración de los demás intelectuales, quienes encararían —cada instante con mayor velocidad— el camino a la extinción serían ahora los egoístas.³⁰ El número de egoístas bajaría muchísimo y llegarían casi al borde de la extinción y aumentaría en proporción el número de intelectuales. No obstante, una vez alcanzado un número muy bajo de egoístas, se daría una situación muy interesante, a saber: los (poquísimos) egoístas comenzarían a gozar —en una sociedad ahora compuesta mayormente por intelectuales que, como dijimos, colaboran *por lo menos una vez* con cualquier desconocido— de una nueva posibilidad de ser acicalados sin tener que devolver el favor, esto es, con el consiguiente ahorro en tiempo y en energía que supondría la colaboración recíproca. Como se puede ver, esta situación modificaría nuevamente los niveles de utilidad, ubicando a los egoístas en promedio por encima de los intelectuales, que deben gastar energía y tiempo en colaboraciones recíprocas por las que fueron beneficiados en el pasado. Entonces, comenzaría a aumentar la proporción de egoístas, primero con lentitud, luego en forma precipitada. Y, cuando el número de los egoístas superase ampliamente al número de intelectuales, nos encontraríamos ante la situación inicial, es decir: el número de egoístas bajaría muchísimo y llegarían casi al borde de la extinción.

Este ciclo se extiende indefinidamente. Se trata de una situación pendular con períodos de mayor estabilidad y con períodos de mayor nivel de cambios, que alterna una población con un alto número de intelectuales (y un bajo número de egoístas muy beneficiados por la conformación de las estrategias sociales) con una población con un alto número de egoístas (y un bajo número de intelectuales muy beneficiados por la conformación de las estrategias sociales). En una simulación computarizada en la que se estudia el devenir de una población compuesta por altruistas, egoístas e intelectuales, DAWKINS (2006: 186) muestra el modo en que, en primer lugar, se da una *desaparición* abrupta de la estrategia altruista, para, luego, pendular en el tiempo las estrategias egoísta e intelectual.

30 Quizá sea correcto aclarar que en el ejemplo que antecede, en el que la población de intelectuales era tan baja que los intelectuales perecían, con posterioridad a la desaparición de los intelectuales se daría la extinción de los egoístas (puesto que ya nadie acicalaría a nadie, y la garrapata terminaría por matar a toda la especie, compuesta en ese punto por egoístas).

V.

En la primera parte de este trabajo anunciamos que retomáramos la cuestión de la colaboración entre pares, luego del recorrido que nos llevó a discutir el dilema del prisionero bajo la óptica de la teoría de juegos. También dijimos que retomáramos la posición kropotkiniana desarrollada en esa parte. Pero antes de avanzar en ese camino conviene aclarar el alcance de la expresión “solución del dilema del prisionero”.

Como se ha podido notar en la Sección II de este trabajo, no resulta preciso el término *solución* aplicado al dilema del prisionero, si se pretende con ello decir que se trata de un modo de “solucionar” la paradoja presentada por esa herramienta, a saber: que el hecho de que los jugadores sean individuos racionales lleva *necesariamente* a que los resultados alcanzados sean, desde el punto de vista colectivo, muy deficientes. Esa necesidad que resaltamos, ese carácter deductivo que posee la conclusión paradójica del planteo, hace imposible hablar de una “solución” del dilema del prisionero, ya que un razonamiento deductivo no concluye como uno querría que concluyera sino que concluye sobre la base de los valores de verdad de las premisas y del funcionamiento de los conectores lógicos empleados. Dicho en un lenguaje más claro: la conclusión de un razonamiento deductivo como el dilema del prisionero se encuentra, de alguna manera, ya contenida en las premisas que lo conforman, de modo que no podría solucionarse el dilema, en el sentido apuntado de solución, sin modificar, de alguna manera, las premisas o la racionalidad empleada por los jugadores que se enfrentan al dilema.

Este último es, justamente, el camino de nuestra propuesta kropotkiniana. Es que, si bien KROPOTKIN no leyó el trabajo de DAWKINS podemos encontrar en sus ideas acerca de la moral una conclusión semejante a la que alcanzamos en el punto anterior. Vimos ya el modo en que el maestro ruso equipara el placer con lo útil para la preservación de la raza (KROPOTKIN, 1977: 183-6). También vimos que esta noción, expresada como el *principio de hacer a los demás lo que quisieras que te hicieran en similares circunstancias*, es presentada en los términos de un consejo que, naturalmente, cualquiera puede tomar o dejar (p. 190). Con D'AURIA (2012), afirmamos que “Kropotkin deriva sus tesis libertarias del evolucionismo darwiniano. Pero a diferencia de los ‘darwinistas sociales’ de su época, Kropotkin pone de relieve al ‘apoyo mutuo’ como factor de evolución de las especies” (p. 128).

Frente a la pregunta por el modo de obtención de esta *colaboración mutua*, la respuesta de KROPOTKIN procede por vía de reafirmación de la libertad del hombre: se presenta un consejo —no una orden jurídica, no una orden moral— que, de ser tomado, daría lugar a esa colaboración. ¿Qué quiere decir, en los términos aquí desarrollados, tomar el consejo de KROPOTKIN de hacer a los demás lo mismo que querrías que te hicieran en similares circunstancias? Simplemente significa adoptar *voluntariamente* una posición colaboracionista como la desarrollada en los puntos anteriores. Esto es: no señalar el carácter irresistible de una estrategia que “determine” las acciones de altruismo, sino la adopción *voluntaria* de una estrategia para los intercambios sociales, capaz de hacer compatible la racionalidad individual con la racionalidad colectiva.³¹ Esto no es tan difícil como en apariencia se presenta.

Supongamos que nuestros jugadores—Priuno y Pridos— se enfrentan con el dilema del prisionero. Mantengamos todos los elementos que hacían a las propiedades de ese juego intactos, pero modifiquemos—siguiendo el consejo de KROPOTKIN— el aspecto que hace exclusivamente al modo en que cada uno de ellos toma sus decisiones, esto es: la racionalidad individual que cada uno aplica. Asumamos el mismo juego enfrentado por dos jugadores racionales kropotkinianos que, adoptando el consejo, toman sus decisiones sobre la base de hacer *exactamente aquello que querrían que les hicieran en similares circunstancias*. Es muy fácil ver que cualquier jugador racional al que se le preguntase qué querría obtener del otro integrante de la sociedad, enfrentados ambos al dilema del prisionero, respondería que todo lo que desea es no ser traicionado por el otro, es decir, un jugador racional respondería que quiere *la colaboración del otro*.

Veamos, entonces, siguiendo el consejo kropotkiniano, el modo en el cual, en lugar de las ilaciones presentadas en la [Sección II](#) de este trabajo, tanto Priuno como Pridos podrían razonar, asumiendo que van a actuar haciendo aquello que cada uno querría que les hiciera el otro en esa situación: “si Pridos confiesa y yo no confieso, él iría un año a la prisión mientras que yo iría veinticinco, y si él confiesa y yo también confieso, él iría diez años a la prisión y yo también. Por lo tanto, para el supuesto de que Pridos confiese sus

³¹ En esta línea, afirma D’AURIA (2012) que “el anarco-comunismo de Kropotkin afirma el principio “de cada uno según su deseo y posibilidad, a cada uno según su necesidad” (p. 131). Resaltamos el extraordinario respeto que esta posición metaética dispensa a los deseos de (o, si se quiere, a la racionalidad empleada por) todos los integrantes de la sociedad.

riesgos son:³² ir diez años a la prisión (para el caso en que yo sí confieso) e ir un año a la prisión (para el caso en que yo no confieso); y, entre diez y uno, naturalmente él prefiere un año de prisión. Por lo tanto, para el caso de que Pridos confiese, me conviene no confesar el robo al banco. Ahora bien, si Pridos no confiesa y yo no confieso, ambos iríamos dos años a la prisión mientras que, si Pridos no confiesa y yo sí lo hago, él iría veinticinco años a la prisión mientras que yo iría solo uno. Por lo tanto, para el supuesto de que Pridos no confiese, sus riesgos son: (1) ir dos años a la prisión, para el caso en que yo tampoco confieso; y (b) ir veinticinco años a la prisión, para el caso en que yo sí confieso; y, entre dos y veinticinco años de prisión, naturalmente él prefiere ir dos años a la prisión. Por lo tanto, para el caso de que Pridos no confiese, me conviene no confesar el robo al banco. Total que, haga lo que hiciere Pridos, lo que me conviene es no confesar el robo al banco”. Como resultado de esta racionalidad se obtiene la colaboración mutua.

Como se viene anunciando, este mismo razonamiento sería el realizado por Pridos, lo que llevaría a ambos jugadores —actuando sobre la base de lo que querrían que les hicieran si estuvieran en similares circunstancias— a colaborar mutuamente, esto es: a no confesar el robo al banco y, consecuentemente, pasarían a sufrir, cada uno, dos años de prisión contra los diez años que sufrirían su fueran egoístas racionales no kropotkinianos. Otro modo de exponer esta situación consiste en señalar que cada jugador decide sus acciones sobre la base de qué es lo mejor para el prójimo.

Este tipo de posición colaboracionista fue expuesta por todo un plexo de pensadores libertarios, y esto ha llevado algunos autores a ensayar críticas a esta posición colaboracionista, fundadas en una antropología ingenua o deficiente.³³ En relación con este punto, nuestro recorrido inicial por el pensamiento de John Stuart MILL y la posición utilitarista nos permite asumir la impertinencia de tales críticas. Desde otro lado, se ha afirmado que nadie actúa racionalmente como un héroe —esto sería, en beneficio de otro— y hasta se ha invocado a alguna autoridad de la filosofía ilustrada para sostener este tipo de crítica. Al respecto, solo podemos aclarar que aquí no se promueve ningún comportamiento heroico—que siempre *debe* ser voluntario—, sino que simplemente se asume la posibilidad de que alguien actúe motivado en el placer. Y este placer, es claro, *siempre* es individual.

32 Nótese la modificación clave: antes cada jugador razonaba teniendo en cuenta sus propios riesgos mientras que, en el caso kropotkiniano, los jugadores razonan atendiendo a los riesgos *del otro* jugador.

33 Al respecto, ver D'AURIA (2012: 130) en donde hay una exposición muy clara de estas críticas.

Así las cosas, lo que aquí se propone es nada menos que lo siguiente: que la búsqueda *individual* del placer, llevada adelante por individuos racionales, no puede sino llevar a la colaboración mutua. Algo así como afirmar que todo egoísta, por el hecho de serlo, debe colaborar. Quizá un camino similar a este haya recorrido Émile ARMAND (2007:47) antes de sostener que "[n]ada nos prueba que el 'egoísmo bien razonado' no pueda llegar a otra cosa que a la colaboración".

Bibliografía

(Excepto en donde se indica una versión en español, todas las traducciones son propias.)

ARMAND, É. (2007) *El anarquismo individualista. Lo que es, puede y vale*. La Plata, Terramar Ediciones.

D'AURIA, A. (2012) *Teoría y crítica del Estado*. Buenos Aires, Editorial Eudeba.

DAWKINS, R. (2006) *The Selfish Gene*. New York, Oxford University Press.

GAUTHIER, D. (1986) *Morals by Agreement*. New York, Oxford University Press.

KROPOTKIN, P. (1977) "La moral anarquista" en *Panfletos revolucionarios*. Madrid, Ayuso, pp. 177-209.

MAYNARD SMITH, J. y PRICE, G. (1973) "The Logic of Animal Conflict" en *Nature*, volumen 246, número 5427, pp. 15-8.

MILL, J. (1993) *Utilitarianism, On Liberty, Considerations on Representative Government*. Londres, Everyman's Library.

— (1997) *Sobre la libertad*. Madrid, Alianza.

REMPEL, M. (2015) "El dilema del prisionero y otras situaciones paradójales. Una breve reflexión a propósito de la racionalidad aplicada al comportamiento colectivo" en ALONSO, J. (comp.), *Racionalidad en el derecho*. Buenos Aires, Eudeba, pp. 567-79.

RESNIK, M. (1998) *Elecciones. Una introducción a la teoría de la decisión*. Barcelona, Gedisa.