

Dilemas sociales I: el dilema del prisionero y sus variantes

Juan Pazos Sierra

1. Dilemas y sus clases

En lógica se entiende por *dilema* el silogismo condicional en el que la proposición mayor es una disyuntiva de suposiciones opuestas de cada una de cuyas alternativas se sigue la misma conclusión. En efecto, para la lógica un dilema es, como las llamas del infierno, un argumento bicornes; es decir, un argumento compuesto de dos proposiciones contradictorias con cualquiera de las cuales, concedida o rechazada por el adversario, se llega a la misma conclusión. El ejemplo más habitual que empleaban los libros clásicos de filosofía para explicar qué es un dilema es el del centinela, que se expresa como sigue: *o abandonaste la guardia o dejaste pasar al enemigo. En ambos casos, mereces la muerte*. Otro ejemplo, también escolar, es el siguiente: *los hombres cometen los asesinatos que traman o no los cometen. Si los cometen, pecan contra la ley divina, y son culpables. Si no los cometen, pecan contra su conciencia moral, y son culpables*. Por consiguiente, cometan o no cometan los asesinatos que traman, son culpables; si traman un asesinato, claro. Más formalmente, en la teoría silogística aristotélica los dilemas, tal y como se muestran en la figura 1, pueden ser de distintos tipos.

$$\text{DILEMAS} \left\{ \begin{array}{l} \text{Simples} \left\{ \begin{array}{l} \text{Constructivo: } \frac{(p \rightarrow q), (r \rightarrow q), (p \vee r)}{q} \\ \text{Destructivo: } \frac{(p \rightarrow q), (r \rightarrow q), (q' \vee r')}{p'} \end{array} \right. \\ \text{Complejos} \left\{ \begin{array}{l} \text{Constructivo: } \frac{(p \rightarrow q), (r \rightarrow s), (p \vee r)}{q \vee s} \\ \text{Destructivo: } \frac{(p \rightarrow q), (r \rightarrow s), (q' \vee s')}{p' \vee r'} \end{array} \right. \\ \text{Clásico: } \frac{([p \rightarrow (q \vee r \vee s)], q', r', s')}{p'} \end{array} \right.$$

Figura 1. Tipos de dilemas.



Un ejemplo patético de dilema constructivo complejo fue el que aplicó el califa Omar para llevar a la hoguera todos los libros de la biblioteca de Alejandría, cuyo enunciado fue: *si estos libros coinciden con el Corán son superfluos. Si no, son perjudiciales. Si los libros son superfluos deben quemarse. Si son perjudiciales, también deben quemarse. Esos libros son superfluos o perjudiciales. Por consiguiente, deben quemarse.*

Sin embargo, hoy se entiende por dilema la situación en la que uno se encuentra cuando se debe decidir entre dos cosas igualmente malas. Este enfoque de *Teoría de Juegos* es el que aquí, ahora, se va a considerar. Es decir, un dilema es un tipo de *juego*, en el sentido von Neumann, denominado de *suma no nula* o de *suma no cero*, en el que ambas partes disponen de opciones en las que las pérdidas de uno no suponen ganancias para el otro, y viceversa.



Figura 2. Von Neumann.

En el folclore africano (Bascom, 1972) abundan disyuntivas excéntricas y difíciles de resolver como la siguiente, que aparece narrada en un cuento tradicional que contiene un dilema, tal y como lo relata el Jeque de Dahomey: *un hombre iba a cruzar un río con su mujer y su madre. De pronto, apareció una jirafa en la orilla opuesta. El hombre sacó su rifle y apunta a la jirafa. Pero ésta, sin preocuparse demasiado, le dijo: si disparas, morirá tu madre. Si no lo haces, morirá tu mujer.* La cuestión surge inmediatamente, naturalmente bajo la premisa de que todo lo que dice la jirafa habladora es verdad; ¿qué hacer?

Por supuesto, este dilema puede replantearse en términos más apropiados a la cultura y lógica occidentales. Por ejemplo, un vesánico ha raptado a un matrimonio, que se ama con locura y a su hija a la que adoran, y les ha encerrado en un *zulo*. En él hay una ametralladora conectada a un botón, que deja al alcance del hombre, y que apunta a la mujer y a su hija, y un reloj que marca los segundos. El loco, le

explica que si oprime el pulsador, la ametralladora apuntará a su mujer y la acribillará. Por el contrario, si no pulsa el botón en un lapso de sesenta segundos, el artefacto encañonará a su hija y disparará, con resultado mortal. Ni que decir tiene que la máquina funcionará inexorablemente. La cuestión que se plantea es tan difícil como dramática: ¿qué debería hacer?

Dilemas de este tipo se discuten a veces en las clases de ética. Evidentemente, no hay una solución satisfactoria. Tampoco cabe inhibirse, bajo la argucia argumental de evadir la responsabilidad con la inacción y la consiguiente inculpabilidad por no hacer nada, dado que la inacción llevaría a la máquina a matar al cónyuge. La alternativa está en decidir a cuál de las personas salvaría, porque la quiere más o, lo que sería más interesante, porque encontró un argumento para decidirlo.

Decisiones de este jaez son, aun si cabe, más peliagudas cuando no es uno solo el que decide, y el resultado depende del conjunto de todas las decisiones tomadas. Gregory Stock (Stock, 1987) plantea, en este sentido, el dilema siguiente: *una pareja de enamorados a tope, se sitúan en habitaciones separadas provistas de sendos pulsadores. Saben que matarán a ambos a no ser que uno pulse el botón antes de una hora. Pero, eso sí, la primera persona que accione el pulsador salvará a la otra, pero morirá inmediatamente. ¿Qué decisión tomar?*

En este dilema, la situación es la siguiente: ahora hay dos personas deliberando sobre su situación que tienen que tomar decisiones independientes. Es necesario que alguien pulse el botón. Lo difícil es decidir cuando uno deberá sacrificarse, dado que ambos están dispuestos a dar la vida por el otro. El dilema obliga a tomar una *decisión tipo bote salvavidas* que afecta a alguien a quien uno quiere y a uno mismo. La cuestión va de suyo: ¿quién debería sobrevivir?

Existen situaciones, ¿muchas?, en las que el altruismo o el amor personal, etc., del ser humano decide salvar a otro a expensas de perder la propia vida. Sea el que fuere el motivo alegado y el criterio usado, y sabiendo que ambas personas no tendrían el mismo, al tomar una *decisión tipo bote salvavidas*, hay los tres desenlaces posibles siguientes: el caso menos angustioso es cuando ambos coinciden en quién debe sacrificarse y quién debe salvarse. Entonces aquél debería pulsar el botón para salvar al otro. Una segunda posibilidad es que ambos decidan salvarse uno al otro. La madre decide salvar a su hija, a la que le quedan más años de vida; la hija decide, asimismo, salvar a su madre, pues ésta le dio la vida. En este caso, se compite por ser el primero en pulsar el

botón. La opción más peliaguda surge cuando *ambas* personas extraen lo peor de ellas mismas y, egoístamente, deciden que es ella misma quien debe salvarse. Entonces nadie pulsa el botón, y el reloj sigue su curso inexorablemente..., llegándose a la situación siguiente: según el reloj han pasado ya 59 segundos. Uno no oprime el botón, con la esperanza de que el otro, que le había manifestado su amor eterno y hasta la muerte, lo hiciera, pero su esperanza fue vana. Supóngase que la noticia de la muerte de uno cualquiera de los protagonistas del dilema es comunicada de inmediato al superviviente en cuanto éste pulsó el botón. Y, por supuesto, cada uno de los afectados ha tenido tiempo suficiente para considerar todas las posibilidades. En efecto, a ciertas personas, indecisas, puede llevarles una hora entera el decidir quién se salvará para reunir el valor suficiente para apretar el botón. Sea como fuere, no hay duda de que la persona a la que uno *adora* ha decidido no ser ella quien se sacrifique.

Por otra parte, no tiene sentido comprometerse a no pulsar *jamás* el botón, ni siquiera en la última fracción de segundo. Por muy egoísta que uno sea, uno no puede salvarse a sí mismo. Inexorablemente, alguien ha de morir. Así es como está planteado el dilema. Si la persona a la que uno ama no se aviene a sacrificarse, por lo menos uno tiene que salvarla. Naturalmente, la situación ideal sería que uno pulsara el botón en el último instante. Es posible que la otra persona piense lo mismo. Precisamente, eso es lo que le da fuerzas a uno para resistir hasta el último momento, pues así se da a la otra persona la oportunidad de apretar el botón en el penúltimo instante. Si no fuera así, entonces y sólo entonces, pulsaría uno su botón. Por supuesto que es posible que el ser amado planee esta *misma* estrategia.

Al llevar la situación al límite de resistir mutuamente hasta el último instante posible, el desenlace dependerá de los tiempos de reacción y de la precisión del reloj. La *malvada* máquina no tiene en cuenta la *intención* de haber pulsado el botón dentro del margen de tiempo estipulado. O bien uno, o bien la otra persona, han de cerrar mediante su actuación el infernal circuito, antes de que la máquina se ponga en marcha y mate a ambos. Además, no se sabe a ciencia cierta si el reloj está perfectamente sincronizado con la máquina. Eso dijo el loco, pero tal vez no se refería al segundero, con el añadido de que los locos son arbitrarios. Es decir, para asegurarse el poder pulsar el botón dentro de un margen de seguridad, uno debería en realidad apretarlo un instante o dos antes. ¡Sin embargo, este requerimiento es contrario a la intención de resistir hasta el último momento!

Uno y la persona amada, que parece aplazar su propia decisión, como uno mismo, están en el mismo barco. Si ambos decidieran pulsar el botón en el último instante, el desenlace sería fruto del azar. Uno de los dos apretaría el botón una fracción de segundo antes que el otro, o bien ambos se pasarían del tiempo límite y por tanto morirían. En suma, el desenlace se determinaría por casualidad. El azar primaría frente a la racionalidad.

En el ámbito de la filosofía, este tipo de situaciones ha despertado el suficiente interés para que se las bautice con el nombre de las *cajas de dilemas*. ¿Por qué tales dilemas llaman tanto la atención? En parte se debe a la curiosidad innata del ser humano ante cualquier situación novedosa. Pero no despertarían tanto interés si sólo fueran rompecabezas que no influyeran en la vida y existencia personal de los seres humanos. De hecho, surgen, y a veces dramática y angustiosamente en la vida real (el dilema nuclear es el caso más paradigmático), gracias a las diversas maneras con que los intereses individuales de unos se enfrentan y debaten con los de los demás y los de la sociedad en general. El caso es que cotidianamente se deben tomar decisiones difíciles, a veces con resultados distintos a los esperados. La cuestión esencial que se plantea es simple y apremiante: ¿existe un comportamiento racional para cada situación?

2. El dilema del prisionero

En 1966, Melvin J. Guyer y Anatol Rapaport, que entonces trabajaban en la Universidad de Michigan, clasificaron todos los juegos simples (Guyer, 1966). En estos juegos, dos antagonistas eligen entre dos alternativas posibles; es decir, son juegos de tipo 2×2 . Existen 78 juegos de 2×2 distintos, que surgen si los resultados sólo se jerarquizan, en vez de asignarles valores numéricos. El caso del *dilema del prisionero* es una de las 78 modalidades distintas de juegos bipersonales con dos estrategias. Es seguro que casi todas ellas deben surgir en la vida real. La mayoría de los 78 casos tienen solución. Si uno se centra en él es porque es un caso que plantea problemas. Tal vez el caso peor de este dilema sea la carrera de armamentos.

En un juego simétrico de tipo 2×2 sólo hay cuatro resultados relevantes. Sea *CC* el resultado, para cada uno de los jugadores en el caso de que ambos cooperen. *DD* para el caso de que ambos deserten; o sea, no cooperen. Cuando uno coopere y el otro deserte, el resultado del cooperador será *CD*, y el del desertor, *DC*. Al determinar un orden de preferencia

dado, entre los cuatro resultados posibles, queda definido un tipo de juego. Por ejemplo, sea la siguiente opción, expresada mediante las desigualdades: $DC > CC > DD > CD$. Este modelo es justamente el del dilema del prisionero.

Lo inquietante del dilema del prisionero es la forma en que el bien común pierde terreno frente a la racionalidad individualista, o mejor al egoísmo del ser humano. Cada participante quiere que el otro coopere, y sin embargo, le tienta a desertar. Esto tiene implicaciones. En términos generales, según la desigualdad, se prefiere el resultado CC al CD. Esto significa que a uno le irán mejor las cosas si el antagonista corresponde a la propia cooperación. Además, se prefiere DC a DD; es decir, uno espera que el adversario mantenga la cooperación aunque uno deserte.

En el mundo actual el dilema del prisionero es el dilema por antonomasia. Y esta es tal vez la principal razón de que sea tan estudiado (Shubik, 1970), (Straffin, 1990). Anthony De Jassay en un ameno ensayo sobre el estado y la lógica del poder político (De Jassay, 1992), afirma que la esencia del dilema del prisionero estriba en la fatalidad de la traición mutua, que es una cuestión de fechas. Dicha fatalidad constituye políticamente una teoría cuyos fundamentos psicológicos fueron explorados por Hobbes con bastante antelación al invento de la psicología.

Hobbes, que al parecer nació antes de tiempo por culpa del pánico que su madre tenía a la Armada Invencible, vivió obsesionado por la guerra civil, cuyo fantasma debió de inspirarle la atractiva teoría contractualista del estado. Ésta se basaba en el sentimiento de autopreservación: todo hombre necesita autoaprobación; por tanto, todo hombre ambiciona la superioridad sobre los demás; ergo, mi prójimo, en su busca de la prominencia, invadirá mi propiedad; por consiguiente, debo atacar la suya primero. El sentimiento de autopreservación conduce pues al enfrentamiento mutuo. He aquí la clase de situación que los teóricos de la decisión y los juegos denominan dilema del prisionero.

De hecho, la obra que se aproxima mucho al dilema del prisionero es el *Leviathán* de Tomas Hobbes (Hobbes, 1989). Su argumento es que la ley y el orden son beneficiosos para todos, no sólo para los legisladores, evitando la desertión. Hobbes argumentó que, en una sociedad sin leyes, el hombre sería lobo para el hombre, y nadie se libraría de caer bajo la explotación de los demás. Se podrían robar las cosechas al agricultor, los animales al ganadero o rancharo y los bienes a los demás y por tanto aquellos tendrían pocos motivos para sembrar los campos,

criar los animales y obtener los bienes lícitamente. Los miembros de la sociedad están en mejor situación si someten su *derecho al pillaje*; es decir, a desertar del bien común, a cambio de tener la seguridad de que no serán las víctimas; esto es, a conseguir un resultado que les recompense.

Es decir, Thomas Hobbes vio el contrato social como un resultado cooperativo exigido. Él después de señalar las desventajas del resultado no cooperativo, sugiere que las muchas decisiones independientes, para cooperar o no, deberían tomarse de la gente que constituye la sociedad. En efecto, ésta debe someterse a un arbitraje obligatorio y el gobierno debería jugar un papel de árbitro.

Según De Jassay, este dilema, implícito en el pensamiento hobbesiano de la guerra, es más natural y menos riguroso que el establecido bajo las convenciones de la teoría formal de juegos, y debiera tener, normalmente, una *solución cooperativa*. Pero no la tiene. Para empezar, el dilema requiere del estado de naturaleza. En este estado, el jugador, antes de realizar cualquier movimiento, puede pronunciar discursos o blandir sus armas. Dependiendo de la reacción del adversario, puede marcharse, si el adversario continúa en su sitio, o golpearlo, porque el adversario pretendía golpearlo primero o porque estaba mirando hacia otra parte; o tomar en consideración una oferta de protección pagada. Imagínense dos contrincantes: Azul y Rojo. Ambos persiguen la preeminencia en el sentido hobbesiano, con este orden de preferencias: victoria en la guerra, desarme, paz armada y derrota en la guerra. ¿Qué ocurre? Azul, que no sabe si Rojo se armará o desarmará, opta por armarse para salvarse de la derrota, con lo cual consigue la paz con coste y una posibilidad de victoria, si Rojo es tonto. Rojo, que razona como Azul, también escoge armarse. Total que Azul y Rojo escogen la paz armada o solución *maximin*; es decir, el mejor caso entre los peores, que es la solución adecuada para jugadores hostiles. Aunque ambos la hubieran preferido, han rechazado la paz sin coste, o *solución cooperativa*, sólo porque la victoria del uno sobre el otro constituía una preferencia mayor. En ausencia de un tercero con capacidad de coerción para imponer el desarme, la paz sin coste es inestable, con lo que el resultado del dilema de la guerra es habitualmente una paz armada, con guerras ocasionales. El mundo desde siempre.

Con todo reconforta saber que las últimas investigaciones sobre el dilema del prisionero indican que la aceptación de la coerción por los jugadores no es condición necesaria para dar con una *solución cooperativa*, siempre que los jugadores no sean estúpidos

ni carentes de previsión. Bastante avisados, los jugadores que tienen mucho mundo cooperan generalmente en los distintos *dilemas del prisionero*.

Más modestamente, el sociólogo George Simmel (Simmel, 1995) reconoció, ya en 1908, que los negocios competitivos se enfrentan frecuentemente con lo que equivale al *dilema del prisionero*. En su planteamiento, hay que destacar que Simmel considera al consumidor como un extraño, quien, aunque esté afectado por lo que hagan los negocios competitivos, no tiene control sobre lo que sucede. En efecto, el consumidor no es un jugador. Mientras la cooperación entre los negocios se traduce fácilmente en ventaja mutua, los efectos pueden ser antisociales en lo referente a la sociedad en conjunto. Y de ese modo la sociedad debe prohibir *juegos cooperativos* en forma de *trust*, *carteles*, fijación de precios, sobornos, etc. Esta cuestión, es decir, la tendencia de los competidores a evitar la competencia de precios mutuamente destructiva, también la trata John Kenneth Galbraith (Galbraith, 1952).

El problema es que la gente actúa con frecuencia irracionalmente. Merrill Flood, a la sazón trabajando en RAND (*Research and Development*) fue el primero que analizó esa irracionalidad desde la perspectiva de la teoría de juegos. Para ello, utilizó juegos curiosos, dilemas y regateos que se daban en la vida cotidiana. Flood informó sobre los resultados de algunas de sus investigaciones en una memoria de investigación (Flood, 1952).



Figura 3. La RAND.

Entre los dilemas de carácter práctico estudiados por Flood y considerados en su trabajo, sin duda el más importante era el tercero, denominado *A non-cooperative pair*, en español *Dos personas que no cooperan*. Siendo éste el primer estudio científico de lo que después se denominó *el dilema del prisionero*, y describe un experimento realizado en enero 1950

por Flood y su colega Melvin Dresher. Este experimento primigenio probablemente no permite comprender de la mejor manera el *dilema del prisionero*, ni es la mejor forma de aproximarse por vez primera al mismo; sin embargo, llamó la atención de la comunidad científica de RAND. Flood recuerda que von Neumann consideró que el juego era estimulante, visto de una manera general, pues cuestionaba la teoría del equilibrio de Nash, pero no se lo tomó muy en serio. Dresher enseñó el juego a otro consejero de RAND, Albert William Tucker. Éste era un distinguido matemático de Princeton y había sido maestro de Nash, el premio Nobel de economía conocido por la película y el libro “Una Mente Maravillosa”.



Figura 4. Flood y Dresher.

En mayo de 1950, el departamento de psicología de la Universidad de Stanford, pidió a Tucker que diera una conferencia sobre teoría de juegos. Él consideró, dado que el público tenía poca idea de teoría de juegos, que era más interesante presentar el juego que le había mostrado Dresher, desde una óptica más amplia. De este modo, decidió que necesitaba presentar el juego inmerso en una historia que se inventó. El relato con dilema incluido es el del *dilema del prisionero*, nombre que el propio Tucker acuñó.



Figura 5. Albert William Tucker.

El ajuste requerido por el juego fue relevante. De hecho, la memoria escrita por Flood, en 1952, era más bien clandestina, ni fue leída por muchas personas, ni tampoco anunciaba el descubrimiento de un nuevo dilema. Flood y Drescher expusieron el experimento del dilema del prisionero como un estudio psicológico. Fabricaron el juego para su uso en el experimento, y en el artículo no hay indicios de que otros juegos parecidos pudieran tener importancia en el mundo real, con independencia de la consciencia que los autores tuvieran de ello. Por eso, Tucker, al plantear el juego como un dilema en el que hay que elegir, y al difundir la historia a través de la comunidad científica, contribuyó mucho al posterior desarrollo de la investigación de los dilemas presentes en la sociedad. En efecto, en una carta a Drescher (Poundstone, 1992), Tucker explicó brevemente el dilema y lo denominó una versión arreglada de un juego como el que me mostró. Ésta fue su descripción:

Dos hombres, acusados de infringir conjuntamente la ley, han sido confinados por la policía en ubicaciones separadas. A cada uno de ellos se les dice que:

- 1) *Si uno de ellos se confiesa culpable, pero el otro no, el primero recibirá una recompensa... Y el segundo será castigado.*
- 2) *Si ambos confiesan, se castigará a los dos.*

Simultáneamente, ambos tienen buenas y fundadas razones para creer que:

- 3) *Si ninguno confiesa, ambos quedarán libres.*

A lo largo del tiempo, la historia se ha adaptado y mejorado. Por ejemplo, actualmente (Davis, 1971), se plantea casi siempre como una cuestión de indulto referente a una condena en la cárcel. Sin duda es más realista negociar el indulto de una sentencia de prisión que ser recompensado en efectivo por confesar la culpa. Por ejemplo, una matriz de la situación podría ser la dada en la figura 6.

	B rechaza la propuesta (Coopera)	B aporta las pruebas (Deserta)
A rechaza la propuesta (Coopera)	1 año, 1 año	5 años, 0 años
A aporta las pruebas (Deserta)	0 años, 5 años	3 años, 3 años

Figura 6. Matriz del dilema del prisionero.

Ambos prisioneros son informados de que al otro se le está haciendo la misma propuesta. A cada implicado, le preocupa sólo su propio bienestar, además de tratar de disminuir, al máximo posible, su propia condena. De este modo, los prisioneros podrían razonar como sigue: *si testifico, pero mi compinche no lo hace, me libero de la cárcel, en lugar de cumplir 3 años de condena. Si canto y el otro también, ambos cumplimos tres años, en vez de 5. En ambos casos, debo confesar.* El problema es que el compinche llegará a la misma conclusión y ambos serán condenados a 3 años, en lugar de sólo a un año, que es lo que ocurriría si ambos no declararan.

Aparentemente, en cualquiera de los dos casos, ¡sale mejor librado confesando! Entonces, ¿dónde radica el problema? He aquí la paradoja. Dos delincuentes honorables irían a la cárcel sólo por un año. Ésta es la posición del Rey Arturo, que, como él era incapaz de engañar a Ginebra, no podía concebir que Ginebra lo engañara a él. El hecho es que Ginebra lo engañó. Por tanto, dos delincuentes ¿inteligentes? o por mejor decir, astutos, teniendo en cuenta que la bondad de la gente es escasa, y siguiendo la máxima de los amigos de lo ajeno que dice: *se cree el ladrón que todos son de su condición*, a la vista de la matriz de pagos, confiesan de plano ambos y se llevan tres años de prisión.

La cosa parece obvia, pero no lo es tanto si se consideran los elementos escondidos de esta cuestión. En efecto, cada uno de ellos tiene dos elecciones básicas:

- 1) Puede actuar cooperativamente; es decir, no declarar.
- 2) No actuar cooperativamente; esto es, acusar al otro; o sea, desertar.

En el primer caso, cada uno consigue mejor resultado que cuando ambos optan por la segunda opción. Es decir, para cualquier estrategia fijada por el primer delincuente, siempre consigue más tomando la opción de desertar; esto es, no actuar cooperativamente. Incluso, en un sistema tan garantista como ingenuo, por decirlo suavemente, como el español.

El único problema de representar el dilema del prisionero en forma de historia, es que aparecen factores emocionales que nada tienen que ver con la cuestión. Por eso el dilema acostumbra a plantearse cómo un juego amistoso sin consecuencias de ningún tipo. A nadie le va a sentar mal que alguien coopere, deserte o abandone. Imagínese que se está en un casino en una mesa donde se juega al dilema del prisionero. Se apuesta para ganar dinero en metálico,

entre dos contrincantes. Cada jugador decide su opción mediante un interruptor oculto, cuando así lo indica el “crupier”. Los dos jugadores no pueden discutir entre ellos su decisión. Además, las reglas de la casa prohíben que dos jugadores participen más de una vez como pareja contrincante. No se considera inmoral ganar lo más posible, como tampoco lo es jugar de la mejor forma admisible al póker, a las veintiuna, en inglés *black jack*, o a cualquier otro juego. Ser un jugador honrado exige jugar a ganar. Los resultados, contabilizados en euros, dólares, libras, fichas o la unidad de apuesta que se prefiera, vienen dados en la tabla de la figura 7.

	B coopera	B deserta
A coopera	4,4	0,5
A deserta	5,0	2,2

Figura 7. Dilema del prisionero.

Lo que se plantea es determinar cuál es la mejor estrategia posible, que le sirve sólo a uno, si el adversario también sólo vela por sus propios intereses. Se debe razonar la respuesta, sea la que sea, con el fin de mejorar las ganancias. No es de recibo cooperar porque eso es lo que está bien. Asimismo, la mejor salida es desertar, independientemente de lo que haga el contrincante. Uno gana 5 unidades en vez de 4, si el otro coopera. Y si no, se lleva 2 unidades, si el otro deserta, en vez de nada. Así de simple. Es decir, ¡el casino le da dos unidades para que deserte! En suma, si ambos adversarios, seres racionales, se percataran de esto, los dos desertarán, y por consiguiente ambos ganarán 2 unidades. En el caso de haber sido menos lógicos, habrían cooperado y se habrían llevado el doble.

Para que una situación o juego pueda calificarse como un dilema del prisionero, sólo es necesario que los resultados de las diferentes elecciones se atengan a una clasificación determinada. En general, un dilema del prisionero tiene la estructura siguiente: existe un resultado recompensa o premio, en el ejemplo 4 unidades, en el caso de que exista cooperación entre ambos adversarios. Este resultado es más conveniente que el resultado de castigo; en el ejemplo, 2 dólares otorgados a ambos si se trata de la situación de no colaboración. Sin embargo, ambos codician el máximo de la tentación, en este caso 5 unidades, que tiene lugar cuando se produce una deserción única. Este resultado proporciona más beneficio que el resultado recompensa. Ambos contrincantes temen ser el que no deserte, y que por tanto le corresponda el resulta-

do de hacer el primo o perder; en este caso, no llevarse nada.

Al analizar el dilema del prisionero, se ve claramente que si el juego se lleva a cabo sólo una vez, los delincuentes no tienen otra opción que confesar. Pero cuando se juega repetidamente, que es lo que importa desde la perspectiva científica, el argumento pierde fuerza. Dicho de otro modo, al considerar repetidamente el dilema del prisionero; esto es, no un número determinado de veces, sino indefinidamente, es cuando la estrategia cooperativa, o sea, la honorabilidad entre delincuentes, se muestra contundente. Y éstas son precisamente las condiciones bajo las cuales se plantea, y surge a menudo, dicho dilema.

El dilema del prisionero tiene, pues, una característica que lo hace peculiar y paradigmático con respecto a otras situaciones similares. Como norma, al analizar una situación competitiva, uno está satisfecho si puede decir lo que harían contrincantes racionales y, en consecuencia, predecir el resultado. Sin embargo, en el dilema del prisionero la estrategia cooperativa es tan poco atrayente que la cuestión que trataría de contestar la mayoría no sería: ¿qué estrategia elegiría una persona racional?, sino: ¿cómo se puede justificar el seguir una estrategia cooperativa?, he ahí la cuestión.

Al expresar los resultados en unidades de valor, se requiere que el resultado premio sea de más cuantía que la media obtenida a partir de los resultados tentación y perder unidos. La cuestión que plantea el dilema del prisionero surge con toda su crudeza al exigir que se obtenga el beneficio común mediante la cooperación. De ahí el comportamiento singular de los jugadores lógicos, pues se juegan el tipo al desertar. En efecto, en la figura 7, los jugadores ganan una cantidad total de $5 + 0 = 5$, si uno coopera y el otro deserta. Es decir, se llevan una media de 2,5 unidades cada uno. Si ambos cooperan, cada uno se lleva 4 unidades. Pero, caso de que la media de los resultados tentación y perder unidos fuera mayor, los adversarios podrían decidir desertar por separado en partidas sucesivas y así, a la larga, ganar más que si colaboraran. Sin embargo, esto es imposible en un verdadero dilema del prisionero.

Todas las particularizaciones del dilema del prisionero finalizan con una contradicción paradójica. Siempre acaba uno preguntándose si ha obrado adecuadamente, sin que importe cómo lo haya hecho. Entonces, la cuestión que se plantea es la siguiente: ¿cuál es la manera adecuada de actuar en el dilema del prisionero? En general, esta pregunta aún no tiene una respuesta, y es probable que no exista. R. Dun-

can Luce y Howard Raiffa, escribían al respecto lo siguiente (Luce, 1957): *la impotencia que se siente ante juegos de este tipo no puede soslayarse insistiendo en el uso de palabras tales como racional e irracional, lo cual es, no obstante, inherente ante la situación planteada.*

En fin, para Poundstone (Poundstone, 1992), a quien se va a seguir en lo que continúa, el dilema del prisionero es difícil porque desafía al sentido común. En efecto, el argumento de sentido común para desertar es el siguiente: un dilema del prisionero es una decisión tomada simultáneamente por dos personas. No existe ningún modo de influir con su decisión en la del otro jugador. Por consiguiente, la solución es sencilla. Haga lo que haga el otro, lo mejor es desertar. Los partidarios de desertar incluso tienen a mano una reputación de lo primero que van a alegar los partidarios de cooperar. Estos preguntarán: ¿qué sucedería si todos actuaran de esa manera? Los desertores replicarían: ¿cómo pensáis, los de la cooperación, que sería una lástima si ambos adversarios desertaran en vez de haber cooperado? ¡Os equivocáis! Recordad que las decisiones no pueden influirse mutuamente. Si el contrincante deserta, lo hace *motu proprio*. La decisión del adversario no habrá tenido nada que ver con eso. Siempre que haya una deserción mutua uno se felicitará de la deserción tomada. Si hubiera cooperado, habría hecho el primo.

El argumento de sentido común para cooperar es como sigue: las situaciones de ambos implicados son idénticas. No es realista que uno de ellos piense aprovecharse del otro mediante la deserción. Si se supone que ambos participantes son racionales, deberán optar por la misma estrategia. Las dos soluciones realistas son la cooperación mutua y la deserción de ambos, y preferirían el resultado de la cooperación. Por esta razón tienen que cooperar. Ahora bien, si se reflexiona un poco se verá que esta argumentación no es muy consistente. En la vida real, nadie garantiza que los implicados vayan a tomar la misma decisión. Unos dejan tirados a sus colegas; otros no. Y la mayoría engaña mediante subterfugios no permitidos; sin embargo, algunos se mantienen fieles a la regla. Por lo tanto, lo práctico es admitir que los cuatro resultados son posibles. El razonamiento más interesante es el que postula la teoría de juegos, usado por los participantes perfectamente racionales. Se puede dar por sentado que sólo hay una manera racional de actuar en el dilema del prisionero. Por tanto, sólo son posibles las estrategias con elecciones iguales: que ambos cooperen o deserten, para un dilema planteado entre seres racionales.

El único punto flaco de la argumentación aparece al suponer que los participantes puedan tener la oportunidad de juzgar cuál de esos resultados es el racional. Supóngase que un millonario excéntrico ofrece a dos reputados matemáticos situados en los antípodas, un premio de un millón de euros multiplicado por el valor numérico del millonésimo dígito del desarrollo decimal del número pi. Este dígito podrá ser cualquiera de los dígitos decimales, del cero al nueve. Por tanto, el premio oscilará entre cero y nueve millones de euros. Lógicamente, los matemáticos, aunque trabajan cada uno por su cuenta, llegarán al mismo valor para el dígito millonésimo de pi. Por supuesto que no influye para nada el hecho de que les gustaría que fuera nueve. Asimismo, no tiene nada que ver que unos participantes perfectamente racionales prefiriesen, por lógica, cooperar. Lo que se plantea es qué acción les exige la lógica que tomen.

Hay otro razonamiento a favor de la cooperación y reza así: el mejor resultado posible es la cooperación mutua frente a los demás. Así pues, lo mejor es estimular la cooperación mutua; el único modo de realizar esto es cooperando. Incluso aunque uno resulte perjudicado esta vez, a la larga es la mejor política. Esta argumentación es bastante aceptable, pero no en la situación aquí planteada. En efecto, si se jugara un dilema del prisionero reiteradamente con el mismo adversario, habría que ocuparse del largo plazo; en este caso, la opción de cooperación ganaría mucha más plausibilidad. Sin embargo, aquí y ahora, se examina el caso del dilema del prisionero que se plantea una única vez. Hay que hacer la apuesta y a lo he hecho pecho. Ya puestos, lo mejor es llevarse cuanto se pueda ganar. En definitiva, dado un auténtico dilema del prisionero, jugado una sola vez, es igual de difícil justificar como resultado racional tanto la cooperación como la deserción mutua. Ahí reside la paradoja.

Tanto Flood como Dresher esperaban que alguien solucionara el dilema y consiguiera una teoría de juegos de suma no nula nueva y mejor. Esta teoría abarcaría el conflicto entre la racionalidad individual y la colectiva, ejemplificado por este dilema. El caso es que no fue así y el dilema permanece como una conclusión negativa, la demostración de lo que falla en la teoría, y desde luego en el mundo real.

Todas las discusiones acerca del dilema del prisionero serían meramente académicas si él fuera un ejemplo excéntrico de la teoría de juegos. Sin embargo, no lo es. El dilema del prisionero es una paradoja con la que los humanos tienen que convivir. Darse cuenta de este dilema es parecido a descubrir que el

aire existe. Siempre ha acompañado a la humanidad, y la gente siempre lo ha notado en mayor o menor medida.

Desde la más remota antigüedad es bastante común encontrarse con recomendaciones éticas basadas en conflictos del tipo del dilema del prisionero. De Confucio (551 - 479 aC) a Séneca (4 aC - 65 dC), pasando por Platón, Aristóteles o Hillel, dieron versiones de lo que después fue considerada como la regla de oro de Jesucristo. En efecto, en el evangelio según San Mateo (a7,12) (San Mateo, 1960) se da dicha regla como sigue: *así que todas las cosas que queráis que los hombres hagan con vosotros, así vosotros también haced con ellos; porque esto es la ley y los profetas*. Probablemente, no es una interpretación demasiado tergiversada afirmar que la regla de oro se refiere a conflictos basados en el dilema del prisionero. Las personas suelen mirar por sus propios intereses. Ciertamente, esta regla sería superflua si se diera cotidianamente o incluso con cierta asiduidad la norma, realmente impopular, que obligase a olvidarse y, o, soslayar el interés propio obvio en aras de un beneficio para todos, posible sólo cuando la gente coopera.

Consejos semejantes existen en la *Crítica de la razón práctica*, de Kant (Kant, 1788), agrupados bajo el nombre de *imperativo categórico*. Kant había llegado a la conclusión de que el comportamiento ético es aquel que puede llevarse a cabo por cualquier persona; es decir, que puede ser generalizado. Es lo mismo que preguntarse siempre: ¿qué pasaría si todo el mundo hiciera lo mismo?

Problemas del tipo del dilema del prisionero, en una u otra forma, han estado por todas partes durante mucho tiempo. La estrategia cooperativa es aceptable, generalmente, como la *idónea*, excepto cuando su efecto es antisocial, a veces por razones éticas. Kant afirmaba que una persona debería decidir si su acto es o no moral, examinando el efecto de todos actuando análogamente; la regla de oro dice casi lo mismo.

En puridad, ninguno de los casos anteriores pueden considerarse precursores del dilema del prisionero; sin embargo, esta tradición es en gran parte similar a la noción intuitiva de la mayoría de las personas de que cooperar es lo correcto y desertar lo incorrecto.

La diferencia entre las suposiciones hechas en los juegos de suma cero y de suma no nula, es aún más clara cuando fallan. Y no hay duda de que la gente fracasa a menudo en jugar cooperativamente en los

dilemas del prisionero de sociedad. En los juegos de suma no nula, cooperar con un compañero que no lo hace con uno conduce al desastre; en los juegos de suma cero, lo peor que puede ocurrir cuando se juega *minimax* es que se pierde la oportunidad de *timar* al contrincante.

3. Otras versiones del “dilema del prisionero”: Marie Roget y Tosca

Símiles muy parecidos al dilema del prisionero son, no obstante, los dos que se van a considerar a continuación, también de la literatura. El primero, debido a Edgar Allan Poe (Poe, 1991), en su relato *El Misterio de Marie Roget*, basado en gran parte en un caso de asesinato ocurrido en New York, en 1842. En él hay una discusión lúcida del dilema, ique además se ciñe bastante bien a la historia de Tucker! En efecto, el detective de Poe C. Auguste Dupin, menciona que se ofreció una recompensa y la garantía de inmunidad al primero de un grupo de delincuentes que confesase su delito. *En esta situación, cada uno de los miembros de la banda no codicia la recompensa, ni ansía escapar, tanto como temen que le traicionen los demás. Trata desesperadamente de traicionar a los otros lo más pronto posible, para así no ser traicionado por los demás*. Como nadie optó por aceptar el trato, según Poe, eso demostraba que el asesinato no fue realizado por la banda.

El caso de Tosca es para Rapoport (Rapoport, 1962), paradigmático al respecto. En la ópera *Tosca* de Giacomo Puccini, compuesta en 1900, y cuyo libreto se basa en una obra teatral escrita, en 1887, por Victorian Sardou, el argumento gira en torno a un claro ejemplo de dilema del prisionero. Y, como es sabido, al desertar ambos contendientes, Tosca y Scarpia, la cosa terminó de la peor manera posible. Más en concreto, la historia es como sigue: el jefe de la policía, Scarpia, ha condenado a muerte al amante de Tosca, Cavaradossi, pero ofrece salvarle a cambio de los favores de Tosca. Tosca acepta el trato, y se decide que la ejecución de Cavaradossi será fingida. Como todo el mundo sabe, Tosca y Scarpia se engañan mutuamente de modo que ella lo apuñala cuando él va a abrazarla, y él no había dado al pelotón de fusilamiento la orden de disparar con balas de fogueo. En suma, que la cosa termina incluso peor que *el rosario de la aurora*. De lo que se trata, entonces, es de saber si realmente lo mejor que pudieron hacer fue el engañarse mutuamente.

Para ello, asígnense valores numéricos a cada uno de los desenlaces posibles de acuerdo con su valor para Tosca y Scarpia, tal y como se muestra en la figura 8. Estos valores, aunque arbitrarios, presentan la situación de un modo razonable. Si ninguno de los dos rompe su compromiso, la satisfacción de Tosca por recuperar a su amado y amante se verá empañada por su entrega al odiado Scarpia. A su vez, la satisfacción de Scarpia por haber conseguido a su admirada Tosca se verá oscurecida por el hecho de tener que perdonar la vida a uno de sus más aborrecibles rivales. Por otra parte, si Tosca engaña a Scarpia y Cavaradossi salva la vida, la ganancia para ella será máxima (+10) y la pérdida de Scarpia también será máxima (-10).

		Tosca	
		S _M	S _E
T	S	T _m +5	T _e -10
	T	T _m +10	T _e -5

(a)

		Scarpia	
		S _M	S _E
T	S	T _m +5	T _e +10
	T	T _m -10	T _e -5

(b)

Figura 8. Formas de razonar de Tosca (a) y Scarpia (b).

Si se engañan mutuamente, ambos salen perdiendo, pero no pierden tanto como perderían si hubieran sido engañados sin engañar ellos a su vez. Por ejemplo, el difunto Scarpia, se supone, se regocijaría en su tumba, al pensar en lo que iba a sentir Tosca cuando al correr hacia su amado caído encuentre que verdaderamente ha sido fusilado. ¿Por qué Tosca no mantuvo el trato? Véase el razonamiento de Tosca para hacer lo que hizo. Tosca no tiene la más mínima duda acerca de la integridad de Scarpia, pero tampoco está segura de lo que éste va a hacer. Por tanto, considera las dos posibilidades: si él mantiene su palabra, lo mejor es engañarle, puesto que, de este modo, tendrá a Cavaradossi sin Scarpia; mientras que si no le engaña, tiene a Cavaradossi con Scarpia. Pero es que si Scarpia la engaña, es evidente que lo mejor es engañarle. Ergo, la conclusión más obvia es matarle, pase lo que pase. Por su parte Scarpia razona exactamente de la misma manera. Si Tosca mantiene la palabra, lo mejor es engañarla, puesto que de ese modo acaba con Cavaradossi; en otro caso, tendrá que enfrentarse a él. Y si ella lo engaña es justo que se vengue. Por consiguiente, la ejecución es lo adecuado. De este modo, ambos consiguen -5, en tanto que si cada uno hubiera mantenido lo pactado, ambos habrían ganado +5.

Las limitaciones del pensamiento estratégico se ponen claramente de manifiesto en este ejemplo. Evi-

dentemente, aparte de calcular las ganancias, hace falta algo más para tomar la decisión adecuada en una situación conflictiva. La teoría de juegos también es capaz de tratar este caso de un modo satisfactorio mediante la idea de coalición. En efecto, si Tosca y Scarpia se dieran cuenta de que ambos saldrían beneficiados si mantienen el trato acordado, y fuesen tipo *Rey Arturo*, no terminarían perdiendo. Sin embargo, las coaliciones también dan lugar a quebraderos de cabeza, pero, eso son otros lópeces.

Lo positivo de la teoría de juegos no reside en las soluciones específicas que puede ofrecer en situaciones altamente simplificadas e idealizadas, que si bien pueden darse en juegos formalizados, no es fácil que se den en la vida real. El valor fundamental de esta teoría está en su capacidad de poner claramente de manifiesto los distintos tipos de razonamiento que se aplican en los diversos tipos de conflictos.

De este modo, en el enfrentamiento entre Tosca y Scarpia, juego de suma no nula, ambos tienen el mismo *minimax* que, en efecto, escogen. Pero el resultado es malo para ambos. La respuesta es que ambos estaban considerando la situación como si fuese de suma nula; es decir, que lo que un jugador gana necesariamente lo pierde el otro. Pero basta examinar la matriz para ver que esto no es así. Ambos, podrían haber aumentado sus ganancias eligiendo, en vez de la solución *minimax*, la solución de coalición; mantener el trato y ganar +5 cada uno. Lo que muestra Tosca y otros juegos, es que las decisiones basadas en el cálculo del propio interés pueden llevar a la catástrofe. Lo más importante de la teoría de juegos radica, para Rapoport, en que el análisis de la misma es capaz de poner de manifiesto sus propias limitaciones.

Por poner un ejemplo de rabiosa actualidad, dada la voracidad del gobierno en lo que concierne e implica a la recaudación de impuestos. Si nadie pagara impuestos, se desmoronaría el Estado y, por supuesto, el gobierno caería en la inanición. Presumiblemente, cada contribuyente, preferiría que todos paguen impuestos, incluido el mismo, ¡faltaría más!, a que nadie los pague. Aún mejor, desde luego, si todos pagaran impuestos, excepto él mismo. A continuación se van a dar dos ejemplos como botón de muestra real. El primero, apócrifo; es decir, no es verdad, pero merecería serlo. El segundo, histórico y muy conocido en España.

Era Vinolandia un país sin grandes problemas y con un monarca verdaderamente preocupado por el bienestar de sus ciudadanos y muy sensible a la cuestión de la justicia distributiva. Así pensó en mejorar la

calidad de vida de los habitantes de su reino reduciendo, ecuánime y equitativamente, la presión fiscal. Para ello, se le ocurrió que si cada una de las familias de Vinolandia, todas ellas poseedoras de excelentes viñedos y bodegas, en mayor o menor medida, aportaba un litro por cada 100 litros que obtuviera se obtendrían los suficientes litros del mejor vino del mundo, cuya venta sufragaría los gastos de todo tipo, incluyendo naturalmente los sociales (sanidad, educación, etc.), de inversión (obras públicas, industriales, etc.) y demás, con creces. De este modo, publicó un decreto por el cual abolía los impuestos, a cambio de que cada Vinolandés aportara su cuota parte en botellas de vino. La noticia fue muy bien recibida y llenó de júbilo a los Vinolandeses.

El día señalado para recaudar el vino estaban los habitantes de Vinolandia haciendo cola para verter su(s) litro(s) de vino en las grandes barricas construidas para ese menester. Uno por uno fueron vertiendo su vino, hasta que se llenaron varias cubas. El rey, que se personó en la bodega al final de la jornada, satisfecho por el resultado obtenido decidió brindar por el aparente éxito y solicitó su copa real de catas del mejor cristal de Bohemia. Se acercó a la primera cuba, llenó su copa y, como experto catador, miró el vino para ver su color, pero algo iba mal, pues era incoloro. A pesar de ello, agitó la copa y la acercó a la nariz para disfrutar de las esencias y aromas de la excepcional bebida. Y, ¡oh sorpresa!, algo no encajaba, pues el supuesto mejor vino del mundo era inodoro. Finalmente, bebió un sorbo y comprobó que el líquido de la copa era insípido. Probó en las demás cubas y obtuvo el mismo sorprendente resultado.

Consultado el enólogo real sobre este fenómeno tan extraño la respuesta, inmediata y clara, fue: *Majestad, es agua. No es cuestión de brujería ni alquimia, es simplemente que sus súbditos son humanos, demasiado humanos. Esta mañana cuando me dirigía hacia aquí escuché a uno que le decía a su mujer que en lugar de traer unos litros de su buen vino, traía unos litros de agua, dado que entre los miles de litros del mejor vino del mundo no se notaría. El problema es que todos pensaron lo mismo.*

Y ahora la historia real. El conde de Romanones era un político poderoso y filántropo benefactor de la mayoría de los miembros de la RAE, y escritor si no excelente, sí digno. Varios miembros de la RAE, que habían sido beneficiados con las donaciones y dádivas del conde, para darle coba y eventualmente aligerarle la cuenta corriente, le instaron a que presentara su candidatura a la RAE y como todos o la mayoría de sus miembros le debían favores mil, su elección

parecía estar más que asegurada. Dicho y hecho. Presenta el conde su candidatura y el día señalado se celebra la preceptiva votación y, ¡oh sorpresa!, el conde no tiene ningún voto a favor. La razón es que todos votaron en contra esperando que alguien, al menos uno, votara a favor y ese voto sería el suyo. Pero como todos pensaron igual el resultado los puso en evidencia. La respuesta del conde ante la esperpéntica situación fue lo mejor de la misma. Simplemente, se limitó a definir, clara y precisamente, a los miembros desagradecidos de la RAE, diciendo: ¡joder, que tropa!

4. Conclusiones

No hay indicio alguno de que von Neumann se hubiera percatado del dilema del prisionero cuando escribió sus trabajos sobre juegos. Sin embargo, pensaba de un modo similar en temas de semejante índole. Éstas fueron sus palabras (Neumann, 1944): *imagínese que se ha descubierto un conjunto de reglas aplicables a todos los participantes, llamadas óptimas o racionales. Estas reglas son desde luego óptimas, en el caso de que los demás participantes se atengan a ellas. Queda por ver lo que sucedería si uno de los participantes no siguiera las reglas. Si de aquí resultan consecuencias beneficiosas para ellos, y en particular perjudiciales para los que cumplen las reglas, se pondría muy en duda la solución mencionada anteriormente... Independientemente de la manera en que formulemos los principios básicos y la justificación objetiva de un comportamiento racional, habrá que tener en cuenta las condiciones que conllevan todos los comportamientos posibles de los otros.*

Por consiguiente, puede decirse que el dilema del prisionero ha sido descubierto, estudiado y olvidado numerosas veces, y normalmente sin ver que se trataba de un problema de alcance universal. Robert Axelrod comentó al respecto (Poundstone, 1992): *creo que se asemeja a comprender el papel que juegan los conceptos de temperatura y calor en física. Puedes decir: hoy hace calor, o bien: este vaso de agua contiene más calor que un vaso más pequeño que está a la misma temperatura. Sin embargo, no podrán plantear las cosas de manera precisa a no ser que sepan distinguir correctamente entre calor y temperatura. En el dilema del prisionero puedes decir que hay un conflicto entre los intereses individuales y del grupo; pero no llegarás mucho más lejos sin el apoyo de la teoría de juegos.*

No es muy difícil crear un dilema del prisionero, ¡de hecho surgen a manta de Dios en la sociedad

actual. El ingrediente esencial es generar la tentación de mejorar los propios intereses, de una forma que llevaría al desastre si *todo el mundo* lo hiciera. Por desgracia, este ingrediente fundamental es fácil de encontrar. Por eso algunos han interpretado el *dilema del prisionero* como el problema fundamental de la sociedad: el problema del *mal*. Las verdaderas tragedias acontecidas a lo largo de la Historia son causadas por el hombre, no las naturales; es decir, las que son consecuencia de las acciones de individuos o grupos contrarios al bien común.

En su forma más cotidiana y común, el *dilema del prisionero* recibe el nombre del *dilema del viajero gratis*; esto es, los que, en cualquier momento y lugar, se cuelan sin pagar. Claro que si todos hicieran lo mismo, no habría metro, ni ferrocarril, etc. Lo mismo sucede cuando se choca a un vehículo y no se le deja una nota con los datos. O si en un puesto de periódicos basado en el *código de honor* uno se lleva el diario sin pagar. En general, el *dilema del viajero gratis* se plantea siempre que el pago o retribución por un producto o servicio se basa en el sistema de *código de honor*; o bien si el pago, en principio obligatorio, es difícil de exigir en la práctica; verbigracia, los que se cuelan en los espectáculos, los que piratean cosas, lo que defraudan a hacienda, etc.

El *dilema del viajero gratis* aún tiene menos perspectivas de solución que el *dilema del prisionero* jugado entre dos personas. Ya no hay quien coopere o abandone a un único compañero de partido. Entran en juego otros muchos participantes. Podría decirse que en un *dilema del viajero gratis*, los desertores abundan entre la muchedumbre. Partiendo de que el comportamiento psicológico del ser humano es variable, no hay que dudar que muchos desertarán y se colarán. Una desertión como la del *dilema del viajero gratis* es una, pero no la única, ni siquiera la más matemática, de las típicas argumentaciones en contra del socialismo marxista o la Utopía de Tomás Moro (Moro, 2003). Si todo el mundo trabajara mucho y se compartieran inteligentemente los bienes, a nadie le faltaría lo esencial, ni se moriría de hambre. Sin embargo, simultáneamente todos tendrían la tentación de holgazanear, pues aún así comerían; su falta personal de implicación no pondría en gran peligro la colectividad. Ahora bien, si todo el mundo se comportase así la comuna fracasaría y la gente se moriría de hambre. Como así ocurre en los países comunistas.

La decisión de aceptar las exigencias de terroristas y secuestradores es, asimismo, una modalidad del *viajero gratis*. Si nadie pagara rescates no habría secuestros. Sin embargo, la realidad es otra. Véanse

los últimos secuestros de españoles: Alakrana, cooperantes, etc. Ahora bien, puede que esté mal colarse en el metro o defraudar a hacienda, pero pocos considerarán inmoral pagar un rescate para salvar a un inocente. En este caso, los únicos que exhiben mal comportamiento son los secuestradores.

Los problemas de la sociedad nunca son fáciles de resolver. Muchos debates sobre temas políticos y militares se enredan en contingencias e imprecisiones, de modo que es posible pasarse toda la vida discutiendo los detalles. Se tiene la sensación de que si se solucionasen estos detalles secundarios, ya no existiría el problema, y surgiría un acuerdo común sobre la línea de acción a seguir. No tiene por qué ser así. Hay muchos casos en los que el dilema principal es auténtico, y aparentemente no tiene solución. Aunque un problema de carácter social plantee un *dilema del prisionero*, no por ello dejará de exigir decisiones angustiosas aun en el caso de que todos los temas secundarios hayan sido debatidos, incluso superados. No habrá una solución *correcta*, siempre existirán puntos de vista razonables pero encontrados.

En el ejemplo del *dilema del prisionero* los pagos se establecen en términos de *años de prisión* o de beneficios netos, más bien que en *utilidades*. Las *utilidades* reales implicadas son realmente sugeridas por aseveraciones tales como: *a cada jugador sólo le concierne su propio interés*, o *cada empresa sólo necesita maximizar sus propios beneficios*. Sin embargo, teniendo esto presente, en el fondo hay ciertos supuestos que, aunque expresados sólo inexactamente, son críticos. Si no son válidos se puede estar jugando a un juego radicalmente distinto del que se piensa. Por ejemplo, si en el *dilema del prisionero* originario, uno de los implicados prefiriese pasar un año en la cárcel junto con su compinche, mejor que ser puesto en libertad, sabiendo que aquel cumplía cadena durante 20 años, el argumento para confesar se desmorona. Pero entonces, difícilmente, puede denominarse *dilema del prisionero*. De hecho la paradoja original no ha sido resuelta. Si un jugador está tan interesado en el pago de su compañero como en el suyo propio, el juego no es un *dilema del prisionero*.



Figura 9. El generalísimo Radetzky.

Por otra parte, las analogías entre las suposiciones que se hacen en juegos de suma cero y en los de suma no nula, tampoco es muy convincente. En juegos de suma cero, puede obtenerse el valor del juego si el oponente es bueno, malo o indiferente; no hay que suponer que sea racional. Claro es que jugando *minimax* se pierde la oportunidad de explotar los errores, y contra un oponente tonto uno no debería conformarse con obtener sólo el valor del juego. En efecto, éste se basa en el principio Radeztky (Pazos, 1987), el invencible generalísimo austriaco a quien Strauss le dedicó la famosa marcha con la que finalizaba el concierto de primero de año de la filarmónica de Viena, y dice: *no consideres a los demás por más estúpido de lo que tú te creas*. En suma, que usando este principio se es eficaz pero no eficiente. Sin embargo, esto no es totalmente cierto, puesto que hay juegos, en los que una estrategia inferior por parte de un jugador y la estrategia de *minimax*, por parte del otro, conducen a un pago que es mayor que el valor del juego para el jugador del *minimax*. Pero incluso en estos juegos en los que jugar *minimax* impide obtener más que el valor del juego, la analogía es discutible. Para explotar la debilidad del oponente no es suficiente saber que se desviará del *minimax*; hay que saber también cómo. Supóngase, verbigracia, que se juega a parear monedas con un lelo y que se supone que no jugará racionalmente. Concretamente, se supone que es muy capaz de jugar a una cara de la moneda con probabilidad mayor que un medio, ¿pero de qué lado? Y ¿cómo puede explotarse su estrategia inferior si no se sabe cuál es? Como norma, cuando se juega una estrategia *minimax* en un juego

de suma cero, no es porque se tenga fe en la racionalidad del oponente, sino porque no se tiene otra alternativa más atrayente, y esto es cierto cuando se sospecha que el oponente es capaz de cometer un error.

Sin embargo, en el dilema del prisionero, el asumir que la pareja cooperará es realmente una suposición. A menos que uno sea un masoquista, si se juega cooperativamente, debe creerse como un acto de que también jugará así el compañero. Incluso si éste coopera y aparentemente justifica el acto de fe, algunos jugadores dudarán todavía en su elección, puesto que se podría haber logrado mejor resultado jugando no cooperativamente. Esta actitud puede parecer egoísta, pero los jugadores no buscan a los teóricos de juegos por sus principios morales; ellos ya tienen los suyos propios. Todo lo que quieren es encontrar una estrategia que sirva para sus propósitos, egoístas o no.

Por último, pero no de último, señalar que el dilema del prisionero es, siguiendo la terminología de Francis Bacon, no sólo “lucífero”; es decir, aclara y explica algunas situaciones dilemáticas sociales, sino que además y sobre todo es fructífero. En efecto, a partir de él se derivan otros dilemas sociales como son: “¡Gallina!”, “El Voluntario” o el atolladero. Y lo que es más importante, de acuerdo con el físico teórico Willian A. Newcomb, es posible plantear la paradoja que, en su honor, lleva su nombre absolutamente relevante para afrontar, nada más y nada menos, que el problema del libre albedrío. Pero como diría Rudyard Kipling esa es otra historia.

Bibliografía

- Bascon, W.: *African Dilemma Tales: An Introduction*. In Dorso, R.M. (Ed.): *African Folklore*. Indiana University Press. Bloomington. Indianapolis, 1972.
- Davis, M.: *Teoría de Juegos*. Alianza Editorial, S.A. Madrid, 1971.
- De Jassay, A.: *El Estado. La Lógica del Poder Político*. Alianza Editorial, S.A. Madrid, 1993.
- Flood, M. M.: *Some Experimental Game*. RM-78M. RAND Corporatiun. Santa Mónica, CA, 1952.
- Galbraith, J. K.: *American Capitalism. The Concept of Countervailing Power*. Houghton Mifflin Co. Boston, Mass, 1952.
- Guyer, M. J. and Rapaport, A.: *A Taxonomy of 2x2 Games*. General Systems, 11. 1966. PP: 203-214.
- Hobbes, T.: *Leviatán o la Materia, Forma y Poder de un Estado Eclesiástico y Civil*. Alianza Editorial, S.A. Madrid, 1989.
- Kant, I: *Kritik der Praktischen Vermunft*. Fr. Hartknoch. Riga, 1788. En español, *Crítica de la Razón Práctica*. En Kant, I.: *Fundamentos de la Metafísica*. Ed. Parrúa. México, 2000.

- Luce, R. D.; Raiffa, H.: *Games and Decisions*. Introduction and Critical Survey. Dover, New York, 1957.
- Moro, T.: *Utopía*. Editorial Planeta. Barcelona, 2003.
- Neumann, J. von & Morgenstern, O.: *Theory of Games and Economic Behaviour*. Princeton University Press. Princeton, N. J., 1944.
- Pazos, J.: *Inteligencia Artificial*. Paraninfo. Madrid, 1987.
- Poe, E. A.: *El Misterio de Marie Roget*. En *Cuentos II*. Anaya, S.A. Madrid, 1991. PP: 5-56.
- Poundstone, W.: *Prisoner's Dilemma: John von Neumann, Game Theory and The Puzzle of the Bomb*. Doubleday. New York, 1992. En español, *El Dilema del Prisionero*. Alianza Editorial, S.A. Madrid, 1995.
- Rapoport, A.: *The Use and Misuse of Game Theory*. Scientific American. December, 1962. PP: 108-114.
- San Mateo: *El Santo Evangelio*. En, Varios Autores: *La Santa Biblia. Antiguo y Nuevo Testamento*. Editorial Vida. Miami, Fl., 1960. PP: 938-974.
- Shubik, M.: *Game Theory, Behaviour, and the Paradox of The Prisoner's Dilemma: There Solutions*. Journal of Conflict Resolution, 14. 1970. PP: 181-193.
- Sigmund, K.: *The Calculus of Selfishness*. Princeton University Press. Princeton, N. J., 2010.
- Simmel, G.: *Conflict and the Web of Group Affiliations*. The Free Press. Glencoe, Ill. 1955.
- Stock, G.: *The Book of Questions*. Wohman Publishing. New York, 1987.
- Straffin, P. D. Jr.: *The Prisoner's Dilemma*. Undergraduate Mathematics and its Applications Project Journal. 1. March, 1990. PP: 102-103.