

# La concepción funcionalista de las normas y la utilidad de no apostar a lo seguro

Adriana Spehrs<sup>1</sup>

Recibido: 9 de marzo de 2019

Versión Revisada: 8 de septiembre de 2019

Aceptado: 18 de septiembre de 2019

---

**Resumen.** Ullman-Margalit emplea la Teoría de Juegos a fin de formular una reconstrucción racional de las situaciones de interacción social en que surgen las normas e instituciones. Tversky y Shafir cuestionan este enfoque del origen de las normas argumentando que en contextos de incertidumbre se registran violaciones sistemáticas de uno de los postulados básicos de aquella teoría. En este trabajo procuramos mostrar que tales transgresiones no solo no afectan a la propuesta funcionalista, sino que además permiten responder a la objeción de que el enfoque funcionalista del origen de las normas opone moralidad y racionalidad.

**Palabras clave:** Funcionalismo – Instituciones – Racionalidad – Decisión.

**Title:** The functionalist account of social norms and the utility of not betting on the sure thing

**Abstract.** Ullman-Margalit uses Game Theory to put forward a rational reconstruction of social interaction situations in which norms and institutions arise. Tversky and Shafir reject this approach, arguing that under uncertainty one of the basic tenets of Game Theory is systematically violated. In this paper, we will try to show that these transgressions do not weaken Ullman-Margalit's approach but would make it possible to confront the objection that the functionalist approach opposes morality and rationality.

**Keywords:** Functionalism – Institutions – Rationality – Decision.

---

## 1. Introducción

Ullman-Margalit defiende un enfoque funcionalista con respecto al origen de las normas sociales y las instituciones, empleando las herramientas conceptuales provistas por la Teoría de Juegos. Este tipo de propuestas ha sido cuestionado por Tversky y Shafir, quienes argumentan que en contextos de incertidumbre se registran violaciones sistemáticas de uno de los postulados básicos de aquella teoría. En este trabajo procuramos mostrar que –contrariamente a lo que sostienen Tversky y Shafir– las transgresiones al denominado “Principio de lo seguro” no solo no erosionan la propuesta Ullman-Margalit sino que la fortalecen. En particular, argumentamos que la evidencia de

---

<sup>1</sup> Universidad de Buenos Aires

✉ [adrianaspehrs@yahoo.com.ar](mailto:adrianaspehrs@yahoo.com.ar)

Spehrs, Adriana (2019). La concepción funcionalista de las normas y la utilidad de no apostar a lo seguro. *Epistemología e Historia de la Ciencia*, 4(1), 56-65. ISSN: 2525-1198 (<https://revistas.unc.edu.ar/index.php/afjor/index>)



tales transgresiones puede emplearse para responder a otra objeción que enfrenta el enfoque funcionalista, la objeción de que esta concepción opone moralidad y racionalidad.

## **2. El papel de la Teoría de juegos en la explicación de la emergencia de normas e instituciones**

En *The Emergence of Norms*, Ullman-Margalit formula una reconstrucción racional de las situaciones de interacción social en las que surgen las normas e instituciones. La autora considera que estas situaciones pueden clasificarse en tres clases paradigmáticas, cada una de las cuales se caracteriza por la estructura común de los elementos que la conforman. Cada clase, además, está asociada a un tipo particular de problemas que deben solucionar las personas involucradas en las situaciones correspondientes a la misma. De acuerdo con la autora, tales tipos de problemas pueden resolverse mediante la adopción de ciertas normas de conducta social. Ullman-Margalit recurre a la Teoría de juegos a fin de mostrar que esas normas son, precisamente, la solución que esta teoría brinda para cualquier juego cuya estructura sea equivalente a la de las situaciones de interacción social de la clase correspondiente.

La Teoría de juegos se ocupa de elaborar modelos formales de situaciones en las que dos o más agentes interactúan estratégicamente, de modo tal que los cursos de acción que adoptan están condicionados no solo por sus preferencias sino también por sus expectativas sobre la conducta de los demás jugadores. En esta teoría se asume que los jugadores procuran maximizar su utilidad –o su utilidad esperada, en caso de que los resultados sean aleatorios–, i.e., que buscan el valor máximo de una función que representa el orden de sus preferencias fundadas en su evaluación de las consecuencias de sus decisiones. El objetivo de la Teoría de juegos es determinar el resultado –o los posibles resultados– de cada juego.

Así, por ejemplo, la autora analiza el caso de dos artilleros que deliberan acerca de la conveniencia de huir ante el avance de las tropas enemigas o permanecer en sus puestos de combate y continuar disparando. Como sus armas están ubicadas en un paso estratégicamente muy importante, si ambos permanecen, es posible que resulten heridos pero es seguro que detendrán el avance enemigo. Si ambos huyen, los enemigos tomarán el paso, avanzarán y ambos serán prisioneros de guerra. Si uno solo permanece y el otro huye, el artillero valiente morirá en batalla mientras que el otro tendrá suficientemente tiempo para escapar y ponerse a salvo. Entonces, si ambos desean sobrevivir, preferentemente sin ser heridos, cada artillero tiene buenas razones para huir, independientemente de la decisión que tome el otro. Pero ambos tendrían mejor suerte si los dos permanecieran en sus puestos.

En algunos casos, como en el de este Dilema de los artilleros<sup>2</sup>, es posible resolver un juego eliminando las estrategias dominadas. Una estrategia de un jugador es estrictamente dominada si hay otra estrategia de ese mismo jugador que le provee una

---

<sup>2</sup> El Dilema de los artilleros es un juego no cooperativo –porque las decisiones se toman en ausencia de acuerdos previos– y estático –porque las decisiones se toman simultáneamente y una sola vez, es decir que el jugador decide sin saber qué harán los otros.

mayor utilidad cualquiera sea la estrategia adoptada por el contrincante. La eliminación de estrategias dominadas se considera un requisito mínimo de racionalidad, ya que en esta teoría un jugador es racional si elige la estrategia conducente al resultado que maximiza la utilidad.

La representación del Dilema de los artilleros en su forma normal o estratégica requiere especificar una lista de jugadores, un conjunto de estrategias puras para cada uno de ellos y una función de pagos o de utilidades, que indica la utilidad que reporta a cada jugador cada combinación de estrategias. Esta información se organiza matricialmente empleando una matriz de pagos como la de la tabla 1, en la que la utilidad del jugador cuyas opciones figuran en las filas –AF– se coloca delante de la coma, y la utilidad del jugador cuyas opciones figuran en las columnas –AC– se coloca detrás de la coma.<sup>3</sup>

		AC	
		Permanece / coopera	Huye / no coopera
AF	Permanece / coopera	2,2	0,3
	Huye /no coopera	3,0	1,1

AF: jugador fila

AC: jugador columna

**Tabla 1.** Dilema de los artilleros

En este juego, cualquiera sea la elección del jugador columna AC, huir (fila 2) aporta una utilidad mayor al jugador fila AF. En efecto, si AC elige permanecer (columna 1) a AF le conviene huir porque obtendría 3 en vez de 2. Si AC elige huir (columna 2), a AF le conviene huir porque obtiene 1 en vez de 0. Similarmente, cualquiera sea la elección del jugador fila AF, al jugador columna AC le conviene huir (columna 2) porque obtiene una utilidad mayor. En consecuencia, permanecer en sus puestos de combate es una estrategia dominada para ambos artilleros, de modo que la solución del juego es que los dos huyan, pero esta no sería la solución socialmente óptima. Así, hay un resultado eficiente en el sentido de Pareto –i.e. que ambos permanezcan en sus puestos y continúen disparando– pero no sería esta la alternativa elegida por los artilleros, si cada uno de ellos decide racionalmente aunque guiándose exclusivamente por el interés en la propia supervivencia. En suma, la característica distintiva del Dilema de los artilleros es que, cualquiera sea la elección del rival, a cada jugador le conviene más no cooperar que cooperar, así que la estrategia dominante es no cooperar. Sin embargo, si los dos jugadores deciden no cooperar, ambos se perjudican más que si los dos cooperaran.

<sup>3</sup> Dado que en los juegos estáticos no es importante el orden temporal, porque las decisiones se toman simultáneamente, sin conocer las elecciones de los otros jugadores, suele ser representado en forma estratégica o normal. En tal caso, para describir el juego es necesario especificar una lista de jugadores, un conjunto de estrategias puras para cada uno de ellos y una función de pagos o de utilidades de qué especifica la utilidad que reporta a cada jugador cada combinación de estrategias. Esta información se organiza matricialmente.

Ahora bien, de acuerdo con Ullman-Margalit, si ambos artilleros comprenden el dilema podrían intentar eliminar esta posibilidad, por ejemplo, encadenándose al cañón. Así, si los artilleros encuentran un modo de asegurarse de que ambos respetarán el acuerdo de permanecer en sus puestos de combate, obtendrían el resultado Pareto-eficiente. Según esta autora, esta situación es estructuralmente equivalente a numerosas interacciones cotidianas en las que hacer lo correcto requiere sacrificar el propio interés, como pagar impuestos, cumplir promesas, honrar deudas y contratos, etc. La autora argumenta, entonces, que así como los artilleros deben implementar alguna medida para impedir que alguno de ellos huya, la sociedad se beneficia con instituciones y convenciones sociales que impidan o hagan muy costoso transgredir las normas. De este modo, Ullman-Margalit provee un análisis funcionalista del origen de las normas, según el cual estas permiten que las personas coordinen sus acciones y cooperen en situaciones en que la mera búsqueda del propio beneficio lo impediría. En otras palabras, la función de las normas consistiría en evitar los resultados subóptimos que se obtendrían si nos guiáramos por la racionalidad individual en las interacciones sociales.

### 3. La evidencia empírica y el Principio de lo seguro

Ahora bien, la evidencia empírica registrada en ciertos experimentos condujo a Tversky y Shafir (1992) a cuestionar esta clase de enfoques en los que se emplean los modelos de la Teoría de juegos para dar cuenta de la emergencia de las normas, las convenciones sociales y las instituciones a partir de las decisiones individualmente adoptadas por los jugadores, en ausencia de acuerdos previos. En particular, estos autores fundan sus críticas en los resultados de un experimento en el cual los participantes intervinieron en juegos de una sola jugada del denominado “Dilema del prisionero”, que es estructuralmente análogo al Dilema de los artilleros.<sup>4</sup> En este experimento, cada jugador creía que jugaba con otra persona a través de una computadora, y cada juego se le planteó en 3 oportunidades. En la primera, el jugador no sabía qué decisión había adoptado su rival imaginario. En la segunda, se le informaba que su contrincante había decidido no cooperar. En la tercera, se le informaba que el rival había decidido cooperar. El resultado obtenido fue que la tasa de cooperación era de 3% cuando los participantes creían que el oponente no había cooperado, y de 16% cuando creían que el contrincante sí había cooperado. Pero cuando no sabían si el rival había cooperado o no –que es lo que generalmente se supone en este juego– la tasa de cooperación ascendía al 37%.

De acuerdo con Tversky y Shafir, este resultado constituye una violación del denominado “Principio de lo seguro”, formulado por Savage (1954). El Principio de lo seguro establece que si un curso de acción G es preferido a otro F cuando se sabe ocurrió el evento C, y también que G es preferido a F cuando se sabe que no ocurrió C, entonces G debería ser preferido a F, aunque no se sepa si C ocurrió o no. La evidencia empírica registrada por estos autores pondría de manifiesto que en contextos de incertidumbre, se

---

<sup>4</sup> La forma más simple del Dilema de los prisioneros es la que se juega una única vez, no hay posibilidad de acuerdos implícitos, inducir reciprocidad, desarrollar una reputación, ni de influir de cualquier otra forma sobre la conducta del competidor.

producen violaciones sistemáticas del Principio de lo seguro. En efecto, si bien la mayoría de los participantes elegían no cooperar cuando conocían la alternativa elegida por el rival (sea cooperar o no), una considerable proporción de los jugadores decidía cooperar –que es la estrategia dominada– cuando no conocían la elección del contrincante.

De acuerdo con Tversky y Shafir, el patrón de preferencias obtenido experimentalmente se debe a que los participantes no toman sus decisiones evaluando las posibles consecuencias que resultarían de estas, sino que buscan una razón que justifique elegir una opción en lugar de la otra. En el caso del juego analizado, los jugadores no cooperan cuando saben que su oponente no cooperó porque consideran que así lo retribuyen justamente. Y tampoco cooperan cuando saben que el contrincante sí cooperó porque así maximizan su beneficio. En cambio, cuando no saben si el oponente cooperó o no, carecen de razones que justifiquen no cooperar, por eso una mayor cantidad de jugadores eligen la estrategia dominada.

Posteriormente, varios autores obtuvieron similares resultados experimentales empleando juegos con la misma estructura que la del Dilema del prisionero. Más aún, Croson (1999) también registró resultados inconsistentes con el Principio de lo seguro empleando una versión de ese juego en la cual los pagos que obtiene cada jugador son diferentes, aunque ambos tengan la misma estrategia dominante –i.e. no cooperar– como en el caso que representamos en la tabla 2:

		JC	
		coopera	no coopera
JF	Coopera	8, 6	3, 7
	no coopera	9, 1	4, 2

JF: jugador fila

JC: jugador columna

**Tabla 2.** Versión asimétrica del Dilema del prisionero

Este tipo de asimetría, con todo, no parece relevante ya que carece de importancia que los montos de los pagos que reciben los jugadores sean diferentes, dado que en Teoría de juegos no se efectúan comparaciones interpersonales de utilidad. Es decir, carece de sentido comparar, por ejemplo, la máxima utilidad del jugador fila –9– con la máxima utilidad del jugador columna –7–. En contraste, sí sería relevante que fueran distintas las diferencias entre las utilidades máximas y mínimas de cada jugador, lo cual no sucede en este caso, pues tal diferencia es de una unidad para ambos jugadores. De todos modos, cabe señalar que Croson reporta haber obtenido similares resultados en variantes del Dilema del prisionero que eran asimétricas en el sentido de que solo uno de los jugadores tenía una estrategia dominada.

Ahora bien, Tversky y Shafir sostienen que este resultado es inconsistente con las explicaciones de la cooperación en términos de normas sociales e imperativos morales

–i.e. las explicaciones según las cuales cada agente está dispuesto a cooperar si el otro coopera. Los autores afirman también que estos resultados permiten cuestionar la Teoría de utilidad esperada que subyace a la Teoría de juegos pues esta es una concepción consecuencialista, dado que asume que las decisiones de los agentes racionales deben estar determinadas por la evaluación de las potenciales consecuencias y de sus respectivas probabilidades. Y, según Tversky y Shafir las decisiones fundadas en una evaluación consecuencialista de los resultados previstos deberían satisfacer el Principio de lo seguro. Los autores consideran que el Principio de lo seguro es un supuesto básico implícito de la Teoría de utilidad esperada porque expresa la intuición fundamental de qué significa que una decisión esté determinada por la previsión de las consecuencias. Por eso, en su opinión, cualquier transgresión del Principio de lo seguro refutaría la Teoría de utilidad esperada. Más aún, si las decisiones se fundaran en una evaluación consecuencialista de las alternativas, el Principio de lo seguro debería cumplirse no solo en los casos en donde la incertidumbre se origina en el desconocimiento de la estrategia que elegirá un rival –como ocurre en la Teoría de juegos– sino también en los casos donde la incertidumbre surge del desconocimiento de cuál será el estado de la naturaleza que tendrá lugar –como sucede en Teoría de la decisión.

Sin embargo, Tversky y Shafir (1992) también registraron violaciones del Principio de lo seguro en experimentos donde la incertidumbre es endógena. En particular, estos autores plantearon a un grupo de 200 estudiantes un problema de decisión que consistía en que debían optar entre aprovechar o no una oferta por tiempo limitado para pasar las vacaciones en Hawaii, luego de haber rendido un examen muy exigente. El resultado obtenido fue que la mayoría de los alumnos que eran informados de que habían aprobado el examen y de los que eran informados de que lo habían desaprobado, decidía aceptar la oferta. En cambio, quienes no sabían si habían aprobado o no, preferían posponer la decisión hasta conocer el resultado del examen, lo que ocurriría cuando ya no estaría vigente la oferta, a menos que pagaran por adelantado una suma adicional de dinero. De acuerdo con los autores, este resultado se debe a que quienes aprobaron tenían una razón para viajar –i.e. festejar la buena noticia– y también la tenían quienes no aprobaron –i.e. buscar consuelo. Pero quienes no sabían si había aprobado o no, carecían de razones que les permitieran preferir una opción a otra. En otras palabras, nuevamente aquí se observó que los jugadores prefirieron G a F cuando sabían que había ocurrido C y también cuando sabían que no había ocurrido C, pero preferían F a G cuando no sabían si C había ocurrido o no. Así, los autores concluyen que cuando la gente delibera en contextos de incertidumbre, no considera apropiadamente todas las alternativas tal como requiere la perspectiva consecuencialista, lo que los lleva a transgredir el Principio de lo seguro.

Tversky y Shafir sostienen que es habitual que estas violaciones del Principio de lo seguro se manifiesten también en situaciones no generadas en un contexto experimental, tal como la registrada en EEUU en 1988. En esa ocasión los mercados financieros permanecieron inactivos y estables en las semanas previas a los comicios presidenciales, pero la semana siguiente a la victoria de G. Bush, el dólar descendió hasta su valor más bajo en diez meses, junto con el precio de los bonos y el índice Dow Jones. Los analistas y operadores bursátiles, sin embargo, consideraban que el mismo declive habría ocurrido si hubiera triunfado Dukakis, puesto que reflejaba la preocupación por

los elevados niveles de déficits comercial y fiscal. En cambio, la inactividad precedente se atribuyó a la cautela de los inversores ante la incertidumbre generada por el resultado de las elecciones presidenciales.<sup>5</sup>

En suma, Tversky y Shafir consideran que estas transgresiones del Principio de lo seguro constituyen evidencia refutadora de la Teoría de utilidad esperada. Sin embargo, contrariamente a lo que ellos suponen, el Principio de lo seguro no se sigue de la lógica clásica, ni del cálculo de probabilidades ni de los axiomas de la Teoría de utilidad esperada. Así que una transgresión del Principio de lo seguro no implica, por sí misma, una refutación de la Teoría de utilidad esperada, en la cual se sustentan tanto la Teoría de la decisión como la Teoría de juegos.

En efecto, el Principio de lo seguro no puede derivarse de los axiomas de la Teoría de utilidad esperada pues involucra relaciones epistémicas, tales como “sabe que C” o “no sabe si C ocurrió o no”. Sin embargo, las relaciones epistémicas como estas no pueden expresarse mediante el cálculo de probabilidades, que es la herramienta básica de la Teoría de la decisión y la Teoría de juegos.

Ahora bien, si se reformulara el Principio de lo seguro de modo tal que pudiera expresarse en los términos del cálculo de probabilidades, las expresiones que refieren a estas relaciones serían transformadas en expresiones que establecen una probabilidad condicionada para el caso de que suceda o no el evento C. Pero no podemos identificar la ocurrencia del evento C con el hecho de que un agente sepa que ocurrió C, y tampoco podemos asimilar la no ocurrencia del evento C con la circunstancia de que el agente no sepa si C ocurrió o no.

Por otra parte, para poder integrar el Principio de lo seguro dentro de cálculo de probabilidades, en lugar de hacer referencia a preferencias por cursos de acción, el Principio de lo seguro debería referirse a preferencias por la ocurrencia de eventos. Pero aun si se lo reformulara de este modo, el Principio de lo seguro no se seguiría del cálculo de probabilidades ni de la Teoría de utilidad esperada. En efecto, como analizaremos a continuación, C. Blyth (1972), demostró que maximizar la utilidad esperada en ciertas circunstancias requeriría transgredir el Principio de lo seguro.

#### **4. La Paradoja de Simpson y el Principio de lo seguro**

Si bien parece imposible la ocurrencia de las reversiones de preferencias prohibidas por el Principio de lo seguro, Blyth probó que en aquellas situaciones en donde se manifiesta la denominada “Paradoja de Simpson”, cumplir con el Principio de lo seguro podría impedir que maximicemos la ganancia esperada. La Paradoja de Simpson, en su versión fuerte, se plantea cuando la asociación estadística entre dos variables A y B revierte su signo si se la condiciona con respecto a una tercera variable C, cualquiera sea el valor de esta.<sup>6</sup> Es decir, si desagregamos los datos en sub poblaciones, cada una de las cuales representa un valor de esta variable C, puede cambiar el signo de la asociación

---

<sup>5</sup> Similarmente, los críticos de la estrategia norteamericana del ataque nuclear preventivo han sostenido que esta es una estrategia inviable en cualquiera de las posibles disposiciones de los misiles rusos, pero insisten en retener esa opción en tanto no se sepa la disposición exacta de los arsenales nucleares rusos.

<sup>6</sup> Versión publicada ya por M. Cohen y E. Nagel (1934).

estadística entre A y B medida en las sub poblaciones con respecto al de la asociación entre A y B medida con los datos agregados.

Consideremos, por ejemplo, el caso de un médico que decide probar un nuevo tratamiento con sus pacientes, la mayoría de los cuales residen en su misma ciudad, aunque unos pocos lo hacen fuera de ella. A fin de comparar la efectividad del este nuevo tratamiento con el tradicionalmente empleado, el médico aplica el nuevo tratamiento al 0.91% de los pacientes de su ciudad y el tratamiento tradicional a los restantes de esa ciudad. Además, aplica el nuevo tratamiento al 0.01% de sus pacientes que no viven en su ciudad y el tratamiento tradicional a los 0.99% restantes. Posteriormente, el médico registra que, de la totalidad de los pacientes tratados, sobrevivieron el 46% de los que recibieron el tratamiento tradicional, mientras que solo lo hicieron el 11% de quienes recibieron el nuevo tratamiento. Si emplamos las letras “A” para representar el evento de que un paciente sobrevive, “B” para el evento de que el paciente recibe el nuevo tratamiento, “¬B” para el evento de que recibe el tratamiento tradicional, “C” para el evento de que el paciente viva en la misma ciudad que el médico y “¬C” para el evento de que viva fuera de ella, este resultado puede expresarse simbólicamente así:  $p(A/B) < p(A/\neg B)$ . Por otra parte, entre los pacientes que viven en la misma ciudad que el médico, la proporción de sobrevivientes que recibieron el nuevo tratamiento duplicó a la de los sobrevivientes que recibieron el tratamiento estándar, es decir:  $p(A/B \wedge C) > p(A/\neg B \wedge C)$ . Y también la proporción de sobrevivientes que recibieron el nuevo tratamiento duplicó a la de los sobrevivientes que recibieron el tratamiento estándar, entre los pacientes que no viven en la misma ciudad que el médico:  $p(A/B \wedge \neg C) > p(A/\neg B \wedge \neg C)$ . Dadas las siguientes cantidades de pacientes que viven en la ciudad en cuestión y la de los que no viven allí, los resultados fueron los que figuran en la tabla 3:

	pacientes C		pacientes ¬C		Total de pacientes	
	Tratados con B	Tratados con ¬B	Tratados con B	Tratados con ¬B	Tratados con B	Tratados con ¬B
Muertos ¬A	9000	950	5	5000	9005	5950
Vivos A	1000 (10%)	50 (5%)	95 (95%)	5000 (50%)	1095 (11%)	5050 (46%)
Total	10000	1000	100	10000	10100	11000

Estos resultados pueden sistematizarse mediante las siguientes desigualdades:

$$p(A/B) = 11\% < 46\% = p(A/\neg B).$$

$$p(A/B \wedge C) = 10\% > 5\% = p(A/\neg B \wedge C)$$

$$p(A/B \wedge \neg C) = 95\% > 50\% = p(A/\neg B \wedge \neg C)$$

**Tabla 3.** Paradoja de Simpson



Aunque este resultado parece paradójico, en realidad puede explicarse si se considera que los pacientes que viven en la misma ciudad que el médico tienen una probabilidad de sobrevivir muy inferior a la de los que no viven allí. Además, el nuevo tratamiento fue aplicado mayoritariamente a los pacientes que viven en la misma ciudad que el médico. Y es claro que un tratamiento tendrá como resultado una tasa de supervivencia inferior si se administra a los pacientes con menores chances de recuperarse. Es decir, la dependencia o interacción entre las variables B –recibir el nuevo tratamiento– y C –vivir en la misma ciudad que el médico– puede generar una inversión en la correlación estadística. Sin embargo, si estas variables fueran independientes, no surgiría la paradoja.

Ahora bien, empleando este caso de la Paradoja de Simpson, Blyth formuló un juego secuencial de apuestas en el cual el curso de acción que viola el Principio de lo seguro arroja un pago mayor que el que se deriva del cumplimiento del Principio de lo seguro. El autor denomina “G” al curso de acción consistente en elegir al azar la ficha médica de un paciente que recibió el nuevo tratamiento y apostar un dólar a que sobrevivió. Y denomina “F” al curso de acción consistente en escoger al azar la ficha de un paciente que recibió el tratamiento estándar y apostar un dólar a que sobrevivió. En estas condiciones, preferiremos F a G porque F aporta mayor probabilidad de ganar que G. Sin embargo, si la ficha se extrajera del conjunto de fichas que corresponden a los pacientes que viven en la misma ciudad que el médico, preferiríamos G en lugar de F, porque en tal caso la probabilidad de ganar con G es el doble que con F. Similarmente, si la ficha se extrajera del conjunto de fichas correspondientes a los pacientes que no vive en la misma ciudad que el médico, preferiríamos G antes que F, porque la probabilidad de ganar con G casi el doble que con F. Pero si desconocemos si la ficha procede del conjunto de fichas de los pacientes que viven en la misma ciudad o del conjunto de fichas de los otros pacientes, adoptaremos un curso de acción que viola el Principio de lo seguro. De modo que el Principio de lo seguro no es aplicable cuando la elección entre los cursos de acción F y G dependen del evento C.

## 5. Consideraciones finales

En suma, las transgresiones del Principio de lo seguro registradas empíricamente no prueban que en tales experimentos se hayan manifestado preferencias irracionales desde la perspectiva de la Teoría de la utilidad esperada, dado que el Principio de lo seguro no se infiere de la Teoría de la utilidad esperada. En otras palabras, los resultados obtenidos por Tversky y Shafir no permiten asegurar que los decisores no maximicen su utilidad, de modo que con esa evidencia no puede considerarse refutada la Teoría de la utilidad esperada.

En este punto, es importante destacar que estas transgresiones del Principio de lo seguro permitirían responder la objeción que habitualmente se formula contra la concepción funcionalista de las normas. Nos referimos a la objeción de que esta concepción opone moralidad y racionalidad. En efecto, como señalamos, desde la perspectiva de la racionalidad individual de cada artillero, lo más conveniente es huir; solo puede detenerlo en su puesto de combate su voluntad de respetar las normas, su

apego a las prescripciones morales. Pero la evidencia empírica registrada por Tversky y Shafir indica que más de la tercera parte de los participantes en el experimento en el que se empleaba el Dilema del prisionero cooperaban cuando desconocían la decisión del rival, y no cooperaban tanto cuando creían que el adversario había cooperado como cuando creían que no lo había hecho. Este patrón de conducta, en el caso del Dilema de los artilleros, correspondería a permanecer en el puesto de combate cuando se desconoce qué decidió el compañero, y huir tanto cuando se sabe que el compañero huirá como cuando se sabe que permanecerá combatiendo.

Ahora bien, aunque esta elección constituya una violación del Principio de lo seguro, no puede calificarse sin más como irracional sino que, contrariamente a lo que suponen Tversky y Shafir, podría ser el resultado de una tentativa por alcanzar el óptimo social. Más aún, como en el caso de los experimentos efectuados por estos autores la elección entre las opciones “cooperar” y “no cooperar” es dependiente de la variable “saber qué curso de acción eligió el rival”, aplicar el Principio de lo seguro no necesariamente conducirá a maximizar la utilidad.

En cualquier caso, dado que el experimento realizado por Tversky y Shafir incluye dos casos en los cuales los jugadores disponen de la información relativa a la decisión adoptada por su adversario imaginario, no puede considerarse como un test de la hipótesis funcionalista de la emergencia de las normas e instituciones. Pues, al menos en la versión de Ullman-Margalit, esta hipótesis se funda en el Principio de lo seguro en su versión estática, i.e. las decisiones se toman simultáneamente y una sola vez, de modo tal que el jugador decide sin saber qué hará su contrincante.

## 6. Bibliografía

- Bagassi, M., & Macchi, L. (2006). Pragmatic approach to decision making under uncertainty: The case of the disjunction effect. *Thinking and Reasoning*, 12, pp. 329-350.
- Blyth, C.R. (1972). On Simpson's Paradox and the Sure Thing Principle. *Journal of the American Statistical Association*, vol 67, n° 338, June 1972, pp. 364-366.
- Croson, R. (1999). The Disjunction Effect and Reason-Based Choice in Games, *Organizational Behavior and Human Decision Processes*, vol. 80, n° 2, November 1999, pp. 118-133.
- Shafir, E. & Tversky, A. (1992). Thinking through uncertainty. *Cognitive Psychology*, 24, pp. 449-474.
- Li, S., Wang, Z., Rao, L. & Li, Y. (2010). “Is There a Violation of Savage's Sure-Thing Principle in the Prisoner's Dilemma Game?”, *Adaptive Behavior*, (2010), Vol 00(0), pp. 1-9.
- Tversky, A., & Shafir, E. (1992). “The disjunction effect in choice under uncertainty”, *Psychological Science*, 3, pp.305-309.
- Von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.
- Ullman-Margalit, E. (1977). *The Emergence of Norms*, Oxford: Oxford University Press.