

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS POLITICAS Y SOCIOLOGIA
Departamento de Sociología I



**COORDINACIÓN Y CONVENCIÓN:
UN ESTUDIO TEÓRICO Y EXPERIMENTAL DE LA
NOCIÓN DE CONVENCIÓN SOCIAL**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Luís Miguel Miller Moya

Bajo la dirección de los doctores:
Fernando Aguiar González y Andrés de Francisco Díaz

Madrid, 2007

ISBN: 978-84-669-3143-4

COORDINACIÓN Y CONVENCION

**Un estudio teórico y experimental de la noción de convención
social**

TESIS DOCTORAL

LUIS MIGUEL MILLER MOYA

Madrid, septiembre de 2006

**Departamento de Sociología 1 (Cambio social)
Facultad de Ciencias Políticas y Sociología**

Programa de doctorado: Dinámicas sociales en la sociedad contemporánea

Coordinación y convención

Un estudio teórico y experimental de la noción de convención social

TESIS DOCTORAL

Luis Miguel Miller Moya

Directores: Fernando Aguiar González y Andrés de Francisco Díaz

Universidad Complutense de Madrid

Septiembre de 2006

ÍNDICE

Índice de Tablas y Figuras	vi
Agradecimientos	vii
CAPÍTULO 1. INTRODUCCIÓN	1
1.1. Objeto de estudio de la tesis	1
1.2. Metodología de la tesis	6
1.3. Estructura general de la tesis	9
PARTE I. LA NOCIÓN DE CONVENCIÓN SOCIAL	13
CAPÍTULO 2. LA TEORÍA ECONÓMICA Y LAS <i>CONVENCIONES DE COMPORTAMIENTO</i>	15
2.1. Introducción	15
2.2. La propuesta original de Hume	16
2.3. David Lewis y las convenciones como regularidades de hecho en el comportamiento humano	22
2.4. Las convenciones sociales como regularidades arbitrarias e indeterminadas	31
2.5. Las convenciones como estrategias evolutivamente estables	37
2.6. El carácter normativo de las <i>convenciones de comportamiento</i>	46
2.7. Resumen	52
CAPÍTULO 3. LA TEORÍA SOCIOLÓGICA Y LAS <i>CONVENCIONES NORMATIVAS</i>	55
3.1. Introducción	55
3.2. La propuesta original de Weber	56
3.3. La recepción de la noción weberiana de convención	62
3.4. La convención como principio	64
3.5. Convenciones y acuerdos: el componente normativo de las <i>convenciones normativas</i>	68
3.6. Resumen	77
CAPÍTULO 4. HACIA UNA DEFINICIÓN INTEGRADA DE LA NOCIÓN DE CONVENCIÓN SOCIAL	79
4.1. Introducción	79
4.2. Definición de convención social	81
4.3. Conclusiones	86

PARTE II. APROXIMACIÓN TEÓRICA A LA NOCIÓN DE CONVENCIÓN SOCIAL	88
CAPÍTULO 5. LA ESTABILIDAD DE LAS CONVENCIONES SOCIALES: UNA PROPUESTA TEÓRICA EVOLUTIVA	90
5.1. Introducción	90
5.2. Clase de juegos: “ <i>The Game of Life</i> ”	97
5.3. Principios normativos	103
5.4. Análisis evolutivo	109
5.5. Interpretación de los resultados formales	120
5.6. Interpretación informal de los resultados	123
5.7. Conclusiones	127
PARTE III. APROXIMACIÓN EXPERIMENTAL A LA NOCIÓN DE CONVENCIÓN SOCIAL	129
CAPÍTULO 6. LA VALIDEZ DE LOS EXPERIMENTOS ECONÓMICOS EN SOCIOLOGÍA	131
6.1. Introducción	131
6.2. Dos concepciones distintas acerca de la experimentación	135
6.2.1. Experimentos de orientación empírica	136
6.2.2. Experimentos de orientación teórica	141
6.3. La validez externa de los experimentos de orientación teórica	145
6.4. La especificidad de los experimentos económicos	149
6.4.1. Uso de Incentivos monetarios	150
6.4.2. Prohibición del engaño (<i>deception</i>)	153
6.5. Conclusiones	154
CAPÍTULO 7. LA DOBLE NATURALEZA DE LAS CONVENCIONES SOCIALES: UN ANÁLISIS EXPERIMENTAL	157
7.1. Introducción	157
7.2. Primer experimento	160
7.2.1. Diseño y conducción del primer experimento	160
7.2.2. Resultados del primer experimento	164
7.2.3. Conclusiones del primer experimento	173
7.3. Segundo experimento	175
7.3.1. Diseño y conducción del segundo experimento	175
7.3.2. Resultados del segundo experimento	180
7.3.3. Conclusiones del segundo experimento	188
7.4. Conclusiones	190
Apéndice 1. Instrucciones del primer experimento	194
Apéndice 2. Instrucciones del segundo experimento	200
Apéndice 3. Juegos realmente jugados en el segundo experimento	207
CAPÍTULO 8. CONCLUSIONES GENERALES	209
REFERENCIAS BIBLIOGRÁFICAS	222

ÍNDICE DE TABLAS Y FIGURAS

TABLAS

Tabla 1.	Tipología de convenciones	84
Tabla 2.	Tabla de principios	109
Tabla 3.	Tabla de éxitos	115
Tabla 4.	Convenciones normativas evolutivamente estables	120
Tabla 5.	Ejemplos de experimentos controlados sobre programas sociales llevadas a cabo en los Estados Unidos entre 1968-1991	139
Tabla 6.	Características de los tratamientos del primer experimento	162
Tabla 7.	Nivel de coordinación alcanzado según condiciones informativas	173
Tabla 8.	Características de los tratamientos del segundo experimento	179

FIGURAS

Figura 1.	Situación pura de coordinación	25
Figura 2.	Juego de la conducción	33
Figura 3.	Situación de negociación	39
Figura 4.	Clase de juegos básica	98
Figura 5.	Juego de coordinación asimétrica	100
Figura 6.	<i>Status quo</i> de desigualdad	102
Figura 7.	Diseño apropiado del experimento	142
Figura 8.	Pantalla del ranking en el primer experimento	163
Figura 9.	Pantalla de toma de decisiones en el primer experimento	164
Figura 10.	Frecuencia de primeros principios	166
Figura 11.	Frecuencia de todos los principios utilizados	167
Figura 12.	Consistencia con la acción prescrita por los primeros principios	168
Figura 13.	Nivel de coordinación alcanzado en cada tratamiento	171
Figura 14.	Nivel de coordinación alcanzado en distintos grupos de períodos	172
Figura 15.	Pantalla de toma de decisiones del tratamiento B en el segundo experimento	177
Figura 16.	Pantalla del ranking en el segundo experimento	178
Figura 17.	Pantalla de toma de decisiones del tratamiento D en el segundo experimento	179
Figura 18.	Principios seguidos realmente en el tratamiento B	181
Figura 19.	Principios ordenados en primer lugar y principios realmente seguidos en el tratamiento P	183
Figura 20.	Consistencia con el comportamiento prescrito en el tratamiento P	184
Figura 21.	Principios ordenados en primer lugar y principios realmente seguidos en el tratamiento D	185
Figura 22.	Coordinación según tratamiento	186

AGRADECIMIENTOS

Una de las primeras disyuntivas que se me plantearon al emprender la redacción final de esta tesis fue la de abrazar una convención ya en desuso que, con objeto de mostrar el carácter humilde del autor, recomendaba utilizar el “plural de cortesía” como sujeto de la misma, o decantarme por emplear la primera persona del singular como muestra del carácter individual de la empresa. Aunque en principio me decanté por esta segunda opción, pronto me di cuenta de que el desasosiego que me provocaba su uso poco tenía que ver con la adherencia a una u otra convención, sino con el hecho de que tantas son las contribuciones académicas y personales a la misma que difícilmente era sostenible el uso del singular. Con estas líneas sólo pretendo mostrar mi agradecimiento a las personas e instituciones que han hecho posible que este trabajo intelectual haya llegado a buen puerto.

Mi primer agradecimiento no puede ser sino para mis directores, los cuales, si bien han desempeñado un papel muy distinto, han sido determinantes para la realización de esta tesis. Andrés de Francisco ha sido, más allá de la relación personal que nos une, ese profesor que consigue hacerte disfrutar con la teoría. Hasta tal punto esto es así que no creo que hubiera sido posible realizar una tesis en la que la teoría ha desempeñado un papel tan importante si Andrés no se hubiera cruzado en mi camino. De él también he aprendido el gusto por el rigor filosófico y metodológico, así como una preocupación casi obsesiva por la claridad y la inteligibilidad de todo cuanto escribo. Fernando Aguiar ha sido un auténtico maestro. Alguien al que he podido acudir diariamente durante los últimos cuatro años para discutir todo tipo de cuestiones personales e intelectuales. Su confianza incondicional, su apoyo continuo y su disponibilidad absoluta, en el terreno personal, así como su rigor científico, su brillantez intelectual y su idea acerca de la necesaria imbricación entre el contenido y la forma de lo escrito, en el plano académico, han hecho que estos cuatro años se hayan convertido en un período de duro mas placentero trabajo. Vaya para ellos dos mi máxima admiración y el mayor de los agradecimientos.

Aun disponiendo de los mejores referentes intelectuales, una tesis difícilmente logra culminarse sin el apoyo institucional adecuado. En este sentido, el Instituto de Estudios Sociales Avanzados (IESA-CSIC) me ha ofrecido todos los medios y el calor humano que uno necesita en esta labor. Así, la tesis se ha beneficiado de los comentarios, discusiones y críticas recibidos tanto en el propio IESA, como en la multitud de congresos y reuniones científicas a las que he tenido la oportunidad de asistir gracias al apoyo económico del IESA. Por ello, vaya también mi agradecimiento para esta institución y, en especial, para sus directores Manuel Pérez Yruela y Eduardo Moyano.

Además del IESA he tenido también la suerte de contar con el apoyo de otras instituciones que me han facilitado la adquisición de los conocimientos teóricos y prácticos que me han permitido completar esta tesis. La Unión Europea, a través de su programa *Training and Mobility of Researchers (TMR) Programme*, me permitió pasar dos meses en la Universidad de Essex. Por su parte, el Max Planck Institute of Economics (Jena, Alemania), en el que buena parte de esta tesis fue concebida y desarrollada, financió tanto mis tres estancias, como las investigaciones experimentales llevadas a cabo en el mismo. A su director, Werner Güth, además de agradecerle las múltiples contribuciones intelectuales realizadas a este trabajo, le agradezco el apoyo

material sin el que esta tesis no hubiera sido posible en su forma actual. El Departamento de Economía de la Universidad de Trento financió la estancia y la asistencia a una escuela de verano sobre teoría de juegos evolutiva y ciencias sociales, cuyas enseñanzas han sido fundamentales para algunas partes de este trabajo. Por último, las visitas periódicas al Departamento de Teoría e Historia Económica de la Universidad de Granada me han permitido mejorar las partes más técnicas de la tesis.

Las contribuciones intelectuales a esta tesis han sido múltiples. A mi llegada al IESA, Francisco Herreros y Henar Criado fueron unos excelentes compañeros en mi aterrizaje en la teoría de juegos y los modelos matemáticos. Dicha labor fue continuada con una de las personas de las que más he aprendido en los últimos años: Vistoria Ateca. En los momentos iniciales de esta investigación, Vicky, además de ser una buena amiga, ejerció la labor de una auténtica codirectora de tesis. En aquellos momentos iniciales también fue crucial la presencia de Gloria Martínez, con la que he mantenido valiosas conversaciones en los últimos cuatro años y de la que he intentado aprender a “escribir bien”. Además de los ya mencionados, las conversaciones mantenidas con algunos compañeros del IESA se han visto reflejadas en distintos apartados de esta tesis, tal es el caso de José Manuel Robles, Ernesto Ganuza, Rafael Serrano, Braulio Gómez y Ana León. Pero, más allá de estos agradecimientos individuales, la relación mantenida con la mayoría de mis compañeros del IESA no sólo ha hecho más llevaderos estos últimos cuatro años de mi vida, sino que, en cierto modo, el propio tema de la tesis no podría ser entendido sin la convivencia con ellos durante este período.

Esta tesis tampoco hubiera sido posible sin la aportación de un grupo de “economistas raros” que me llevaron desde muy pronto a la fascinación que en un sociólogo joven provocaban materias como la economía del comportamiento, la economía experimental, la imbricación entre economía y psicología o, incluso, la neuroeconomía. En este sentido, de nuevo Victoria Ateca, Antonio Morales, Enrique Fatás y, sobre todo, Pablo Brañas se empeñaron en mostrarme con buenas razones la pertinencia de estos enfoques novedosos para abordar teórica y empíricamente las ideas inconexas que poblaban mi cabeza. Pablo también me dio el primer empujón para irme a Alemania, algo que cambió profundamente esta tesis, así como posiblemente mi propia carrera investigadora.

De mis varias visitas a Alemania sólo puedo decir que pocas veces uno tiene la oportunidad de trabajar “codo con codo” con uno de los máximos representantes de una disciplina científica, en este caso la economía experimental. El profesor Werner Güth transformó, en apenas veinticuatro horas, algunos esquemas apenas desarrollados que llevaba en mi mochila en un marco teórico coherente y susceptible de ser abordado empíricamente. De su aportación sólo puedo decir que tardé un año en asimilar unas ideas que supuestamente estaban en mi mente, pero que nunca hubieran tomado forma sin su ayuda. También han contribuido a esta tesis, con ideas, críticas e incluso ayuda logística en la realización de los experimentos, otros miembros del *Strategic Interaction Group* del *Max Planck Institute of Economics*, como Vittoria Levati, Torsten Weiland, Christoph Vanberg, Luis González y Susanne Büchner. También tuve la oportunidad de recibir unas valiosísimas críticas a algunas partes de este trabajo por parte del profesor Hartmut Kliemt.

Estos agradecimientos no estarían completos sin referirme a aquéllos que han estado más cerca personalmente durante estos últimos años. Enrique Moreno y, sobre todo, Javier Hidalgo han hecho que la vuelta a Córdoba, lejos de ser una experiencia traumática como yo imaginaba, me haya permitido vivir los cuatro años más intensos de mi vida. A Javi le debo también la insistencia en buscar ejemplos simples y cotidianos para ilustrar los argumentos abstractos de la tesis. Los cientos de conversaciones que hemos mantenido en los últimos años han enriquecido sobremanera el resultado final de la misma. Ana León, por su parte, ha conseguido que en ese momento final de la redacción de la tesis, en el que uno sólo está pensando en terminar el trabajo, todavía sacara fuerzas para imprimirle algo más de cariño y pasión a aquello que escribía. Esta tesis seguro que hubiera sido menos “fría” si hubiera caído antes en sus manos.

He dejado para el final a quienes, de un modo constante durante todo este período, han soportado la vida poco convencional de alguien que se empeñaba en hacer una tesis teórica sobre las convenciones sociales. A mis padres, por su comprensión ilimitada y la educación que me han dado, así como a mi hermana, porque en el fondo siempre está ahí, les dedico esta tesis.

CAPÍTULO 1

INTRODUCCIÓN

1.1. Objeto de estudio de la tesis

No podríamos hablar de sociedad sin convenciones sociales. Sin ellas los humanos nos sentiríamos perdidos en toda interacción social, sin saber cómo actuar en cada encuentro con los demás, desbordados por la necesidad de reinventar cada pauta de comportamiento en la sociedad. Así, resulta imposible imaginar un contrato de compraventa sin medio de intercambio alguno, como el papel moneda, o una celebración – digamos una boda – en la que los participantes no conocieran la indumentaria que han de llevar. Difícil sería emprender cualquier búsqueda en un diccionario o enciclopedia, si no existiese un modo de ordenar los volúmenes o si las palabras estuvieran dispersas en lugar de seguir un abecedario. Tan obvia nos resulta la existencia de convenciones sociales, que rara vez caemos en la cuenta de nuestra experiencia cotidiana con ellas¹. Asumimos que existen sin preguntarnos acerca de las condiciones de su existencia; sin cuestionarnos, además, si existen otras alternativas a las mismas. Pero lo cierto es que sería difícil escapar de una convención sin caer en otra o sin caer en otro marco regulativo que nos ayude a desenvolvernó en nuestra sociedad.

¹ En realidad, normalmente sólo nos damos cuenta de que nos hallamos en presencia de una convención social cuando nuestras creencias o nuestro comportamiento chocan contra ésta. La literatura está plagada de ejemplos en este sentido. Por citar sólo uno bastante revelador, hemos elegido el siguiente fragmento de *Lolita*, la universal obra de Nabokov, en la que el protagonista Humbert Humbert se refiere del siguiente modo a su esposa: “Charlotte, que no se daba cuenta de la intrínseca falsedad de las convenciones y normas de conducta a las que se atenía” (Nabokov, 2002: 105).

A pesar de este contacto diario con las convenciones sociales, no resulta tan sencillo el hecho de explicarlas desde un punto de vista teórico. Como veremos en esta tesis, explicar la emergencia, la evolución y el ocaso de las convenciones sociales ha sido un objetivo perseguido por un buen número de filósofos y científicos sociales² sin que, por el momento, dispongamos de una teoría que haya cerrado el debate acerca de qué son y por qué existen. Este trabajo pretende realizar una pequeña aportación a dicha discusión. Pero antes de entrar de lleno en la misma, presentaremos en esta introducción los fundamentos teóricos y metodológicos de la tesis que vamos a desarrollar, así como la estructura de la misma. Intentaremos mostrar el camino seguido para desvelar los interrogantes que nunca nos planteamos en nuestra experiencia cotidiana con las convenciones sociales.

Como punto de partida diremos que una convención es un tipo específico de institución humana³. En otras palabras, dentro del conjunto más amplio de las instituciones humanas se encuentra el subconjunto de las convenciones sociales. Si bien esta primera observación podría ser aceptada por la mayoría de teóricos sociales, la cuestión de qué características definen el subconjunto de las convenciones sociales no parece tan clara. En el resto de la tesis tendremos oportunidad de contrastar la mayor parte de las teorías contemporáneas de las convenciones sociales. Sin embargo, antes de esto, y ya que hemos definido las convenciones como un tipo específico de institución, parece lógico preguntarse qué es una institución.

² Desde clásicos del pensamiento filosófico como David Hume, hasta filósofos analíticos de la segunda mitad del siglo XX, como David Lewis, Edna Ullmann-Margalit o Margaret Gilbert; desde clásicos de la teoría sociológica como Ferdinand Tönnies o Max Weber, hasta sociólogos contemporáneos como los pertenecientes a la escuela francesa de la Economía de las Convenciones; desde clásicos del pensamiento económico como John Maynard Keynes o Thorstein Veblen, hasta teóricos de juegos contemporáneos como Ken Binmore o Robert Sugden.

³ En la segunda mitad del siglo pasado asistimos a un renacimiento de propuestas de marcos teóricos integrados que pretenden contribuir a resolver la cuestión de la emergencia, la evolución y el mantenimiento de las instituciones sociales. Ver, entre otros, Schotter (1981), Williamson (1985), North (1990), Hechter, Opp y Wippler (1990) o Knight (1995).

En un artículo seminal, Crawford y Ostrom (1995: 582) afirman que “las instituciones son regularidades duraderas en la acción humana en situaciones estructuradas por reglas, normas y estrategias compartidas, así como por el mundo físico.” Como veremos, esta afirmación no puede estar más en consonancia con el planteamiento de nuestra propuesta teórica. Así, en toda institución pueden ser distinguidos, al menos desde un punto de vista analítico, dos componentes: un componente de regularidad duradera y estable en el comportamiento humano y un componente estructurante⁴, constituyente o normativo que, a modo de prescripción, “ordena” dicha regularidad de comportamiento.

Esta consideración de la “doble naturaleza” de las instituciones, tan clara en el planteamiento de Ostrom (Crawford y Ostrom, 1995; Ostrom, 2005), y central para nuestra definición de las convenciones sociales, no puede decirse que haya calado hondo en las propuestas teóricas en torno a dicho tema procedentes de las distintas ciencias sociales. Muy al contrario, ante la pregunta acerca de qué es una institución, diversas corrientes teóricas han aportado respuestas diferentes, a veces contrapuestas. Sin embargo, a pesar de ello, y gracias al trabajo de Crawford y Ostrom (1995) mencionado anteriormente, es posible agrupar tales propuestas en torno a dos grandes bloques.

El primero de estos bloques es el que considera las instituciones como *equilibrios sociales*; basado en las propuestas teóricas de autores como Menger (1963) o Hayek (1945; 1967) y en elaboraciones contemporáneas como las de Riker (1980), Schotter

⁴ La noción de institución que aquí estamos empleando es bastante similar a la noción de “sistema” empleada por Giddens (1979; 1984).

(1981) o Sugden (1986). Este primer grupo de propuestas, cuya influencia sobre la teoría económica de las instituciones ha sido determinante, pone el énfasis en el componente de regularidad de las instituciones y, de forma muy resumida, entienden éstas como pautas estables de comportamiento en sociedad.

El segundo grupo de propuestas es más heterogéneo e incluye a dos subgrupos a los que Crawford y Ostrom (1995: 582) han denominado “enfoque de las *instituciones-como-normas*” y “enfoque de las *instituciones-como-reglas*”. El primero de ellos incluye a autores como Lewis (1969), Ullmann-Margalit (1977) o Coleman (1987; 1990a; 1990b), mientras que el segundo puede encontrarse en los trabajos de V. Ostrom (1980; 1987; 1991), Williamson (1985), North (1986; 1990) o E. Ostrom (1986; 1990). A pesar de las diferencias entre estos dos subgrupos, ambas aproximaciones “se centran en los constreñimientos lingüísticos (hablados, escritos o prescripciones o consejos entendidos tácitamente) que influyen sobre las preferencias y el comportamiento optimizador de los actores de forma conjunta” (Crawford y Ostrom, 1995: 582).

Nuestro objetivo es, en el caso concreto de las convenciones sociales, integrar en un único marco teórico y analítico ambas propuestas, dando importancia tanto al componente conductual como al componente normativo de las mismas. ¿Qué propuestas teóricas pueden servirnos para levantar el andamiaje de nuestra construcción teórica? Al igual que en el caso de las instituciones, las teorías que han tratado de dar cuenta de las convenciones son múltiples y, en multitud de ocasiones, contrapuestas. Por ejemplo, en la teoría económica que se apoya en la teoría de juegos una convención es una solución estable a un tipo concreto de problemas de interacción social: los problemas de coordinación. En este sentido, las convenciones se distinguirían de las

normas sociales porque, mientras que las primeras resuelven problemas de coordinación, las segundas resuelven otro conjunto clásico de problemas de interacción social: los denominados dilemas sociales⁵. Para otros autores, como Elster (1989), las convenciones serían una subclase de normas sociales. En realidad, también se podría decir lo mismo del primer enfoque presentado, aunque en el mismo hay una diferencia más clara entre normas y convenciones dependiendo de la estructura de interacción de la que cada una de estas instituciones emerge. Un tercer grupo de autores, cuyo principal exponente es Lewis (1969) defienden una concepción mucho más restringida de las convenciones sociales, circunscribiendo a estas a los denominados problemas puros de coordinación⁶. Si la de Lewis, como veremos en el capítulo 2, es una de las versiones más restringidas del alcance de la noción de *convención*, posiblemente sea la de Weber (1964) una de las más amplias. Para Weber una convención social es toda institución que no cae dentro del terreno de la moral o del ámbito de la ley.

Como se puede observar, la tarea de combinar las distintas definiciones de la noción de convención social resulta sencilla. En el capítulo 4 de esta tesis intentaremos abordar dicho problema, aunque ya avanzamos que nuestra posición se encuentra más cerca de la de Weber que de las anteriores.

Dicho esto, a continuación haremos una breve referencia a la metodología de la tesis, así como a la estructura general de la misma.

⁵ Para una defensa sociológica de este esquema ver el capítulo 10 de la obra de Coleman (1990), basado en la seminal obra de Ullmann-Margalit (1977).

⁶ No nos detenemos aquí a definir todos estos conceptos ya que serán ampliamente tratados en el resto de la tesis.

1.2. Metodología de la tesis

Antes de avanzar las herramientas concretas que van a ser empleadas en la tesis, diremos algo acerca del marco teórico y metodológico general en el que ésta se encaja. Esta tesis se apoya sobre el conjunto de críticas y reformulaciones que tanto desde la sociología, como de otras ciencias sociales se han hecho a la teoría de la elección racional en su versión más estándar en las últimas tres décadas. Así, se aprovecha de los avances producidos por la revolución que ha supuesto la “economía del comportamiento” (*behavioral economics*) en el campo de la economía (Camerer, 1999) o de su traslación a disciplinas como la ciencia política (Ostrom, 1998). Del mismo modo, pretende continuar la labor de proporcionar micro-fundamentos a las explicaciones de fenómenos sociales – en este caso, las convenciones sociales – enlazando con propuestas como la de Coleman (1990), Elster (1989) o, más recientemente, la indagación sobre “mecanismos sociales” concretos (Hedström y Swedberg, 1998) susceptibles de ser investigados tanto teóricamente y formalmente, como empíricamente. Una tercera fuente de inspiración es la necesidad de buscar explicaciones que integren conocimientos procedentes de diferentes disciplinas académicas⁷. Como se comprobará en los capítulos posteriores, esta tesis se vale de teorías, metodologías o herramientas de todo tipo procedentes tanto de la sociología, como de otras disciplinas afines como la economía, la ciencia política, la antropología, la psicología social e incluso la biología evolutiva. Por último, en la misma también se apuesta por una recuperación del método experimental como una herramienta privilegiada para el contraste de teorías sociales⁸. Dicho esto, a continuación

⁷ Sobre la posibilidad de una unificación de las explicaciones procedentes de las distintas ciencias del comportamiento, ver Gintis (2006).

⁸ Para una reflexión sociológica al respecto, ver Kanazawa (1999a).

introduciremos brevemente las herramientas concretas que se emplearán en las diversas partes de la tesis.

Un problema tan complejo como el estudio teórico de la noción de convención social requiere de una selección cuidadosa de las metodologías que van a ser empleadas, así como de la combinación que se hace de las mismas. Por ello, con el objetivo de buscar el mayor refinamiento metodológico posible, en esta tesis se van a realizar tres aproximaciones distintas a nuestro objeto de estudio: aproximación analítica, aproximación teórica y aproximación experimental.

Realizaremos una aproximación analítica al estudio de las convenciones sociales del mismo modo que ya lo han hecho con anterioridad autores como Ullmann-Margalit (1977), Elster (1989) o Coleman (1990) para el caso de las normas sociales. Para ello emplearemos la teoría de juegos estándar, así como ejemplos cotidianos, para reconstruir a modo de tipos ideales los conceptos que serán empleados en la tesis⁹. Con el objetivo de una mayor comprensión de los mecanismos que se hallan en la base de la generación y el mantenimiento de las convenciones sociales, emplearemos continuamente matrices de juegos y argumentos que no se apartan de la teoría de la elección racional; también señalaremos – allí donde sea necesario hacerlo – los límites de esta teoría para dar cuenta de las convenciones sociales. De este modo, no dudaremos en incorporar las matizaciones que tanto desde las teorías de la “racionalidad acotada”

⁹ Además de la noción weberiana de “tipo ideal”, nuestra aproximación es deudora de otras propuestas analíticas como el concepto de “reconstrucción racional” empleado por filósofos como Carnap (1962: 576-577) o Ullmann-Margalit (1977: 1). Esta última justifica su propuesta del siguiente modo: “Por reconstrucción racional de un evento, o con más exactitud, de una descripción de un evento, no debe entenderse una reconstrucción de las circunstancias históricas específicas en las que este evento ocurrió; ésta es la tarea del historiador o del arqueólogo. Más bien, es una descripción de las características esenciales de las situaciones en las que *dicho* evento *podría* ocurrir: es una historia de como algo podría suceder – y, cuando se trata de acciones humanas, del fundamento de de que acontezcan de esa manera – no de lo que realmente tuvo lugar.”

(*bounded rationality*) – de procedencia psicológica –, como desde la “teoría de juegos evolutiva” (*evolutionary game theory*) – cuyos resortes se hallan ligados a los avances producidos en la moderna biología evolutiva – se han realizado a la propia teoría de la elección racional.

La aproximación teórica al objeto de la tesis estará basada en la moderna teoría de juegos evolutiva¹⁰ que servirá para darnos algunas pistas teóricas acerca del porqué de la existencia determinados tipos de convenciones sociales. Antes de proceder a nuestro análisis evolutivo, justificaremos en qué sentido creemos que la teoría de juegos evolutiva es la herramienta óptima para abordar nuestro problema de investigación. La originalidad de nuestro análisis, en este sentido, reside en incorporar cuestiones de interés para la sociología al debate teórico en torno a la aplicación de modelos matemáticos procedentes de la biología al estudio de fenómenos económicos y sociales. En cuanto a la técnica en sí, ésta resultará bastante estándar y se ajusta en buena medida a la propuesta original de teoría de juegos evolutiva realizada por John Maynard Smith (1982).

Por último, las hipótesis extraídas a partir de la aplicación de las técnicas anteriores serán puestas a prueba experimentalmente. Debido a la escasa utilización del método experimental por parte de la sociología, el capítulo 6 de la tesis estará dedicado a argumentar en qué sentido el método experimental puede suponer un complemento importante para el conjunto de métodos y técnicas empleados actualmente por la disciplina sociológica.

¹⁰ En el capítulo 5 expondremos las diferencias fundamentales entre la teoría de juegos estándar y la teoría de juegos evolutiva. Sobre teoría de juegos evolutiva, ver el texto clásico de Maynard Smith (1982) o el más reciente de Weibull (1995).

Como se puede observar, la aportación novedosa que pretende realizar esta tesis, más allá del intento de combinación de explicaciones económicas y sociológicas, reside en el hecho de aplicar un conjunto de métodos y técnicas *en boga* en otras ciencias del comportamiento a un “problema sociológico fundamental”. En última instancia, en esta tesis, además de avanzar en la comprensión de la noción de convención social, pretendemos poner a prueba herramientas, como la teoría de juegos evolutiva o la economía experimental, que apenas están teniendo eco en las discusiones sociológicas contemporáneas. La validez del trabajo que aquí se presenta depende en gran medida de la paulatina incorporación de este conjunto de herramientas al arsenal teórico y metodológico propio del sociólogo. Por ello, antes de aplicar cada una de las mismas, haremos explícitos los supuestos que se hallan detrás de éstas para, de este modo, contribuir a que el resultado final pueda ser evaluado con la mayor transparencia posible.

1.3. Estructura general de la tesis

La presente tesis consta de tres partes y ocho capítulos. Tras el capítulo 1, que corresponde a esta introducción, la parte I abarca los capítulos 2, 3 y 4 y está dedicada a una aproximación analítica a la noción de convención social.

El capítulo 2 aborda la noción de convención social como *convención de comportamiento* predominante en la teoría económica contemporánea, partiendo de la noción original de convención en la obra de Hume (1978 [1740]). En este capítulo se reconstruye dicha noción a partir de los argumentos propuestos por autores como Lewis (1969), Ullmann-Margalit (1977), Schotter (1981), Sugden (1986; 1989), Binmore

(1994; 1998) o Young (1993; 1996; 1998a). Para exponer dichos argumentos se hace uso de matrices de juegos simples. El capítulo concluye con una exposición de las características básicas de la primera noción de convención social que estamos considerando.

El capítulo 3 reconstruye la noción de convención social como *convención normativa* propia de la teoría sociológica. Si en el capítulo anterior el punto de partida era Hume, en éste la reflexión parte de la definición de convención aportada por Weber (1964: 27). A partir de la recepción de ésta por autores como los pertenecientes a la denominada escuela francesa de la Economía de las Convenciones y, sobre todo, por la filósofa Margaret Gilbert, construimos nuestro tipo ideal de *convención normativa*. El capítulo concluye con un repaso a las características básicas de esta noción sociológica de convención.

El objetivo del capítulo 4 es proporcionar una definición de convención social capaz de integrar los argumentos expuesto tanto por economistas, como por sociólogos. Volviendo a la idea de los dos componentes inseparables de las instituciones, nuestra definición trata de tener en cuenta tanto los aspectos conductuales, como los aspectos normativos de las convenciones sociales. El capítulo concluye con una breve tipología de convenciones sociales derivada a partir de la definición propuesta. Este capítulo cierra la primera parte de la tesis.

La parte II está compuesta por un único capítulo, el capítulo 5, y en ella proponemos un marco teórico integrado que suponga una formalización del esfuerzo de aclaración conceptual realizado en la primera parte. En el mismo presentamos nuestra

*clase de juegos básica*¹¹ que, desde ese momento, se convertirá en una concreción formal de los argumentos de la tesis y que será empleada tanto en esta segunda parte, como en los experimentos que serán llevados a cabo en la tercera. Tras formalizar los distintos componentes de nuestra definición de convención social, llevamos a cabo un análisis evolutivo simple que nos permite obtener algunos resultados teóricos a partir del marco analítico definido. El capítulo concluye con una interpretación informal de los resultados alcanzados.

La parte III presenta los resultados obtenidos en dos experimentos realizados con el objetivo de poner a prueba los argumentos desarrollados con anterioridad. Pero antes de proceder a ello en el capítulo 7, hemos considerado conveniente realizar, en el capítulo 6, una descripción general del método experimental, así como una evaluación de su uso en sociología. De hecho, el capítulo 6 supone un paréntesis metodológico en una tesis dedicada íntegramente al estudio de las convenciones sociales. Sin embargo, la escasez de referencias metodológicas al uso de experimentos en sociología nos ha hecho decantarnos por dedicar un capítulo entero al mismo.

Una vez presentadas las consideraciones metodológicas generales acerca de la validez de los experimentos económicos en sociología, en el Capítulo 7 presentamos el diseño y los resultados de los dos experimentos a los que hacíamos referencia con anterioridad. En aras de la claridad expositiva, utilizamos la misma estructura para presentar ambos experimentos: primero el diseño, después los resultados y, por último, una discusión de los mismos a la luz de nuestra propuesta teórica. El capítulo finaliza con unas conclusiones que pretenden integrar los resultados obtenidos en ambos

¹¹ Por *clase de juegos básica* entendemos la representación formal – a través de una matriz de juegos – de la estructura de interacción subyacente al tipo de situación social que pretendemos analizar.

experimentos y que, en última instancia, se plantearon con el fin de poner a prueba empíricamente cuestiones relacionadas con la incertidumbre que rodea la naturaleza de las convenciones sociales.

El capítulo 8 de conclusiones cierra esta tesis. El objetivo del mismo no es otro que el de integrar los resultados obtenidos a partir de las distintas metodologías aplicadas, así como el de plantear las distintas líneas de investigación que pueden haber quedado abiertas a partir de esta tesis.

PARTE I

LA NOCIÓN DE CONVENCION SOCIAL

En el conjunto de las ciencias sociales contemporáneas no es posible identificar una definición técnica inequívoca y comúnmente aceptada de lo que intuitivamente calificamos como *convención social*. Así, preguntas del tipo: ¿qué regularidades observables en el comportamiento humano pueden ser calificadas como convenciones sociales?, ¿cómo surgen, se mantienen y evolucionan las convenciones sociales?, ¿cuál es la fuente de obligatoriedad de éstas?, entre otras muchas, sólo encuentran respuestas parciales procedentes fundamentalmente de la teoría económica, de la filosofía analítica y, en menor medida, de la teoría sociológica¹².

A pesar de lo anterior, la aparición en 1969 de la obra del filósofo analítico norteamericano David Lewis, *Convention. A Philosophical Study*, ha supuesto una importante recuperación del uso del término *convención*, al menos por dos motivos. En primer lugar, porque numerosos autores, fundamentalmente economistas, aceptaron rápidamente la definición de Lewis como válida e incorporaron el concepto de convención con el significado aportado por éste¹³. En segundo lugar, porque, más recientemente, un grupo de filósofos y sociólogos ha puesto en duda la validez del

¹² Una aproximación bastante estándar a la noción de convención desde la teoría económica puede encontrarse en Sugden (1998); Vanderschaarf (1998) supone un buen ejemplo de la aproximación al concepto de convención desde la filosofía analítica; Biggart y Beamish (2003) revisan la teoría de las convenciones sociales desde la sociología económica.

¹³ No es extraño que la propuesta de Lewis (1969) fuera bastante bien acogida por los economistas ya que, como veremos más adelante, su definición de convención está basada en la teoría de juegos de estrategia, especialmente la versión de ésta que se deriva del influyente libro *The Strategy of Conflict*, del recientemente galardonado con el premio Nobel de Economía Thomas Schelling (1960).

planteamiento lewisiano retomando una noción de convención anclada en los fundamentos metodológicos de la sociología de Max Weber¹⁴.

El objetivo de esta primera parte de la tesis es reconstruir, a modo de tipos ideales, estas dos aproximaciones a la noción de convención, para, a partir de ellas, proponer una definición, siempre provisional, de lo que en esta tesis entendemos por convención social. Por motivos que pronto se harán evidentes, utilizaremos la etiqueta de *convenciones de comportamiento* para referirnos a la primera de las aproximaciones referidas en el párrafo anterior y la de *convenciones normativas* para la segunda. Así, esta primera parte se articula del siguiente modo. En primer lugar, abordaremos la noción de *convención de comportamiento* predominante en la teoría económica. A continuación nos detendremos en la crítica, fundamentalmente sociológica, a esta aproximación conductual a la noción de convención, contraponiendo a ella una aproximación normativa, a la que denominamos *convención normativa*. Por último, a partir de las similitudes y diferencias entre ambas aproximaciones, propondremos nuestra propia definición de convención social.

¹⁴ Una de las paradojas que descubriremos a lo largo de esta parte de la tesis es que la crítica que estos filósofos y sociólogos contemporáneos realizan a la noción de convención predominante en la teoría económica actual es similar a la que el propio Weber realizaba a algunos filósofos del derecho alemanes de principios del siglo XX, tales como Stammler (1914). Sobre este hecho ya tendremos oportunidad de detenernos en las próximas páginas.

CAPÍTULO 2

LA TEORÍA ECONÓMICA Y LAS *CONVENCIONES DE COMPORTAMIENTO*

2.1. Introducción

Como hemos señalado anteriormente, la noción de convención entra en la teoría económica de la mano de la obra de Lewis (1969)¹⁵. Sin embargo, tanto éste como otros teóricos como Sugden, Skyrms o Binmore, justifican su posición recurriendo a la definición de convención aportada por el filósofo de la Ilustración escocesa David Hume, principalmente en su *Tratado de la naturaleza humana* (1978 [1740]). Lewis (1969: 3) lo argumenta del siguiente modo: “No obstante, en el fondo, la teoría de juegos sólo es el andamiaje. Puedo volver a plantear mi análisis de la convención sin ella. El resultado es una teoría en línea con la de Hume, en su discusión del origen de la justicia y la propiedad.” Sugden (1986: 33) afirma: “Si se me acusa de utilizar el término ‘convención’ de una manera excéntrica, puedo apelar al precedente de David Hume, quien – argumentaré – lo utiliza de una forma bastante similar a la mía.” Skyrms (1996: vi) abre su libro *Evolution of the Social Contract* con una cita del pasaje en el que Hume se refiere al lenguaje como un ejemplo de convención. Aunque Skyrms no comenta más adelante esta cita, asumimos que este autor también suscribe la posición

¹⁵ No todos los economistas estarían de acuerdo con esta afirmación y quizá necesite ser aclarada. La afirmación no tendría más problemas si sólo nos referimos a la corriente principal de la economía contemporánea (*mainstream economics*), fundamentada en los supuestos de la teoría de la elección racional y la teoría de juegos. Sin embargo, Latsis (2005: 710) se refiere al precedente en el uso de una noción de convención bien distinta por autores como Keynes (1973: capítulo 12) o Veblen. El uso que estos clásicos del pensamiento económico dan al concepto de convención poco tiene que ver con el que nos disponemos a desarrollar aquí bajo la etiqueta de *convenciones de comportamiento*. Para una interpretación reciente de la noción de convención en Keynes, ver Bibow, Lewis y Runde (2005).

humeana¹⁶. Por último, quizá sea Binmore (1998: 4) el que ha expresado de una forma más clara la reminiscencia humeana presente no sólo en la aproximación de estos teóricos económicos al estudio de las convenciones sociales, sino también en su concepción más general de los asuntos humanos: “Aquí, me uno a Hume, y a investigadores contemporáneos como Axelrod, Bicchieri, Gibbard, Hammond, Kliemt, Schotter, Skyrms y Sugden, quienes ven la teoría de juegos de Von Neumann y Morgerstern como el vehículo apropiado para expresar las ideas de Hume.”

La insistencia en el fundamento humeano de la noción de convención manejada por este conjunto de autores hace necesario que nuestra exposición comience con los párrafos en los que el propio Hume introduce su definición de convención. A continuación nos detendremos en la noción humeana de convención, como un primer paso para desgranar las principales características de lo que denominamos *convenciones de comportamiento*.

2.2. La propuesta original de Hume

Aunque el término convención, como su etimología latina muestra¹⁷, procede de la época romana¹⁸, fue David Hume (1978 [1740]) el que, en la parte segunda del libro

¹⁶ Así también lo han señalado recientemente tanto Binmore (1998: 4), como Cubitt y Sugden (2003: 181).

¹⁷ El término convención (del latín *conventio -onis*) era frecuente en el lenguaje jurídico romano con un significado similar al de acuerdo, tuviera éste un carácter tácito o explícito. Por ejemplo, apelando a la convención como un acuerdo tácito, la expresión *tacita civium conventio* era utilizada en el lenguaje jurídico para referirse a la fuerza vinculante de la costumbre. Como un acuerdo explícito el término convención puede ser encontrado en *conventio in manum*, una norma por la que la mujer es adquirida como una cosa, mediante la compra *per aes et libram*, entrando así en posesión del marido.

¹⁸ Que el término *convención* sea de procedencia romana no quiere decir que no existiera la noción de *convención* con anterioridad. Según Mosterín (2006: 150), “Respecto a las costumbres y las normas morales y políticas, nadie tenía una idea tan clara de su relatividad como los sofistas, viajeros infatigables y observadores ligeramente cínicos de la vida política y social de su tiempo. Ellos contribuyeron más que nadie a establecer la distinción entre convención o ley – *nómos* –, por un lado, y naturaleza – *Phycis* –, por otro.”

tercero de su *Tratado de la naturaleza humana*, presenta la que puede ser considerada como la primera noción moderna de convención¹⁹. El contexto en el que se produce este hecho es la crítica que Hume realiza a la idea de que la justicia debe quedar establecida por un contrato vinculante (Latsis, 2005: 710). De esta forma, el filósofo escocés toma postura en el debate a propósito del problema hobbesiano del orden social, planteando una solución a dicho problema diferente a la planteada por los teóricos del contrato. La solución planteada por Hume es lo que a partir de ese momento se conocerá como *convención*, una solución que “no se deriva [...] de la naturaleza, sino del artificio”²⁰; o bien, hablando con más propiedad: la naturaleza proporciona un remedio en el juicio y en el entendimiento para lo que resulta irregular e inconveniente en las afecciones” (Hume, 1978 [1740]: III.ii.2). Por tanto, la convención supone un remedio genuinamente *social* al problema hobbesiano²¹; pero, ¿cuáles son las características que definen la noción de convención propuesta por Hume?

A partir de la recepción de la propuesta humeana por los teóricos a los que nos hemos referido anteriormente, y de nuestra propia interpretación de los textos originales de Hume, podemos identificar cinco características definitorias de la noción humeana de

¹⁹ Las referencias explícitas al concepto de convención pueden ser encontradas en dos obras de Hume. Por una parte, en el ya citado *Tratado de la naturaleza humana* (Hume, 1978 [1740]), en el que la noción de convención aparece en la discusión acerca de dos asuntos, a saber: la emergencia de la propiedad privada y la fuente de obligatoriedad de las promesas. Por otra parte, también encontramos la noción de convención en el Apéndice Tercero de su *Investigación sobre los principios de la moral* (Hume, 1975 [1777]). En este último caso el término convención aparece ligado al origen de la justicia.

²⁰ Aquí, el sustantivo *artificio* debe ser entendido como *sociedad*; es decir, la convención es un concepto eminentemente social. En este sentido, utilizar el término *convención social* es redundante, en cuanto que toda convención es social y no podría existir una convención que quedara al margen de la interacción social entre los humanos. En todo caso el apellido *social* puede servirnos para distinguir el término convención en el sentido que se utiliza en esta tesis de otra de las acepciones habituales del mismo: la convención como una asamblea de representantes de un país, de un partido político, etc. Como esta segunda acepción no tiene cabida en este trabajo, en el mismo utilizaremos de forma indistinta los términos *convención* y *convención social*.

²¹ La contraposición entre la solución hobbesiana al problema del orden social y aquella propuesta por Hume radica en el hecho de que mientras el primero recurre a la idea de contrato para explicarlo, Hume recurre al argumento de la emergencia de la convención de forma gradual y paulatina, como veremos a continuación. Dicho esto, ambos autores postulan sendas soluciones “artificiales” al problema del orden. El primero a través del Leviatán u “hombre artificial” (Hobbes, 1993 [1651]), mientras que el Segundo a través de la convención que surge del “artificio” (la sociedad) (Hume, 1978 [1740]: III.ii.2).

convención, a saber: *conocimiento común, utilidad, reciprocidad, independencia respecto del lenguaje y emergencia gradual*.

En primer lugar, el supuesto de *conocimiento común*, el cual supondrá una de las características fundamentales de la definición de convención de Lewis (1969), puede ser encontrado en los propios textos de Hume, como se comprueba a continuación:

Y esto no puede hacerse de otra manera que mediante la convención, en la que participan todos los miembros de la sociedad, de conferir estabilidad a la posesión de estos bienes externos, dejando que cada uno disfrute pacíficamente de aquello que pudo conseguir gracias a la laboriosidad o suerte. De esta manera, todo el mundo sabe lo que le es posible poseer con seguridad, y las posesiones se ven restringidas en sus movimientos partidistas y contradictorios. (Hume, 1978 [1740]: III.ii.2)

La primera frase de este texto delata lo que poníamos de manifiesto con anterioridad: para Hume, la convención es el único remedio posible al problema del orden social, en general, y al establecimiento de un sistema que garantice la estabilidad de las posesiones propias y ajenas, en particular. Pero, no sólo es un sistema en el que todos los miembros de la sociedad participan, y que supone un modo de interacción pacífica entre los mismos, sino también uno en el que “todo el mundo sabe” de su funcionamiento. Este conocimiento común o compartido de la convención será un elemento crucial para entender las interpretaciones posteriores del planteamiento humeano. No abundaremos más sobre el mismo en este momento.

En segundo lugar, la convención surge a partir de un “sentimiento general de interés común”; en otras palabras, la convención está sustentada sobre un criterio de *utilidad*. La referencia a dicho “sentimiento general de interés común” puede ser localizada en el siguiente texto:

La convención consiste únicamente en un sentimiento general de interés común: todos los miembros de la sociedad se comunican mutuamente este sentimiento, que les induce a regular su conducta mediante ciertas reglas. Yo me doy cuenta de que redundará en mi provecho el que deje gozar a otra persona de la posesión de sus bienes, dado que esa persona actuará de la misma manera conmigo. (Hume, 1978 [1740]: III.ii.2)

Como ha señalado Young (1998b: 822), “Hume reconoció, además, que tales convenciones sólo pueden ser sostenidas si constituyen un equilibrio de un juego definido apropiadamente.” En la moderna teoría de juegos, el denominado equilibrio de Nash está formado por un conjunto de estrategias individuales que constituyen la mejor respuesta de cada individuo frente a las posibles estrategias del resto. Alcanzado ese equilibrio ningún individuo tiene incentivos para cambiar de estrategia (Morrow, 1994: 80). Para hacer más explícita la equiparación de la noción humeana de convención a un equilibrio de Nash, observemos el siguiente fragmento. En el mismo aparecen dos elementos – constitutivos de la noción de equilibrio – que serán fundamentales en la interpretación posterior de Hume, a saber: en primer lugar, al seguir la convención no tenemos por qué desviarnos de nuestro interés particular; en segundo lugar, la convención favorece tanto nuestros intereses como los de los demás.

En lugar de abstenernos de la propiedad ajena apartándonos de nuestro propio interés o del de nuestros amigos más íntimos, no hay mejor modo de atender a ambos intereses que mediante una convención tal, porque es de ese modo como se sostiene la sociedad, tan necesaria para la buena marcha y subsistencia de los demás como para la nuestra. (Hume, 1978 [1740]: III.ii.2)

Si las citas anteriores no son suficientes para hacer evidente la importancia de la noción de *utilidad* para la definición humeana de convención, en la única referencia a la misma que encontramos en la *Investigación sobre los principios de la moral* (1975 [1777]) Hume hace una referencia explícita a la “utilidad pública” de la convención:

Pero si por convención quiere aquí decirse un sentido de interés común, sentido que cada hombre experimenta en lo hondo de su pecho, que nota en sus prójimos, y que le lleva, en

conurrencia con otros, a un plan general o sistema de acciones que tiende a la utilidad pública, entonces debe admitirse que, en este sentido, la justicia surge de convenciones humanas. Pues si admitimos (lo cual es, desde luego, evidente) que las consecuencias particulares de un acto particular de justicia pueden hacer daño tanto al público general como a los individuos, se sigue de ello que todo hombre que abraza esta virtud debe tener los ojos puestos en la totalidad del plan o sistema, y debe esperar que sus prójimos concurren en el mismo comportamiento o conducta. (1975 [1777]: App. 3)

En tercer lugar, la noción de convención en Hume requiere de la presencia de cierta *regla de reciprocidad* (Binmore 2005: 77). Como se leía en un párrafo ya introducido anteriormente, Hume lo expresa del siguiente modo: “Yo me doy cuenta de que redundará en mi provecho el que deje gozar a otra persona de la posesión de sus bienes, dado que esa persona actuará de la misma manera contigo.” También podemos encontrar implícito el recurso a un criterio de reciprocidad en el siguiente texto:

Y esto es lo que puede ser denominado con bastante propiedad convención o mutuo acuerdo, aun cuando no exista la mediación de una promesa, dado que las acciones de cada uno de nosotros tienen referencia a las del otro y son realizadas en el supuesto de que hay que realizar algo a favor de la otra parte. (Hume, 1978 [1740]: III.ii.2)

La cuarta característica expresada anteriormente, la *independencia respecto del lenguaje*, ha sido definida por Latsis (2005: 712) del siguiente modo: “[la idea humeana de convención] está basada en un tipo de razonamiento inductivo, de prueba y error, donde el lenguaje no es esencial.” Más adelante afirma que para Hume la convención supone un medio de “coordinación no lingüística en torno a una regularidad de comportamiento mutuamente beneficiosa.” Esta idea ha servido a Lewis (1969: 160) para introducir su concepción de “convenciones de lenguaje”, en la que, como el propio Hume hace en el siguiente texto, señala que existen convenciones que preceden al propio lenguaje y no necesitan ser formuladas de forma lingüística para que sean seguidas dentro de una población. Dicha argumentación puede encontrarse especialmente en la insistencia humeana en la distinción entre convención y promesa: “Esta convención no tiene la naturaleza de una *promesa*, pues hasta las mismas

promesas – como veremos más adelante – surgen de convenciones humanas” (Hume, 1978 [1740]: III.ii.2). Más adelante, hablando del carácter obligatorio de las promesas, dirá: “*que una promesa no sería inteligible antes de ser establecida por las convenciones humanas; y que, aun en el caso de ser inteligible, no estaría acompañada de ninguna obligación moral*” (Hume, 1978 [1740]: III, ii, 5).

En quinto lugar, de los fragmentos considerados de Hume se deriva que las convenciones emergen de forma *gradual*; actualmente diríamos que surgen a través de un proceso *de prueba y error* prolongado en el tiempo. Para ejemplificar este hecho, hemos seleccionado uno de los extractos que han sido citados en un mayor número de ocasiones para introducir la noción humeana de convención:

Quando dos hombres impulsan un bote a fuerza de remos lo hacen en virtud de un acuerdo o convención, a pesar de que nunca se hayan prometido nada mutuamente. No menos se deriva de las convenciones humanas la regla de la estabilidad en la posesión, por surgir gradualmente e ir adquiriendo fuerza mediante una lenta progresión, y porque experimentamos repetidamente los inconvenientes que resultan de transgredirla. (Hume, 1978 [1740]: III.ii.2)

Es la experiencia repetida de los inconvenientes que resultan de desviarnos de la convención lo que hace que sigamos respetándola. En esta idea podemos encontrar el germen de la moderna interpretación de la convención como una pauta de comportamiento estable en el tiempo, surgida a partir de mecanismos evolutivos y de aprendizaje adaptativo. No es extraño que uno de los principales valedores de la concepción evolutiva de las convenciones, Brian Skyrms, haya elegido este fragmento para abrir su libro *Evolution of the Social Contract* (1996: vi).

Esta brevísima referencia al concepto humeano de convención nos permitirá comprender la procedencia de algunas de las ideas que se introducen en este apartado. A

partir de este momento nos ocuparemos de lleno de la reconstrucción ideal de lo que hemos denominado *convención de comportamiento*.

2.3. David Lewis y las convenciones como regularidades de hecho en el comportamiento humano

Una de los supuestos que no admiten discusión en el seno de la teoría económica contemporánea es que una convención supone una regularidad social, en tanto que implica la pauta de comportamiento efectivamente seguida por una población al encontrarse ante una situación recurrente de interacción social (ver, por ejemplo, Schotter, 1981: 10; Young, 1996: 105). Así, las convenciones son pautas de comportamiento empíricamente observables. Por ejemplo, saludar por la mañana con un simple “buenos días” es la convención existente en mi centro de trabajo; ahora, si el encuentro se produce fuera del trabajo, la convención es “dar dos besos” si una de las partes es una mujer y “estrechar las manos” cuando ambas partes son hombres. Generalmente todas las reglas relacionadas con el “decoro social” y con las “buenas costumbres” son consideradas convenciones, ya que suponen la regularidad de comportamiento predominante en una situación de interacción social determinada. Pero, las convenciones no se circunscriben a cuestiones relacionadas con la educación o el civismo, sino que podemos encontrarlas en cualquier momento y lugar. Pensemos en el siguiente ejemplo, donde varias convenciones desempeñan un papel muy importante en el ámbito económico.

Durante décadas era *costumbre* entre los propietarios de tierras en el sur de España reclutar campesinos en *la plaza del pueblo* para que trabajaran en sus tierras a cambio

de un *jornal*. No existía ningún tipo de contrato, ni era normal la existencia de una convocatoria explícita; es más, aunque el contacto entre propietarios y *jornaleros* se producía de forma regular en un lugar concreto – “la plaza del pueblo” – y en un momento determinado del día – “al amanecer” –, esto no implicaba que existiera ninguna obligación para las partes y cada día el encuentro podía producirse o no llevarse a cabo. Pero el hecho es que dicha pauta de comportamiento se repetía y, regularmente, tanto propietarios como *jornaleros* se encontraban en un *mercado de trabajo* que, aunque improvisado, estaba provisto de multitud de convenciones referentes a los espacios, los tiempos, los intercambios de toda índole, etc. ¿Cómo era posible la emergencia de tales reglas sin la concurrencia de acuerdos explícitos referentes a cada uno de los asuntos en cuestión? En términos generales, esta es la pregunta que se halla en la base del trabajo de Lewis (1969). La solución que este autor propone – como veremos – requiere de la existencia de dos elementos, a saber: un *precedente* que marca la pauta a seguir en un problema de interacción social recurrente y *conocimiento común* del mismo por parte de los participantes en la regularidad que se quiere explicar.

Para Lewis (1969), las convenciones emergen a partir de situaciones *no triviales*, es decir, situaciones que representan un problema cuya solución no es trivial – la solución no se deriva directamente de la propia estructura del problema –, como es el caso de los múltiples problemas de coordinación que tenían que resolver propietarios y jornaleros en el ejemplo introducido anteriormente. En general, los denominados problemas de coordinación representan situaciones en las que “los actores tienen intereses similares y, aunque puede no importarles qué solución será la impuesta, están de acuerdo en que algún tipo de solución es necesaria” (Wilson y Rhodes, 1997: 767). Así, tanto los propietarios de tierras, como los jornaleros no preferían encontrarse en la

plaza por encima de encontrarse en el parque, por ejemplo, aunque ambos preferían, desde luego, coincidir en un punto de encuentro. Un problema similar es el que se da en cualquier mercado entre vendedores y compradores, o en cualquier mercado de trabajo entre empresarios y trabajadores. De ahí la emergencia de ciertas regularidades en el establecimiento espacial de determinados comercios – como la presencia de firmas de moda internacionales en los Campos Elíseos de París o la concentración de empresas de tecnología punta en Silicon Valley –, o de regularidades relativas a tiempos y horarios – como la elección del domingo para montar numerosos mercadillos callejeros. Este tipo de situaciones pueden ser definidas como situaciones puras de coordinación (Schelling, 1960; Lewis, 1969: 14); un tipo de situaciones donde predomina la coincidencia de intereses²². Dado este tipo de situaciones, el objetivo de Lewis es plantear que los acuerdos explícitos no suponen, desde un punto de vista lógico, el único mecanismo para la resolución de problemas de coordinación. Aunque está claro que un acuerdo supone la forma más directa de generar el conjunto de *expectativas mutuas concordantes* necesarias para la resolución de problemas de coordinación (Miller, 2007), este tipo de expectativas también pueden ser generadas a través de las convenciones, entendidas éstas como regularidades de hecho en el comportamiento humano. Veamos esto a través de la representación formal del ejemplo de la coordinación entre propietarios y jornaleros, para, posteriormente, adentrarnos en la definición de convención aportada por Lewis (1969: 47), así como en las implicaciones de la misma.

²² Esta “coincidencia predominante de intereses” debe ser aclarada. Cuando afirmamos que tanto propietarios como jornaleros prefieren coincidir en un punto de encuentro sólo estamos diciendo que prefieren coincidir en esto y nada se dice sobre la coincidencia o el conflicto en cualquier otro problema de interacción que pueda surgir. Lo mismo se puede decir de compradores y vendedores, de trabajadores y empresarios, etc.

Para simplificar, supongamos que tanto jornaleros como empresarios pueden dirigirse cada mañana a dos lugares del pueblo para ponerse en contacto: la plaza y el parque. Si ambos coinciden en alguno de estos dos lugares obtienen un pago, derivado de la relación que se establece entre ambos, y si no coinciden no obtienen ningún pago. Dada esta estructura de preferencias, ambos son indiferentes entre encontrarse en la plaza o encontrarse en el parque, siempre que cada mañana esté claro dónde han de encontrarse (ver Figura 1)²³. Pero, dada dicha estructura de pagos, ¿cómo podemos explicar la emergencia de una convención que suponga que la regularidad de comportamiento es encontrarse en la plaza y no en el parque?

Figura 1: Situación pura de coordinación²⁴

		<i>Jornalero</i>	
		<i>Plaza</i>	<i>Parque</i>
<i>Propietario</i>	<i>Plaza</i>	1, 1	0, 0
	<i>Parque</i>	0, 0	1, 1

Como avanzábamos, la solución aportada por Lewis a la pregunta anterior es la siguiente: los actores resuelven problemas de coordinación recurriendo al *conocimiento común* de un *precedente* que indica cuál es la solución *establecida* del problema en cuestión. Es decir, si un propietario quería disponer de mano de obra en un momento determinado, el hecho de reconocer un *precedente*, tanto si había sido experimentado

²³ Alguien podría decir que el hecho de representar la relación entre propietarios y jornaleros como un juego simétrico está completamente alejado de la realidad, debido a la enorme desigualdad que en términos reales existe entre ambos. Esto no supone ningún problema para el argumento que estamos desarrollando, ya que aunque el hecho de ponerse en contacto reporte al propietario una utilidad enorme – digamos 100 – y al jornalero siga reportándole 1, ambos prefieren encontrarse a dirigirse a lugares por separado.

²⁴ Las combinaciones de estrategias que suponen un equilibrio aparecen en negrita. Este juego también tiene un equilibrio en estrategias mixtas (ENEM) en el que cada jugador elige dirigirse a la plaza con probabilidad $\frac{1}{2}$ y dirigirse al parque con esta misma probabilidad. Como veremos más adelante, al hablar de la estabilidad de las convenciones, la interpretación de las convenciones en los términos de una combinación probabilística de estrategias resulta bastante problemático.

por él mismo como si se lo había transmitido una tercera persona, hace que sepa que la *convención establecida* es ir a buscar a los jornaleros a la plaza del pueblo. Además, la existencia de dicho precedente debe ser de *conocimiento común* entre los actores implicados del siguiente modo: el propietario debe conocer que la convención supone que los jornaleros esperan la llegada de los propietarios – o de los empleados de estos que actúan de capataces – en la plaza del pueblo, los jornaleros deben conocer también este hecho, los propietarios deben saber que los jornaleros conocen este hecho, los jornaleros deben saber que los propietarios saben que los jornaleros conocen este hecho y, así, *ad infinitum*²⁵.

El supuesto de la existencia de un precedente permite a Lewis solucionar problemas de coordinación como el presentado en la Figura 1. El supuesto de conocimiento común le permite dotar a la convención establecida – por el precedente – de la estabilidad propia de un equilibrio de Nash: cualquier desviación de la misma supone la obtención de un resultado subóptimo. Desviarse de la convención, en este sentido, no sólo perjudica al resto de participantes en la misma, sino que principalmente perjudica al propio actor que se desvía. De este modo, la convención supone una regularidad que se refuerza a sí misma – recordar la característica de *utilidad* de la convención en Hume. La definición exacta que Lewis aporta es la siguiente:

Una regularidad R en el comportamiento de los miembros de una población P siendo agentes en una situación recurrente S es una *convención* si y sólo si, en cualquier caso de S

²⁵ No está muy claro a quién hay que atribuir la primera formulación de la noción de conocimiento común en la Historia de la Teoría de Juegos. Parece aceptarse que la primera caracterización informal de este concepto fue realizada por el propio Lewis (1969); aunque, el también filósofo Robert Nozick afirma que este concepto apareció por primera vez en su tesis doctoral en 1963 (ver Cubitt y Sugden, 2003). Sin embargo, la noción de conocimiento común que se ha impuesto como la estándar en la teoría de juegos es la formulación matemática de la misma realizada por Aumann (1976), así como la axiomatización llevada a cabo por Milgrom (1981). Para una representación de ésta como un “sistema interactivo de creencias” ver O’Neil (2000); para una interesante representación “gráfica” del proceso de generación de conocimiento común ver Chwe (2001: apéndice).

entre los miembros de P , (1) todos se conforman con R ; (2) todos esperan que todos los demás se conformen con R ; (3) todos prefieren conformarse con R a condición de que los otros lo hagan, dado que S es un problema de coordinación y la conformidad uniforme con R es un equilibrio propio de coordinación en S . (Lewis, 1969: 47)

Como se observa, la definición técnica aportada por Lewis contiene los elementos que han sido introducidos intuitivamente hasta el momento. La convención es una regularidad en el comportamiento de los miembros de una población determinada, en la que: (1) no existen incentivos individuales para desviarse de la misma; por tanto, (2) todo el mundo tiene la expectativa de que el resto actuará de acuerdo con la convención; y, así, (3) la convención se convierte en una situación estable, debido a que el seguimiento de la misma – condicionado a que el resto de participantes también la sigan – supone una situación individualmente óptima, es decir, una situación de equilibrio²⁶.

Las tres características presentadas de la propuesta lewisiana – convención como regularidad, existencia de un precedente y conocimiento común – han sido objeto de numerosas críticas. En este momento sólo nos vamos a referir a los aspectos más problemáticos que plantean cada una de ellas; realizaremos una crítica más profunda en el siguiente capítulo.

La primera crítica se refiere a la afirmación por la cual *una convención es una regularidad social en tanto que supone la pauta de comportamiento realmente seguida por una población al encontrarse ante una situación recurrente de interacción social*. Como argumentaremos más adelante, ésta es una de las características más

²⁶ La presentación realizada de la propuesta de Lewis supone sólo un desarrollo esquemático de su definición de convención, haciendo hincapié en los puntos que más han influido en la noción de convención manejada por los teóricos económicos contemporáneos. Nada se afirma aquí de otros objetivos de su obra, tales como su propuesta de “convenciones de lenguaje.”

problemáticas de la definición de convención social asumida por los teóricos económicos y teóricos de la elección racional en general. Así, ¿implica esta condición que para que una regularidad de comportamiento en una población de tamaño n sea considerada una convención social tiene que ser seguida por n personas? Y si la regularidad es seguida por $n-1$ personas, ¿seguiríamos hablando de convención social? Sin entrar en más detalles por el momento, diremos que, en general, los defensores de esta definición de convención se conforman con un supuesto más modesto que no exige la unanimidad en el seguimiento de la convención, sino, más bien, un cierto seguimiento mayoritario de la misma. El propio Lewis pone este hecho de manifiesto en la última de las definiciones de convención propuestas en su libro:

Una regularidad R en el comportamiento de los miembros de una población P siendo agentes en una situación recurrente S es una convención si y sólo si es verdad que, y es conocimiento común en P que, en casi cualquier caso de S entre los miembros de P , (1) casi todos se conforman con R ; (2) casi todos esperan que casi todos los demás se conformen con R ; (3) casi todos tienen aproximadamente las mismas preferencias en relación con todas las posibles combinaciones de acciones; (4) casi todos prefieren que cualquier otro se conforme con R , a condición de que casi todos los demás se conformen con R ; (5) casi todos preferirían que cualquier otro se conforme con R' , a condición de que casi todos se conformen con R' , donde R' es alguna regularidad posible en el comportamiento de los miembros de P en S , tal que casi ninguno en casi ningún caso de S entre los miembros de P podría conformarse con ambas R' y R . (Lewis, 1969: 78)

Con esta nueva definición Lewis se adelanta a algunos de los autores que años más tarde han criticado su concepción de las convenciones como regularidades de comportamiento: no es necesario la conformidad unánime con la convención para que ésta sea considerada como tal. Sin embargo, la insistencia de igualar convención a regularidad de comportamiento sigue dejando una cuestión abierta: ¿cómo es posible que la simple expectativa de que un fenómeno aparezca con cierta regularidad genere automáticamente la adhesión y la conformidad de los miembros de una población? La respuesta que Lewis – también Hume – daría a esta pregunta podría ser la siguiente: *porque está en su interés* – porque es útil, porque supone un equilibrio. Para que esto

sea posible Lewis necesita el que ahora se ha convertido en el supuesto (3) de su definición: que todo el mundo tenga aproximadamente las mismas preferencias, es decir, que nos encontremos aproximadamente ante una situación pura de coordinación como la descrita en la Figura 1. Pero intuitivamente no está claro que esto tenga necesariamente que ser así. Más bien parece que se trate más de una cuestión susceptible de ser abordada empíricamente, que de un supuesto teórico. Gilbert (1989: 342) se muestra contundente a este respecto: “las convenciones sociales, concebidas intuitivamente, pueden cubrir situaciones de toma de decisiones interdependientes que no suponen problemas [puros]²⁷ de coordinación como los definidos por Lewis.” Por tanto, necesitaremos de *algo más* que de regularidades predecibles de comportamiento para dar cuenta de la fuerza vinculante que intuitivamente parecen tener las convenciones sociales, ya que “... si las convenciones sociales no tienen por qué cubrir lógicamente los problemas [puros] de coordinación, su necesario aspecto normativo difícilmente puede ser expresado en los términos de dicha estructura subyacente” (Gilbert, 1989: 355).

La segunda crítica al planteamiento de Lewis se refiere al argumento según el cual *la convención queda establecida por la existencia de un precedente*. El problema en sí no es que no seamos capaces empíricamente de correlacionar la existencia de una regularidad actual con la presencia de un precedente en el pasado, ni que no podamos

²⁷ Es importante tener en cuenta que la crítica de Gilbert va contra la idea de que las convenciones sólo emergen a partir de problemas puros de coordinación, es decir problemas – como el representado en la Figura 1 – donde existe una coincidencia total de intereses entre los participantes en lo que se refiere al resultado del juego. En esta tesis vamos a seguir considerando los problemas de coordinación como estructuras de interacción que se hallan en la base de la emergencia de las convenciones sociales, pero no nos restringiremos sólo a problemas puros de coordinación, sino que fundamentalmente nos ocuparemos de los denominados problemas de coordinación asimétrica, donde ambos jugadores pueden tener preferencias distintas en relación con el resultado del juego. En este sentido sí que estamos de acuerdo con Gilbert que para que la teoría lewisiana de las convenciones tuviera un carácter general debería dar cuenta de los mecanismos que garantizan la estabilidad de las convenciones en entornos asimétricos. Ésta será una de las principales preocupaciones teóricas de las partes segunda y tercera de esta tesis.

trazar la línea que une ambos fenómenos. El problema surge cuando nos preguntamos por el *cómo* y el *cuándo* de la emergencia del precedente, sin caer en una regresión hasta el infinito, donde cada precedente quedaría justificado por la existencia de un precedente anterior. Hay que tener en cuenta que, en el lenguaje de Lewis, un precedente supone solamente un caso específico de un mecanismo psicológico más general denominado “prominencia” (*salience*). Decimos que un equilibrio – combinación de estrategias – es prominente cuando “sobresale del resto por su singularidad en algún aspecto conspicuo” (Lewis, 1969: 35). Este recurso explicativo a la prominencia es bastante frecuente en la teoría de juegos contemporánea. Cuando no se sabe muy bien cómo dos actores han logrado coordinar sus estrategias de forma satisfactoria es frecuente recurrir al argumento de que el equilibrio alcanzado es *prominente*, o lo que es lo mismo, que constituye un *punto focal*. Pero, como ha señalado Gilbert (1989: 334): “dentro del marco de la teoría de juegos el supuesto de que algún punto es prominente no nos lleva a ningún lugar. Es decir, no lleva el razonamiento de los actores hacia la siguiente conclusión: actúa de esta forma.” Por tanto, la vinculación de la existencia de la convención a la existencia de un precedente, aunque empíricamente plausible, no nos proporciona una explicación última acerca de la emergencia de las convenciones sociales.

La tercera crítica del planteamiento lewisiano en la que nos vamos a detener en este momento es la *necesidad de la existencia de conocimiento común entre los participantes en la convención*. En realidad, una vez puesto en duda el argumento de la convención basada en un precedente, el supuesto de conocimiento común no tiene tanto sentido en el planteamiento de Lewis, ya que la función del supuesto de conocimiento común aquí es la de perpetuar una combinación de estrategias que han surgido a partir

del reconocimiento común de la existencia de un precedente que establece cuál es la convención en una situación de interacción social concreta. Un segundo problema más de fondo es lo que Latsis (2005: 718) denomina *falacia del modelo mental*, que implica que “para que se genere una convención, un agente debe conocer, o ser consciente de lo que él o ella está haciendo.” No parece que nuestra experiencia cotidiana de las convenciones implique que seamos conscientes en todo momento de la convención que estamos siguiendo. Por tanto, el supuesto de conocimiento común deberá ser, cuando menos, relajado.

La referencia a la obra de Lewis nos ha permitido dar un paso más en la propuesta que derivábamos de Hume. A la misma, el filósofo norteamericano aporta varios elementos: la convención como una regularidad de comportamiento social; la interpretación de la misma como un equilibrio en un juego puro de coordinación; y un mecanismo explicativo basado en el conocimiento común de un precedente, el cual ha sido puesto en duda por autores posteriores. En el siguiente apartado, avanzaremos en la reconstrucción de las *convenciones de comportamiento*, a partir de otra de las características fundamentales de las mismas: el carácter *arbitrario e indeterminado* de las convenciones sociales.

2.4. Las convenciones sociales como regularidades arbitrarias e indeterminadas

La corriente principal de la teoría de juegos, al menos desde la seminal obra de Von Neumann y Morgenstern (1947), ha concentrado sus esfuerzos en plantear los juegos como interacciones estáticas entre jugadores perfectamente racionales, en un entorno en el que la función de utilidad y el comportamiento racional de cada uno de los jugadores

son de conocimiento común. Dicho intento analítico está encaminado a la persecución de un objetivo claro y preciso: mostrar que el análisis racional prescribe una única estrategia para cada jugador en un juego determinado (Sugden, 1989: 88). Sin embargo, esta estrategia analítica, cuyo máximo exponente es el trabajo conjunto de los premios Nobel Harsanyi y Selten (1988), se ha visto seriamente obstaculizada por la dificultad que entraña analizar situaciones donde el puro análisis de las acciones racionales de los jugadores no conduce a una solución única²⁸. Como ya hemos visto, cuando la interacción entre dos o más jugadores perfectamente racionales da lugar a más de una solución en potencia, nos encontramos ante los denominados *problemas de coordinación*. En general, los problemas o juegos de coordinación, como veíamos anteriormente, representan situaciones en las que varios actores tratan de alcanzar uniformidad en sus acciones haciendo cada uno lo que espera que haga el resto. Más arriba, utilizábamos como ejemplo la coordinación espacial y temporal entre propietarios de tierra y jornaleros en el mundo rural. A continuación, seguiremos avanzando en la construcción de la noción de *convención de comportamiento* a través de uno de los ejemplos más recurrentes en la literatura económica sobre el tema: la coordinación entre conductores.

Supongamos que estamos conduciendo por una carretera comarcal donde no existe separación entre carriles en alguna provincia española y que nos encontramos de frente con otro coche que se aproxima hacia el nuestro. Existe el espacio justo para que, a duras penas, puedan pasar dos coches al mismo tiempo. ¿Por qué lado de la vía

²⁸ Aunque el propósito analítico de la corriente principal de la teoría de juegos ha perseguido el objetivo de la definición de una única solución racional para cada juego, incluso los “padres fundadores” de la misma asumían que, más allá del análisis teórico-ideal, en la práctica, esta empresa no era posible. Así, “Von Neumann y Morgenstern sostuvieron que cuando quiera que existe más de una solución teórica a un juego, la solución que es realmente alcanzada depende de normas y convenciones (o, como ellos lo expresaron, de los ‘estándares de comportamiento’) en la comunidad a la que los jugadores pertenecen” (Luce y Raiffa, 1957: 205, 219; Ullmann-Margalit, 1977: 15).

deberíamos conducir? En España conduciríamos por la derecha, esperando que el otro conductor haga lo mismo. Pero, ¿cuál es la fuente de dicha expectativa? Si no se tratara de una situación real, sino de una situación puramente ideal, atemporal y descontextualizada, como suelen ser representadas en la teoría de juegos estándar (ver Figura 2), cualquier elemento que permitiera a ambos conductores coordinar sus expectativas cumpliría la función de resolver una situación que desde un punto de vista formal resulta irresoluble. Tanto es así que, si en vez de tratarse de dos conductores circulando por una carretera comarcal española, se tratara de dos conductores irlandeses conduciendo por una carretera de su país, ambos elegirían circular por la izquierda con la misma naturalidad que nosotros lo hacemos por la derecha.

Como ocurría con el ejemplo del encuentro entre propietarios de tierra y jornaleros, cada una de las soluciones del juego representado en la Figura 2 supone una convención *en potencia* y aquella que es la predominante en un espacio social determinado – circular por la derecha en España – será la convención *establecida*.

Figura 2: Juego de la conducción²⁹

		<i>Conductor A</i>	
		<i>Izquierda</i>	<i>Derecha</i>
<i>Conductor B</i>	<i>Izquierda</i>	1, 1	-1, -1
	<i>Derecha</i>	-1, -1	1, 1

Esta es la forma más simple de dar cuenta de la noción de *convención de comportamiento* que podemos encontrar: en una situación social donde existe una multiplicidad de pautas de comportamiento posibles, la convención supone aquella que

²⁹ Como en la Figura 1, los equilibrios aparecen resaltados en negrita. Este juego también tiene un equilibrio en estrategias mixtas (ENEM) similar al del juego de la Figura 1. Ver comentario sobre el mismo en la nota 25.

es seguida en la práctica. Ejemplos recurrentes de convención en este sentido son el hecho de adherirse a un determinado modo de vestir – como vestir de negro en un funeral –, la utilización de palabras con su significado convencional o numerosos aspectos relacionados con el mercado y la actividad económica, tales como: el uso de distintos tipos de dinero, el uso de estándares industriales y tecnológicos o las reglas de contabilidad³⁰.

Los teóricos de juegos han empleado el concepto de arbitrariedad (*arbitrariness*) para hacer referencia a esta característica de las convenciones que implica que cada convención supone sólo una de las múltiples regularidades de comportamiento que pueden emerger a partir de una determinada situación de interacción social. Por ejemplo, Binmore (1994: 290) afirma: “Afortunadamente, las sociedades humanas tienen abundantes *convenciones* [...]. En puridad, estas convenciones son completamente arbitrarias. La gente no queda en el ‘punto de reunión’ de un aeropuerto porque piensa que allí la reunión será más placentera, sino porque es convencional hacerlo. Lo mismo se puede decir de otras convenciones como ‘circular por la derecha’ o ‘poner cosas en orden alfabético’”. Otras referencias explícitas al carácter arbitrario de las convenciones pueden ser encontradas en Lewis (1969: 70) y en Coleman (1990: 248), entre otros.

Las convenciones son arbitrarias en el sentido de no ser necesarias, como arbitrarias son las normas legales o la propia moral de cada grupo social. Por tanto, a

³⁰ El éxito de mercados legales, como la prostitución, o directamente ilegales, como el tráfico de drogas sólo puede ser entendido a partir de la comprensión de las numerosas convenciones en que se sostienen. Por ejemplo, existen convenciones acerca de cómo ponerse en contacto con un vendedor de droga al por menor, sobre cómo negociar el precio de la misma con él, incluso sobre el tipo de lenguaje que se ha de utilizar. Al no tratarse de un mercado regulado, parece evidente que las convenciones desempeñan un papel fundamental en la articulación de los conocimientos tácitos de los participantes en este tipo de transacciones económicas.

pesar de lo que parece deducirse de buena parte de la teoría económica que se ha preocupado de las convenciones, la arbitrariedad no es una característica distintiva de las mismas. En todo caso, podríamos decir que las convenciones tienen un *rango limitado de arbitrariedad*, en el sentido de que, aunque podemos imaginar convenciones alternativas a una convención establecida, el número de éstas no suele ser ilimitado. Así, la convención alternativa a “conducir por la derecha” sería “conducir por la izquierda”, la alternativa al saludo mediante dos besos en España sería la convención de dar tres besos en Francia o la de dar uno en Estados Unidos. Por tanto, afirmar, como hace Binmore, que las convenciones son “completamente arbitrarias” tampoco parece lo más correcto.

En realidad, como ya hemos señalado anteriormente, cuando Binmore y otros teóricos hacen referencia al carácter arbitrario de las convenciones sólo están poniendo el énfasis en la existencia de posibles convenciones alternativas a la convención establecida. En este sentido sería más apropiado hablar del carácter *indeterminado* de las convenciones sociales. Su propia condición de soluciones estables a problemas de coordinación y, por tanto, la multiplicidad de soluciones potenciales a éstos, hacen que no podamos predecir teóricamente cuál es la convención que se establecerá en un contexto social determinado. En definitiva, el problema de qué convención emergerá como solución a un problema de coordinación concreto nos remite a una cuestión eminentemente empírica.

El carácter arbitrario de las convenciones hace que éstas aparezcan como el producto de una historia determinada y que, cuando nos disponemos a dar cuenta de la emergencia, la evolución o el ocaso de una convención concreta no podamos separar

ésta de las condiciones históricas particulares en las que se dan estos procesos. Sobre esto volveremos más adelante, al discutir la noción de *convención normativa*, ya que en la teoría económica el único elemento que remite a este concepto de arbitrariedad es la representación de las convenciones a partir de problemas de coordinación, de lo que se deduce que cada convención establecida presupone, por definición, la existencia de al menos otra convención en potencia, una regularidad “alternativa a la convención”, en palabras de Lewis (1969: 68).

Hasta el momento, a través del recurso a los problemas de coordinación, hemos intentado argumentar en qué sentido una convención como regularidad de comportamiento es limitadamente arbitraria e indeterminada por definición. También, al menos indirectamente, hemos planteado por qué una convención, una vez establecida como regularidad de hecho en el comportamiento social, se refuerza a sí misma. Así, tanto en la Figura 1 como en la Figura 2, cualquier desviación unilateral de la convención conduce a una situación individualmente subóptima, por lo que podemos afirmar que la propia estructura de la situación ofrece incentivos suficientes para que no sea necesaria ninguna intervención externa que haga efectiva la conformidad con la convención³¹. Sin embargo, lo que todavía no ha quedado muy claro es cómo emergen las convenciones y cómo evolucionan hasta convertirse en regularidades de comportamiento aceptadas por la mayoría de la sociedad. ¿Pueden los modelos formales proporcionarnos un modo de explicar por qué entre todas las convenciones *en potencia* sólo una es la convención *establecida*? Como ya hemos avanzado, el intento de la teoría de juegos más estándar ha sido infructuoso en este sentido. Ha tenido que ser una alternativa teórica a ésta, la teoría de juegos evolutiva, la que ha permitido avanzar en la

³¹ A esto es a lo que se refirió anteriormente Hume cuando afirmaba aquello de que “experimentamos repetidamente los inconvenientes que resultan de transgredirla.”

comprensión de los mecanismos que hacen que una determinada regularidad de comportamiento se convierta en una convención. En concreto, como argumentaremos a continuación, la teoría de juegos evolutiva nos permite plantear sendas respuestas (seguramente sólo provisionales) a tres preguntas que se derivan de lo dicho hasta el momento, a saber: (1) ¿bajo qué condiciones surgen las convenciones sociales?, (2) ¿cómo evolucionan las convenciones sociales?, y (3) ¿por qué son estables las convenciones sociales?

2.5. Las convenciones como estrategias evolutivamente estables

Interpretar las convenciones a partir de la interacción de individuos perfectamente racionales nos ha llevado al problema teórico de encontrarnos con varias soluciones formales a una situación donde sólo existe una en la práctica. Una alternativa sería considerar las convenciones a partir de las limitaciones que tiene la teoría de juegos para abordar problemas *no triviales*. ¿Es posible avanzar hacia un modelo explicativo que sitúe a las convenciones más como una solución a un problema que como el problema mismo? Por otro lado, si no se afronta el problema de la emergencia y la evolución de las convenciones sociales, el planteamiento estaría muy cerca de un argumento funcionalista *ad hoc* que explicara cada convención simplemente como la mejor solución a cada problema de coordinación social concreto. Para no caer en dicha falacia explicativa, los teóricos de juegos han adoptado la perspectiva de la teoría de juegos evolutiva para modelar la emergencia de las convenciones sociales³².

³² Es obvio que no todas las convenciones sociales son el resultado de un largo proceso evolutivo. En multitud de ocasiones es una autoridad central la que determina cuál es la convención. Por ejemplo, después de la Revolución Francesa, los revolucionarios establecieron nuevas convenciones en relación con las unidades de peso y medida (el sistema métrico) e inventaron un nuevo calendario, con nuevos períodos de vacaciones y el cambio de la semana de siete días por una “década” de diez días (Chwe, 2001: 23). Sin embargo, una convención puede convertirse en una pauta de comportamiento social tan

El trabajo de Sugden (1986) fue el primero en utilizar herramientas importadas de la biología evolutiva con el objetivo de definir las convenciones como soluciones o equilibrios estables en problemas de coordinación³³. A continuación describiremos intuitivamente cómo la teoría de juegos evolutiva modela las convenciones sociales. Para ello, consideremos el siguiente ejemplo.

Cuando tomamos el autobús o el metro, entramos en un cine o un teatro donde los asientos no están numerados o intentamos comprar algo en una tienda abarrotada de clientes, generalmente respetamos una regla no escrita que dice que la prioridad sobre recursos escasos es asignada según el orden de llegada. Así, por ejemplo, si el autobús dispone de veinte asientos, sólo los veinte primeros viajeros que accedan al mismo tendrán derecho a sentarse. Para comprender cómo dicha regla no escrita, o *convención de comportamiento* en la terminología que empleamos en esta tesis, puede emerger en una situación determinada consideraremos la situación más simple, una situación en la que sólo existe un asiento y dos viajeros dispuestos a sentarse en el mismo. La estructura de dicho problema es similar a un problema de negociación como el representado en la Figura 3.

estable que hasta el gobierno más autoritario del mundo tendría dificultades en cambiarla. Al menos eso es lo que tuvo que pensar el gobierno comunista chino en plena revolución cultural, tal y como queda descrito en la narración de Kristof y Wudunn (1994: 70): “Durante un tiempo, se obligó a los coches a avanzar con las luces [de los semáforos] en rojo y a detenerse con las luces en verde, debido a que el rojo era un color revolucionario sinónimo de acción. Dicho plan fue retirado cuando, al no llegar el mensaje a un número suficiente de conductores, se produjeron accidentes múltiples en los principales cruces.”

³³ No es éste el lugar para realizar un repaso a las aplicaciones de la teoría de juegos a la biología, ni de cómo a partir de ésta un buen número de economistas han incluido las explicaciones de corte evolutivo en su arsenal teórico. Sobre esto segundo sí que se darán algunas pinceladas en las páginas siguientes. A propósito de la teoría de juegos evolutiva, ver el texto clásico de Maynard Smith (1982) o el más reciente de Weibull (1995).

Figura 3: Situación de negociación

		<i>Viajero A</i>	
		<i>Sentarse</i>	<i>Ceder el asiento</i>
<i>Viajero B</i>	<i>Sentarse</i>	-2, -2	2, 0
	<i>Ceder el asiento</i>	0, 2	1, 1

Ante tal situación, cada viajero puede elegir entre dos estrategias posibles: proceder a “sentarse” en el único asiento disponible o “ceder el asiento” a la otra persona. Si ambos muestran una clara intención de sentarse, el resultado de la interacción entre ambos será una disputa que, independientemente del resultado que produzca, supondrá un pago negativo para ambos jugadores. Si sólo uno reclama el asiento, y el otro cede, el primero recibirá la mayor utilidad posible en este juego, mientras que el segundo no recibirá utilidad alguna. Si ambos deciden no sentarse reciben una utilidad intermedia, inferior a la utilidad de ocupar el asiento, pero superior a la resultante de verse implicado en una disputa por el mismo. A continuación presentaremos brevemente cómo dicho juego puede ser interpretado en términos evolutivos.

Sea $UE(I, J)$ la utilidad esperada de un jugador de usar la estrategia I cuando el otro jugador utiliza la estrategia J . La estrategia I es una estrategia evolutivamente estable (EEE) cuando se cumplen las dos condiciones siguientes:

- $UE(I, I) \geq UE(J, I)$ para toda estrategia J
- o $UE(I, I) > UE(J, I)$, o bien $UE(I, J) \geq UE(J, J)$

La primera condición, o *condición de equilibrio*, se corresponde con el concepto de equilibrio de Nash utilizado por la teoría de juegos estándar. Este juego tiene tres equilibrios. Dos de ellos son equilibrios en estrategias puras, es decir situaciones en las que cada jugador emplea cierta estrategia con certeza. En el juego de la Figura 3 dichos equilibrios se corresponden con situaciones en las que uno de los jugadores elige la estrategia “sentarse” y el otro la estrategia “ceder el asiento”. Además, existe un tercer equilibrio, el denominado equilibrio en estrategias mixtas, en el que la estrategia “sentarse” es seleccionada con probabilidad $\frac{1}{3}$ ³⁴. Teóricamente, cualquiera de los tres equilibrios puede ser la solución real del juego.

La segunda condición, o *condición de estabilidad*, garantiza que una población que sigue la estrategia *I* no será satisfactoriamente invadida por otra población que emplee una estrategia distinta. Ésta es la condición genuinamente evolutiva e implica que una estrategia evolutivamente estable ha de ser la mejor respuesta a sí misma en relación con el conjunto de estrategias disponibles. La intuición que se halla detrás de esta condición es que una población alcanzará en su seno, a través de un largo proceso de prueba y error, una distribución de estrategias estable, en el sentido de resistente a la introducción de comportamientos diferentes a los ya presentes en dicha población. En otras palabras, si una distribución de estrategias cumple esta segunda condición, la

³⁴ El equilibrio de Nash en estrategias mixtas (ENEM) puede ser interpretado de dos modos distintos. En primer lugar, la teoría de juegos más estándar lo interpreta como una situación en la que cada jugador, individualmente, emplea una combinación de estrategias. Así, una persona que se encontrara de forma repetida con el juego de la Figura 3, elegiría “sentarse” en un tercio de las ocasiones y “ceder el asiento” en el resto. Una segunda interpretación, quizá más realista, propia de la teoría de juegos evolutiva, es interpretar el ENEM como una distribución de estrategias dentro de una población. En este caso deberíamos interpretar la situación de equilibrio como una en la que un tercio de la población elige “sentarse” y dos tercios “ceder el asiento”. Esta última interpretación es la que parece estar imponiéndose como la interpretación más aceptada en el seno de la teoría de juegos.

situación resultante estará preparada para no dejarse invadir por las posibles “mutaciones” en el comportamiento que se produzcan en pequeños grupos dentro de la misma³⁵.

La teoría de juegos evolutiva modela situaciones, como la representada en la Figura 3, en las que los individuos de una población interaccionan por parejas de forma aleatoria. En principio, estos individuos no saben qué estrategia emplear y, de este modo, comienzan utilizando una estrategia o una combinación probabilística de estrategias y van ajustando paulatinamente sus comportamientos mediante un proceso de aprendizaje. Los equivalentes de la selección natural cuando nos referimos al comportamiento social son la imitación y el aprendizaje. Esto tiene implicaciones importantes a la hora de pensar en el período de tiempo necesario para que una convención emerja o “mute”; cuando hablamos de “evolución social” o de “evolución cultural” no es necesario pensar en los términos de la “evolución biológica” o “genética”, sobre todo porque, según afirma la mayoría de biólogos evolutivos, la estructura genética del ser humano apenas se ha modificado desde las sociedades de cazadores y recolectores. “En la ‘evolución cultural’, la reproducción se produce a través de la formación de roles, de la instrucción, de la influencia social, de la imitación y del refuerzo delegado [en otras personas]” (Macy, 1997: 437).

La evolución de situaciones como la representada en la Figura 3 y la emergencia de convenciones de comportamiento en las mismas pueden ser caracterizadas de dos modos: a través de un proceso de evolución simétrica o de evolución asimétrica. En un

³⁵ No es extraño que uno de los campos donde se está intentando aplicar la teoría de juegos evolutiva sea la evolución y el cambio de las convenciones en ámbitos geográficos que reciben un volumen importante de inmigración. Sobre la coexistencia de diferentes convenciones en el corto y en el largo plazo, ver Bhaskar y Vega Redondo (2004).

proceso evolutivo simétrico la población converge hacia una situación en la que todos los individuos emplean la misma estrategia o combinación probabilística de estrategias. En nuestro ejemplo, el proceso convergería hacia la solución siguiente.

Sea p la probabilidad de encontrarse con un individuo que sigue la estrategia “sentarse”. De aquí se deduce:

$$UE(\textit{sentarse}) = p(-2) + (1 - p)2 = 2 - 4p \quad (2.1)$$

$$UE(\textit{ceder}) = p(0) + (1 - p)1 = 1 - p \quad (2.2)$$

Dadas las ecuaciones (2.1) y (2.2), cada individuo encuentra más beneficioso seleccionar la estrategia “sentarse” cuando $p < \frac{1}{3}$ y “ceder el asiento” cuando $p > \frac{1}{3}$.

Diremos que $p = \frac{1}{3}$ es EEE, ya que, al tratarse de uno de los equilibrios del juego cumple la primera condición y, además, es la única estrategia que cumple la segunda condición.

La solución propuesta no parece muy realista en ninguna de sus dos posibles interpretaciones, a saber: bien existe un tercio de la población que siempre reclama su derecho a sentarse y dos tercios que siempre ceden el asiento, o bien cada individuo reclama el asiento en un tercio de las ocasiones y lo cede dos tercios de las veces. En realidad, la situación es poco realista porque obvia una de las características más relevantes de la interacción social: la presencia de asimetrías. ¿Tiene sentido en la situación descrita tener presente las posibles asimetrías existentes?, y, ¿qué resultado depararía la evolución si consideráramos una situación asimétrica?

En el simple ejemplo introducido en la Figura 3 es posible introducir una asimetría que se observa en la realidad. Supongamos que nos encontramos en una población cargada de estereotipos de género y que como consecuencia, en el ejemplo que estamos utilizando, se pueden detectar dos comportamientos distintos, según el sexo de los individuos. Supongamos que el viajero A es una mujer, mientras que el viajero B es un hombre. Ahora nos encontramos ante un juego asimétrico, fruto de la introducción de un elemento ajeno a la información contenida en la matriz del juego: el tipo o rol de los participantes en el mismo. En un juego asimétrico, los jugadores aprenden a diferenciarse y a asignar a cada persona un rol diferente. A partir de dicho momento se produce un aprendizaje específico de rol. En este sentido, el enunciado de una estrategia en este tipo de juegos debe incluir la estrategia prescrita para cada rol, por ejemplo: “si eres jugador del tipo B juega X ; si eres jugador del tipo A juega Y .” Veamos a qué resultado conduce este aprendizaje específico de rol.

Sea p la probabilidad de que el jugador con rol B juegue “sentarse” y sea q la probabilidad de que el jugador con rol A juegue “sentarse”. Del mismo modo, sea $E(X / K)$ el pago esperado para un jugador con la característica K al elegir la estrategia X . De aquí se deduce que:

$$E(\text{sentarse} / B) = (-2)q + (1 - q)2 = 2 - 4q \quad (2.3)$$

$$E(\text{ceder} / B) = (0)q + (1 - q)1 = 1 - q \quad (2.4)$$

$$E(\text{sentarse} / A) = (-2)p + (1 - p)2 = 2 - 4p \quad (2.5)$$

$$E(\text{ceder} / A) = (0)p + (1 - p)1 = 1 - p \quad (2.6)$$

De las ecuaciones anteriores se deriva que de los tres equilibrios que tiene el juego de la Figura 3, sólo dos son estables en una situación de evolución asimétrica como la descrita. Los dos equilibrios en estrategias puras son estables debido a que la mejor respuesta ante un individuo que elige la estrategia “sentarse” es “ceder el asiento” y viceversa. Sin embargo, el ENEM, en un escenario asimétrico, viola la segunda de las condiciones que debe cumplir una estrategia evolutivamente estable, ya que podría ser invadido por cualquier otra combinación de estrategias. En otras palabras, cualquier otra combinación de estrategias es al menos tan buena como la prescrita por el ENEM.

El concepto de EEE a partir de un juego asimétrico parece acercarse más a la noción cotidiana de convención. Así, la convención “ceder el asiento a una mujer” cumple los requisitos formales de una estrategia evolutivamente estable. El problema es que la teoría no nos había prescrito sólo dicha estrategia, ya que también la convención “ceder el asiento a un hombre” sería evolutivamente estable según lo argumentado hasta el momento³⁶. Hasta aquí es hasta donde la teoría de juegos evolutiva puede ayudarnos a comprender el origen y la evolución de las convenciones: nos proporciona las condiciones mínimas que requiere una convención para sobrevivir a las fuerzas evolutivas. Sin embargo, no nos ofrece un argumento sólido acerca del porqué del establecimiento de una convención determinada a partir del conjunto de convenciones que serían evolutivamente estables.

Para explicar por qué nos coordinamos en una convención y no en otra, la teoría de juegos, y aquí también tenemos que incluir su versión evolutiva, se ve obligada a

³⁶ El hecho de que ambas regularidades sean evolutivamente estables nos recuerda de nuevo el carácter arbitrario e indeterminado de las convenciones. Del mismo modo que podríamos tachar de arbitraria la convención de ceder el asiento a una persona “de raza blanca” en los Estados Unidos de mitad del siglo XX.

recurrir a elementos ajenos a la teoría, tales como los conceptos de *prominencia*, o *punto focal*, como veíamos anteriormente. Por ejemplo, la convención “ceder el asiento a una mujer” podría haber evolucionado a partir de la *analogía* con otro tipo de situaciones sociales en las que también se asigna prioridad en el uso de recursos a las mujeres, como al atravesar una puerta o al servir los platos a los comensales en una cena³⁷. Pero también alguien podría haber comenzado a realizar tal práctica y haberse extendido a través de un proceso de “bola de nieve”. Por ejemplo, recordemos la práctica regular de los propietarios de tierra del sur de España de recoger a campesinos en la plaza del pueblo para que trabajaran en sus tierras a cambio de un jornal. Pues bien, si buscamos dicha regularidad en las mismas zonas rurales en la actualidad vemos que, si bien la práctica sigue existiendo, los grupos sociales que toman parte en ella han cambiado. Actualmente los propietarios ya no reclutan a campesinos autóctonos sin tierra, sino que reclutan a inmigrantes, sobre todo magrebíes. ¿Se ha producido una transmisión de conocimiento de los campesinos autóctonos del sur de España a los inmigrantes? Si es así, ¿cuál es la forma concreta que adopta dicha transmisión? La teoría de juegos evolutiva todavía nos aporta pocas soluciones a este tipo de cuestiones. Además, al menos por el momento, cualquier conjetura de este tipo se escapa de la propia teoría, por lo que ésta no nos aporta una explicación última de la regularidad que observamos. Por tanto, no es aventurado concluir que aunque la teoría de juegos nos ofrece las herramientas adecuadas para representar el escenario de condiciones necesario para la emergencia de las convenciones sociales, dar cuenta de éstas en toda su complejidad requerirá de *algo más*.

³⁷ El recurso explicativo a la presencia de analogías ha sido ampliamente empleado por Sugden (1986; 1989).

Ese *algo más* necesario para explicar por qué una convención concreta, de entre todas las convenciones posibles, acaba perpetuándose en el tiempo nos remite a la pregunta de por qué cada individuo se conforma con cada *convención de comportamiento* que observamos en la realidad social. En otras palabras, nos conduce a preguntarnos por el carácter normativo de las convenciones sociales. Este aspecto normativo, que será central cuando reconstruyamos el tipo ideal de *convención normativa* que presentaremos en el siguiente apartado, también ha recibido algunas consideraciones por parte de los teóricos de las *convenciones de comportamiento*. A continuación, repasaremos algunas de estas consideraciones.

2.6. El carácter normativo de las *convenciones de comportamiento*

Si bien todas las definiciones de lo que hemos denominado *convenciones de comportamiento* coinciden en las tres características mencionadas anteriormente – *regularidad de hecho*, *arbitrariedad* y *estabilidad* – no todas coinciden en la justificación última de la conformidad con la convención por parte de los participantes en la misma. Por ejemplo, mientras que en la definición de Lewis (1969: 47) citada anteriormente los agentes tienen una preferencia por la conformidad con la convención – eso sí, condicionada a la conformidad del resto –, Sugden (1986: 33) rechaza plenamente esta necesidad. Mientras que Lewis (1969: 14) afirma que su análisis de las convenciones se concentra en “situaciones en las que predomina la coincidencia de intereses”³⁸, la mayor parte de los ejemplos empleados por Sugden (1986; 1989)

³⁸ A continuación, explica desde un punto de vista técnico a qué se refiere con este predominio de la coincidencia de intereses, a saber: “las diferencias entre los pagos de los agentes en cualquier casilla [combinación de estrategias] son pequeñas comparadas con las diferencias entre los pagos de distintas casillas [distintas combinaciones de estrategias]” (Lewis, 196: 14).

remiten a situaciones mixtas donde se da tanto la coincidencia como el conflicto de intereses.

Aquí se puede ver cómo la herencia de Hume ha dado lugar a distintas interpretaciones. Si pensamos en el ejemplo humeano de cómo dos remeros coordinan sus movimientos para hacer que el barco avance, desde luego que encontraremos coincidencia de intereses. Del mismo modo, también la encontraremos en situaciones que responden al orden de preferencias de los juegos de las Figuras 1 y 2, tales como: multitud de convenciones en la carretera, la utilización de signos convencionales para coordinarnos, la definición de un punto convencional de encuentro, etc. Sin embargo, si pensamos en la discusión humeana acerca de la convención como una regla que resuelve el problema de la propiedad, está claro que aquí los intereses no sólo son diferentes, sino en multitud de ocasiones contrapuestos. Pero, ¿por qué es importante determinar la relación entre los intereses de los actores participantes en la convención? Pues porque de ella dependerá la fuente de legitimidad y la aceptación de la misma y, por tanto, el modo en el que los actores individuales conforman su comportamiento a la convención. Por ejemplo, si la convención establecida en un centro de trabajo es acudir cada mañana con un atuendo correcto, pero informal, uno no gana nada desviándose de la misma vistiendo bien con ropa deportiva o bien con esmoquin. Por el contrario, si la convención establecida en el mismo centro es que la antigüedad en el mismo determina la posibilidad de acceder a mejores condiciones de trabajo, uno, como joven titulado, podría estar interesado en que la convención fuera, por ejemplo, que los más cualificados accedieran a dichas mejores condiciones. En este último caso la conformidad con la convención es más problemática. Si se produce la misma más que de conformidad hablaríamos de conformismo, es decir, aceptar sin más una convención

que va contra nuestros propios intereses o, al menos, aceptarla a sabiendas de la existencia de una posible pauta de comportamiento alternativa que pudiera satisfacer mejor las preferencias propias³⁹.

Varios autores han ligado el problema del conformismo con la emergencia de las convenciones, en general, y con la generación de éstas a partir de problemas de coordinación social, en particular. Para Ullmann-Margalit (1977: 93), “un conformista es alguien que considera la mayoría de situaciones [...] como si fueran problemas de coordinación⁴⁰.” Esto es, el conformista en vez de elegir su opción preferida, elige aquella que considera que va a ser elegida por los demás. En este sentido, las convenciones suponen, en cierto modo, un ahorro considerable de costes cognitivos para los conformistas: les informan acerca de la acción que es probable que siga la mayoría de un grupo en un entorno social concreto.

No nos detendremos a profundizar más aquí sobre conformidad y conformismo, ya que ambos conceptos sólo tocan tangencialmente el tema central de esta tesis. De todos modos, volviendo a Ullmann-Margalit (1977: 96), esta aproximación al conformismo puede suponer el camino contrario “de un procedimiento bastante común en las ciencias sociales, donde muchas discusiones en apariencia a propósito de las normas sociales pasan directamente a una discusión sobre el conformismo.” Lejos de establecer un vínculo directo y necesario entre normas y convenciones, por una parte, y

³⁹ Con respecto al ejemplo de una convención establecida en la que chocaran los intereses de jóvenes y no tan jóvenes uno podría recurrir para su aceptación a argumentos de tipo intertemporal, como el siguiente: asumiendo que en $t+1$ dicha pauta me va a beneficiar, me conformo con ella en t . No diremos mucho aquí en contra de dicho argumento, aunque creemos que sólo supondría una justificación más para conformarme efectivamente con la misma.

⁴⁰ Aunque la metáfora es sugerente, la propia Ullmann-Margalit señala que el conformista no afronta un problema de coordinación como los que se han introducido anteriormente, sino lo que podría ser denominado como un *problema de coordinación unilateral*, es decir, sólo afrontado por aquél que conforma su comportamiento al de los demás, sin necesidad de que ocurra a la inversa.

conformismo, por otra, esperamos que esta breve discusión acerca de uno de los mecanismos que se pueden hallar tras el acto social del conformismo sirva para repensar éste desde el punto de vista de las acciones de los individuos.

Más allá del fenómeno del conformismo, la cuestión de la coincidencia o no de intereses en los problemas de coordinación que se encuentran en la base de la emergencia de las convenciones sociales es importante por un segundo motivo: el carácter normativo de las convenciones sociales.

Desde el comienzo de este capítulo, y a partir de la distinción entre el tipo ideal de convención predominante en la teoría económica y el propio de la teoría sociológica, hemos distinguido entre las *convenciones de comportamiento* – regularidades sociales arbitrarias y estables – y las *convenciones normativas* – de las que nos ocuparemos en el siguiente capítulo. Es por ello que el carácter normativo de las convenciones debería ser abordado más atendiendo al segundo tipo que al primero. Sin embargo, los teóricos económicos, y teóricos de la elección racional en general, tampoco han dejado de plantearse el *porqué* de la conformidad con las convenciones sociales. Incluso estos autores se sienten un tanto insatisfechos con la respuesta de que un actor individual sigue las convenciones sociales sólo porque el resto lo hace. A continuación presentaré algunas de las respuestas que se han dado al problema normativo desde esta primera definición de las convenciones sociales.

En primer lugar, cabe preguntarse por qué merece la pena recurrir a argumentos normativos para explicar comportamientos que son óptimos desde un punto de vista individual – dado el comportamiento esperado del resto de individuos – y, por tanto,

gozan de un refuerzo intrínseco – suponen equilibrios en el lenguaje de la teoría de juegos . Ullmann-Margalit (1977: cap. 3) ofrece tres argumentos de peso:

- (1) “Una norma es capaz de regular y canalizar las expectativas [...] de participantes anónimos.” Por ejemplo, supongamos que en nuestra empresa existe la siguiente norma de coordinación⁴¹: *ante la limitación de plazas de aparcamiento propiedad de la empresa, dichas plazas serán ocupadas por los empleados mediante un sistema de turnos*. Sólo la existencia de una norma que regule esta actividad garantiza que los nuevos trabajadores que se incorporen a la empresa puedan aprender el significado concreto de esta pauta de comportamiento.

- (2) “Mientras una regularidad extraída de eventos pasados podría ser continuada de más de una forma, una norma proporcionará el principio de continuidad que resolverá las posibles ambigüedades en la mayoría de situaciones futuras.” En el ejemplo anterior, la norma es necesaria para regular las acciones individuales de forma correcta, por ejemplo la norma podría indicar cuántas semanas en un período de tres meses uno puede utilizar una plaza.

- (3) “Existe un mayor nivel de articulación asociado con una norma que con una mera regularidad de comportamiento.” Debido al hecho de que una norma

⁴¹ Ullmann-Margalit (1977) utiliza el concepto de norma de coordinación para referirse tanto a “convenciones”, como a “decretos.” Por convención entiende una regularidad de comportamiento como la que hemos descrito en este apartado. Por decreto entiende la instauración de una norma por parte de un poder centralizado que resuelve un problema de coordinación; por ejemplo, los cambios introducidos por los revolucionarios franceses a los que nos referíamos en la nota 33. La distinción es interesante, aunque no abundaremos más en ella para no complicar más la terminología empleada en esta tesis. Salvo que se especifique de otro modo, cuando hablemos de normas de coordinación nos estaremos refiriendo a las propias convenciones.

puede ser transmitida y explicada, la presión grupal indicará qué solución (equilibrio) es prominente⁴².

Poco hay que objetar a la función de las convenciones como soluciones a problemas de coordinación. Sin embargo, si nos limitamos a describir el carácter funcional de las mismas no habremos avanzado mucho. En realidad, las convenciones pueden ser consideradas como un tipo especial de normas: normas de coordinación. Son especiales en el punto más problemático a la hora de abordar las normas sociales desde una perspectiva individualista: ¿por qué los individuos cumplen las normas? Las convenciones, en este sentido cumplirían una función más cognitiva que motivacional a la hora de coordinar las expectativas de múltiples actores. Sacconi y Moretti lo expresan del siguiente modo:

No nos interesa aquí la explicación de cómo las normas inducen puntos de equilibrio en los juegos debido a sus efectos estructurantes sobre los pagos y los incentivos. Los puntos de equilibrio son entendidos como posibles combinaciones de estrategias pre-existentes a una situación de juego dada. Las normas, por otra parte, deben ser principalmente vistas como un conjunto de sentencias normativas que prescriben una acción particular para cada jugador en cualquier situación del juego perteneciente a una clase de juegos dada. Por lo tanto, dichas sentencias simplemente piden a los jugadores que adopten un comportamiento de equilibrio dado dentro del conjunto de equilibrios existente de una situación de juego dada. Su efectividad no reside en el refuerzo proporcionado por algún conjunto de acciones a través de incentivos o sanciones externas, sino más bien, y sólo, en su funcionamiento como mecanismos cognitivos que inducen el sistema de expectativas apropiado sobre los comportamientos en equilibrio. Dado dicho sistema de expectativas apropiado, los jugadores tienen incentivos endógenos para adoptar el comportamiento prescrito, en tanto que esperan que los otros jugadores realicen la acción correspondiente. (Sacconi y Moretti, 2002: 1)

Así, de acuerdo con la visión extendida en la teoría económica, mientras que las normas que surgen como soluciones a dilemas sociales⁴³ – normas de dilema del

⁴² En este juego habría tantos equilibrios como número de combinaciones posibles para distribuir las plazas de parking limitadas entre todos los empleados en un espacio de tiempo concreto.

⁴³ Por dilema social se entiende una situación en la que todos los componentes de un grupo, o la mayor parte del mismo, actuando de acuerdo con sus intereses particulares, obtienen un resultado peor del que hubieran obtenido si hubieran ignorado sus propios intereses. Para una extensión de esta definición, ver

prisionero en la terminología de Ullmann-Margalit (1977) o normas esenciales en la terminología de Coleman (1990: 248) – siempre necesitan de alguna instancia sancionadora que garantice el cumplimiento de las mismas, las convenciones o normas convencionales sólo cumplen una función cognitiva: informar a los actores de algo que se ajusta a su propio interés. Las *convenciones de comportamiento*, en el sentido de la reconstrucción realizada en este apartado, no serían más que una *herramienta heurística*⁴⁴ de carácter social que permite a los individuos coordinar sus expectativas y comportamientos. Ésta es la conclusión que puede ser derivada del marco teórico empleado por los teóricos de juegos para representar las convenciones sociales. Como veremos en el siguiente apartado, justo éste será uno de los puntos donde los teóricos de lo que hemos denominado como *convenciones normativas* se mostrarán más críticos.

2.7. Resumen

En este apartado hemos intentado desgranar los elementos que componen una de las interpretaciones posibles de la noción de convención social, a la que hemos denominado *convención de comportamiento*. A partir de la definición ofrecida por David Hume, dicha interpretación ha sido la predominante entre los teóricos de la economía y los teóricos de la elección racional en general. El tipo ideal de la noción de *convención de comportamiento* queda definido por tres características:

Orbell y Dawes (1991: 38), Van Lange, Liebrand, Messick y Wilke, (1992: 3-4), Komorita (1994: 8), Kollock (1998: 183-184) o Miller (2004), entre otros.

⁴⁴ Regla heurística en el sentido empleado por los teóricos de la racionalidad acotada (*bounded rationality*). Como afirma Robles (2005), “Genéricamente los estudiosos de la racionalidad acotada han denominado heurísticos a los algoritmos (conjunto finito de operaciones que nos permiten resolver un problema) utilizados para resolver nuestras restricciones cognitivas.”

- (1) La *convención de comportamiento* es una *regularidad* empíricamente observable en el comportamiento humano. En otras palabras, una regularidad que ha de ser observada de forma recurrente en situaciones de interacción social.

- (2) La *convención de comportamiento* es una regularidad *indeterminada* y que tiene un *rango limitado de arbitrariedad*, en tanto que los participantes en la misma se mostrarían indiferentes entre dicha convención y otra pauta de comportamiento social alternativa si esta existiera.

- (3) La *convención de comportamiento* es una regularidad *estable*, ya que una vez establecida los sujetos individuales no tienen incentivos para desviarse unilateralmente de la misma. Dicha estabilidad también puede ser interpretada como el resultado de procesos evolutivos en el largo plazo.

Por último, nos hemos referido a dos cuestiones relacionadas que debe responder cualquier teórico que afronte el problema de la emergencia y evolución de las convenciones sociales: ¿por qué aceptamos y nos conformamos con las convenciones sociales?, y, ¿qué poder normativo tienen las convenciones sociales? En respuesta a ambas, la corriente teórica a la que nos hemos referido asume que sólo aquellas regularidades que emergen de situaciones donde predomina la coincidencia de intereses pueden ser calificadas como convenciones sociales. En dichas situaciones la convención desempeña el papel de un *heurístico social* que reduce los costes cognitivos propios de la toma de decisiones en situaciones de interacción social.

En el siguiente capítulo nos ocuparemos de la segunda de las interpretaciones de la noción de convención social, a la que denominamos *convención normativa*.

CAPÍTULO 3

LA TEORÍA SOCIOLÓGICA Y LAS *CONVENCIONES NORMATIVAS*

3.1. Introducción

En el año 1989, dos décadas después de la publicación de la influyente obra de Lewis (1969), se produjeron dos hechos importantes en la historia moderna del concepto de convención social. Por una parte, la filósofa Margaret Gilbert publicó su libro *On Social Facts*, cuyo capítulo sexto está dedicado a realizar una crítica de la propuesta de Lewis, en primer lugar, y a la reconstrucción de una noción de convención social anclada en los fundamentos metodológicos de Max Weber, en segundo. La referencia al concepto de convención en Weber se hace explícita en la primera página de este capítulo sexto (Gilbert, 1989: 315)⁴⁵. Por otra parte, en este mismo año, un grupo de sociólogos y economistas franceses, conocidos posteriormente como la Escuela Francesa de la Economía de las Convenciones, publican un número monográfico de la *Revue Economique* sobre “Economía de las Convenciones”. En éste se puede encontrar una introducción de carácter programático escrita por algunos de los autores más destacados de esta escuela, tales como: Jean-Pierre Dupuy, François Eymard-Duvernay, Oliver Favereau, André Orléan, Robert Salais y Laurent Thévenot (Dupuy *et al*, 1989)⁴⁶. Este grupo de pensadores franceses comparten con Gilbert la crítica del concepto estratégico de convención desarrollado por Lewis y la propuesta de un concepto de convención social más cercano al ya desarrollado por Weber (1964 [1922]).

⁴⁵ Gilbert también hace referencia a la noción de convención en Weber en artículos anteriores a *On Social Facts* (ver, por ejemplo, Gilbert, 1983: 242).

⁴⁶ Aunque los argumentos defendidos por diferentes autores dentro de esta corriente son variados, tres compilaciones publicadas recientemente pueden dar una imagen bastante completa de lo que esta escuela de pensamiento supone. Se trata de Batifoulier (2001), Favereau y Lazega (2002) y Orléan (2004).

La referencia explícita al concepto weberiano de convención no se reduce a estos autores. Por ejemplo, en un artículo reciente, Biggart y Beamish (2003) realizan un repaso a la noción sociológica de convención en el que también enlazan ésta con la propuesta de Weber. Como ya ocurría en el apartado anterior a propósito de Hume, la coincidencia de diversos autores en su reivindicación del concepto weberiano de convención social hace que sea importante detenernos en los principales elementos de la propuesta que realiza el sociólogo alemán, para posteriormente elaborar nuestro tipo ideal de *convención normativa*.

3.2. La propuesta original de Weber

El término *convención* aparece en el capítulo de “Conceptos sociológicos fundamentales” de *Economía y Sociedad*, específicamente en el punto 6, en el que Weber presenta la distinción entre *convención* y *derecho* como los dos tipos ideales de orden legítimo. Antes de adentrarnos en la definición weberiana de convención, así como en las implicaciones teóricas de este concepto, vamos a detenernos brevemente en el contexto teórico en el que Weber propone su noción de convención.

Weber introduce su concepto de convención en la discusión a propósito de la distinción entre la *costumbre* y el *derecho* presente en los filósofos del derecho y sociólogos alemanes de principios del siglo XX. Entre los primeros cita a Jhering (1916) y Weigelin (1919), mientras que la referencia sociológica es Tönnies (1909). No es éste el lugar para profundizar sobre la influencia positiva de cada uno de estos autores sobre la noción de convención presentada por el propio Weber. Por el contrario,

sí que le dedicaremos algunas líneas a la crítica que Weber realiza a la noción de convención propuesta en Stammler (1914). Como señalábamos en la nota 9, dicha crítica coincide con la crítica que determinados filósofos y sociólogos contemporáneos hacen a la noción de convención predominante en la teoría económica. Weber afirma:

Cuando Stammler pretende distinguir el derecho de la convención por el carácter enteramente “libre” de la sumisión en esta última, deja de estar de acuerdo con el uso corriente del lenguaje y ni siquiera es exacto en el caso de sus propios ejemplos. La observancia de la “convención” (en el sentido corriente de la palabra) - por ejemplo: del saludo en uso, del vestido conveniente, de los límites de forma y contenido en el trato humano - se *exige* muy seriamente al individuo como obligación o modelo y en modo alguno - como en la simple “costumbre” de preparar de cierta manera un plato culinario - se le deja a su libre elección. Una falta contra la convención (*costumbre estamental*) se sanciona con frecuencia con mucha más fuerza que la que pudiera alcanzar cualquier forma de coacción jurídica, por medio de las consecuencias eficaces y sensibles del boicot declarado por los demás miembros del propio estamento. (Weber 1964 [1922]: 27)

El contenido de este extracto no puede estar más lejos de la noción de convención social a la que hemos denominado *convención de comportamiento* y que ha sido reconstruida idealmente en el capítulo anterior. Primero, los individuos no siguen las convenciones libremente, sino que las cumplen como una obligación o como un modelo a seguir. Segundo, la trasgresión de la convención, al contrario de lo que ocurre con la costumbre, es penalizada por un “grupo de referencia” (estamento) del propio individuo. Llegados a este punto nos preguntamos: ¿cómo es posible que Weber introduzca un concepto de convención tan diametralmente distinto al de la tradición humana? El propio Weber nos ofrece una posible respuesta, en otra de sus críticas al trabajo de Stammler:

[En Stammler] se confunden la “regulación” *normativa* de una conducta mediante reglas, que de hecho son tratadas por una pluralidad de hombres como “debiendo valer”, con las regularidades efectivas de la conducta humana. Ambas cosas deben ser separadas con rigor conceptualmente. (Weber, 1964 [1922]: 264)

En otras palabras, que una cosa son las regularidades que observamos en el comportamiento humano, y otra, conceptualmente distinta, la “regulación” normativa de éstas, por ejemplo mediante una convención. Tal distinción, central para el argumento de esta tesis, puede también ser encontrada en los trabajos de algunos teóricos actuales. Por ejemplo, el filósofo alemán Hartmut Kliemt ha señalado con respecto a este asunto lo siguiente:

Las estructuras superficiales de las instituciones sociales se reflejan en la conciencia de los hombres. Debido a su fantasía y a su imaginación, los individuos humanos tienen la capacidad de reinterpretar como estándares las regularidades institucionales observables y aceptarlas como razones de su propia acción. Pueden desarrollar en ellos mismos – y así lo hacen por cierto continuamente – la disposición a seguir determinadas pautas del comportamiento regular, no sólo en virtud del cálculo individual o de intereses en el caso particular, sino *porque* aceptan las pautas correspondientes como razones *independientes* de su comportamiento. [...] Las instituciones surgidas son “bifrontes”. Poseen, por una parte, el carácter de regularidades de comportamiento determinadas por intereses, que surgen de las disposiciones de comportamiento. Por otra, estas regularidades son concebidas por los individuos también como estándares obligatorios o reglas y normas en sentido estricto. Desde el punto de vista terminológico, este segundo aspecto de las instituciones merece ser considerado por separado. (Kliemt, 1986: 169)

Este carácter “bifronte” de las instituciones – entre las cuales podemos distinguir a las convenciones sociales – es fundamental para entender la propuesta que aquí estamos realizando. Sobre todo para no confundir esas regularidades arbitrarias, indeterminadas y estables de comportamiento social, a las que los teóricos económicos llaman convenciones, con la fuente de obligatoriedad de las mismas.

Como un último ejemplo del argumento weberiano, nos referiremos a una cita de Giner (1997: 89) en la que, tal como hacía Weber con Stammler y Kliemt con los teóricos de la elección racional más estándar, nos recuerda las limitaciones de un enfoque puramente voluntarista para dar cuenta tanto de las convenciones, como de otras instituciones sociales. En este caso, Giner discute la propuesta de “individualismo institucionalista” de Agassi (1960). El fragmento concreto es el siguiente:

Agassi propone un “individualismo institucionalista”. Éste entendería la sociedad como un medio (componente objetivo) convencional (componente subjetivo) de coordinación de acciones individuales. Aunque Agassi da a la palabra “convencional” un sentido específico que se refiere al hecho de que los individuos optan por seguir (o no) ciertas pautas aceptadas de conducta, [...] su tratamiento se inclina en demasía hacia una visión muy permisivista de las estructuras sociales. Bien está que las “instituciones puedan explicarse como medios impersonales de coordinación” – lo son –, pero ya es más dudoso que ello sea así porque “sean como actitudes que son aceptadas convencionalmente o por acuerdo.” (Giner, 1997: 89)

Todas estas citas pueden servirnos para vincular la propuesta weberiana con ese *algo más* que necesitaban los teóricos económicos para explicar el porqué de la emergencia y el mantenimiento de una convención concreta. Pues bien, para ir concretando ese *algo más* diremos que el primer elemento de nuestro tipo ideal de *convención normativa*, derivado de la definición weberiana de convención social, es la *presencia de cierta noción de obligatoriedad que acompaña a cada convención y que no puede ser reducida a un conjunto de preferencias y expectativas acerca del comportamiento de los demás*. Hecha esta primera aproximación intuitiva a la noción weberiana de convención, nos preguntamos: ¿qué entiende exactamente Max Weber por convención social?

Para Weber (1964: 27), la convención es un tipo de orden; un orden cuya “validez está garantizada externamente por la probabilidad de que, dentro de un determinado círculo de hombres, una conducta discordante habrá de tropezar con una (relativa) *reprobación* general y prácticamente sensible.” Si las *convenciones de comportamiento*, en su versión evolutiva, se retroalimentaban a través de un incremento en la probabilidad de que una determinada regularidad se encuentre en una población concreta, las *convenciones normativas* se refuerzan a medida que aumenta la

probabilidad de la reprobación de una conducta contraria a la misma⁴⁷. Más adelante, Weber completa la definición aportada anteriormente:

Convención debe llamarse a la “costumbre” que, *dentro de un círculo de hombres*, se considera como *válida* y que está garantizada por la reprobación de la conducta discordante. En contraposición al derecho (en el sentido que usamos esta palabra) falta el cuadro de personas especialmente dedicado a imponer su cumplimiento. (Weber, 1964 [1922]: 27)

De estas definiciones extraemos el segundo y tercer elemento de la definición de convención social de Weber. No sólo sabemos que las convenciones van acompañadas de cierta noción de obligatoriedad, sino también que *esta obligatoriedad se deriva de la probabilidad de reprobación de la conducta discordante con la convención* y que *dicha reprobación no es realizada por un “cuadro de personas especializado”*. Pero si, al contrario que en el derecho, no existe ese cuerpo de personas especializado en “castigar” a aquéllos que se desvían de la norma legal, ¿quién ejerce ese papel de controlador en el caso de las convenciones? Con respecto a esto, Weber señala lo siguiente:

Para el concepto de "convención", según lo entendemos, no es suficiente que una acción de una clase determinada sea "aprobada" por muchos individuos y "desaprobada" la acción contraria, sino que se de la probabilidad de una actitud semejante en un "medio específico" del actor, con lo cual, naturalmente, no se quiere significar un medio espacial. Esto es: debe darse alguna nota que limite el círculo de personas correspondientes que forman este "medio", sea profesional, de parentesco, de vecindad, estamental, étnica, religiosa, política o de cualquier otra clase, y sea la pertenencia tan oscilante como se quiera. (Weber, 1964 [1922]: 262)

La reprobación de las conductas discordantes es una tarea que, en la convención, es realizada por un “medio específico del actor”, es decir, por un grupo de personas que, siendo también partícipes en la convención, penalizan a aquel individuo que se

⁴⁷ Ni que decir tiene que ambas formas de retroalimentación no son excluyentes, sino más bien complementarias.

desvía de ella. Por ejemplo, cuando un adolescente se comporta de forma contraria a las convenciones imperantes en su entorno – relativas al lenguaje, el atuendo, los gustos, etc. – es su propio “grupo de pares” el que pone en marcha los mecanismos sancionadores oportunos para disciplinar a quien no se conforma con estas convenciones. Estas sanciones pueden ir desde una mirada sutil de desaprobación al “ostracismo social” (el “boicot declarado”, en palabras de Weber).

Resumiendo lo anterior, la tres características fundamentales de la noción de convención en Weber son: (1) la presencia de cierta noción de obligatoriedad, (2) derivada de la aprobación o desaprobación por parte de los otros, (3) los cuales no suponen un grupo especializado de personas encargadas de administrar incentivos y sanciones, sino que pertenecen al “medio específico del actor.” Como última forma de delimitar el concepto weberiano de convención social, nos detendremos a continuación en la distinción entre *costumbre*, *derecho* y *convención*, como medio para comprender dónde debemos establecer los límites conceptuales de esta última. Detengámonos un momento en el siguiente fragmento:

Por oposición a la *convención* y al *derecho*, la *costumbre* aparece como una norma *no* garantizada exteriormente y a la que de hecho se atiene el actor “voluntariamente”, ya sea “sin reflexión alguna” o por “comodidad”, ya por otros fundamentos cualesquiera, y cuyo probable cumplimiento en virtud de tales motivos puede esperar de otros hombres pertenecientes al mismo círculo. La *costumbre*, en este sentido, carece de “validez”; por nadie está exigido que se la tenga en cuenta. Naturalmente, el tránsito a la *convención* válida y al *derecho* es absolutamente fluido. Por doquier, lo que de hecho se viene haciendo es padre de lo que luego pretende validez. (Weber, 1964 [1922]: 24)

Así, la *costumbre* no requiere de validez y la adhesión a la misma no puede considerarse sujeta a coerción alguna. Por el contrario, los mecanismos de aprobación y desaprobación, los incentivos y las sanciones sociales dotan a la *convención* y al *derecho* de validez. La diferencia entre estos dos últimos es la presencia de un cuerpo

especializado de personas que administran las sanciones en el caso del *derecho* y la ausencia de éste en el caso de la *convención*. Sobre la distinción entre *costumbre* y *convención*, Weber introduce el siguiente ejemplo:

Es costumbre hoy un determinado tipo de desayuno; pero jamás es obligatorio (excepto para los huéspedes de un hotel) y no siempre fue costumbre. Por el contrario, los modos de vestir, aunque nacieron como “costumbre”, son hoy, en gran medida, no sólo costumbre sino convención. (Weber, 1964 [1922]: 24)

Una vez que hemos delimitado el uso que Weber hace del concepto de convención, el resto de este apartado se dedicará a la reconstrucción ideal de la noción de *convención normativa*.

3.3. La recepción de la noción weberiana de convención

Como ha sido puesto de manifiesto en la introducción de este apartado, diversos autores y corrientes teóricas han ido más allá de la explicación de las convenciones sociales en los términos de la teoría de juegos, recurriendo a la noción weberiana de convención expuesta anteriormente.

Así, los teóricos de la escuela francesa de la *Economía de las Convenciones*, a los que ya nos referíamos, han señalado la necesidad de entender las convenciones sociales no sólo como reglas de comportamiento, sino también como modelos de evaluación que permiten la interpretación de dichas reglas. En este sentido, las convenciones cumplirían una doble función: por un lado, la de coordinar las acciones individuales – como hemos visto en el apartado anterior; por otro, la de coordinar los principios que sirven como marco interpretativo para la ejecución de las propias acciones. Así, emergen dos niveles en la definición de la convención, a saber: el nivel de los principios, compuesto por un

conjunto de principios normativos que otorga legitimidad a las acciones; y el nivel de las reglas convencionales, que responde al componente estratégico de las convenciones sociales⁴⁸. Este segundo nivel ya debe resultarnos conocido: la teoría de juegos evolutiva nos ofrece una forma intuitivamente plausible de aproximarnos al mismo. Sin embargo, hasta ahora hemos dicho poco acerca de los fundamentos teóricos del primero. De ello nos ocuparemos en este capítulo.

La autora que ha tratado de una forma más precisa el componente normativo de las convenciones sociales ha sido la filósofa Margaret Gilbert. Ésta, en una serie de artículos publicados a lo largo de la década de los ochenta, discute algunos de los aspectos más controvertidos de la noción de convención social como regularidad de comportamiento (especialmente la que procede de la obra de David Lewis discutida en el apartado anterior)⁴⁹. Dichos trabajos culminaron en la publicación en 1989 de su libro *On Social Facts*, cuyo capítulo sexto constituye la formulación más acabada de lo que hemos denominado *convención normativa*⁵⁰.

Para Gilbert (1989: 337), una convención social no supone necesariamente una regularidad empíricamente observable en el comportamiento humano, sino más bien “un principio de acción aceptado conjuntamente, un *fiat* de grupo con respecto a cómo debe actuar uno en determinadas situaciones.” De forma intuitiva, y antes de adentrarnos en las especificaciones técnicas de la propuesta de Gilbert, diremos que si la

⁴⁸ Batifoulier y Larquier (2001) han vinculado este primer y segundo nivel con las tradiciones weberiana y humeana, respectivamente, en el estudio de las convenciones sociales.

⁴⁹ Resulta sorprendente ver como el trabajo de Gilbert ha sido pasado por alto por la mayoría de los autores de la escuela francesa de la Economía de las Convenciones, a pesar de que ambos parten de una crítica similar al concepto de convención introducido por Lewis (1969). Mediante una comunicación personal, el sociólogo danés Soren Sommer Jagd, estudioso de esta escuela francesa nos indicaba que, hasta donde él conoce, sólo Orléan (2004) establece un vínculo explícito entre el trabajo de Gilbert y la citada escuela.

⁵⁰ En aras de la claridad expositiva, nuestro argumento en este punto seguirá, en general, el de Gilbert, sólo recurriendo a otras referencias de forma secundaria.

convención de comportamiento recomendaba seguir determinada pauta de comportamiento debido a que constituía la práctica habitual en una situación social dada, la *convención normativa* recomienda seguir determinada pauta de comportamiento debido a que es lo que *está prescrito* hacer en una situación social concreta. Eso sí, no hay que confundir el *deber* que implica la convención con el *deber* intrínseco propio de los principios morales. De esto nos ocuparemos más adelante.

Del mismo modo que ya hacíamos en el apartado anterior, dedicaremos las próximas páginas a reconstruir idealmente los principales elementos de la noción de *convención normativa*. Paralelamente, discutiremos por qué la definición de convención aportada por la teoría de juegos necesita de esta segunda consideración normativa para dar cuenta de toda la complejidad de las convenciones sociales.

3.4. La convención como principio

Dedicaremos este subapartado a delimitar el significado que tiene el término *principio* para Margaret Gilbert. La noción de principio empleada por Gilbert puede ser formulada del siguiente modo: “dondequiera que un miembro de [la población] *P* esté en la circunstancia *C*, él o ella va a realizar la acción *A*” (Gilbert, 1989: 373). De este modo, un principio actúa como un simple *fiat*, es decir, como lo que se supone que uno va a hacer en una situación social determinada. Según Gilbert, este principio, o *fiat*, es aceptado por una población de forma tácita y conjunta.

La aceptación de un principio normativo, o *fiat*, es tácita en tanto que no es necesario hacer explícito el principio en cuestión para que éste sea aceptado por la

población que participa del mismo. De este modo, el principio *P* es de conocimiento común entre dicha población; en otras palabras, todo el mundo sabe que *P* es el principio que rige en la situación *S*, todo el mundo sabe que todo el mundo sabe que *P* es el principio que rige en la situación *S*, y así sucesivamente. Por tanto, aunque todo el mundo conozca el principio que regula determinada acción, su aceptación y el seguimiento de sus prescripciones pueden tener lugar sin que los miembros de la población en cuestión se lo comuniquen los unos a los otros. Es importante que tengamos clara la diferencia entre el supuesto de conocimiento común invocado por Gilbert (1989) y el utilizado por Lewis (1969). La diferencia fundamental entre ambos es la presencia del lenguaje en el caso de Gilbert (1989). Así, no es necesario que los participantes en la convención tengan *conocimiento común* sobre un *precedente* que determina cuál es la convención establecida; lo importante es que sean capaces de reconocer el *principio* – conjunto de sentencias normativas que indican cuál es la acción que se va a emprender – que rige en una situación social determinada. Por tanto, en este caso, y al contrario de lo que ocurría en la tradición humeana, la convención no puede ser anterior al lenguaje; no puede ser anterior al acto verbal de reconocer, aunque sólo sea interiormente, el principio al que nos referimos.

Además, el principio es aceptado de forma conjunta en el sentido que queda expresado en la siguiente oración: “nosotros aceptamos el principio *P*, por tanto, en tanto que yo soy uno de nosotros, debo conformarme con el principio *P*.” Este segundo elemento supone una de las características definitorias de la noción de convención como principio formulada por Gilbert: la convención implica necesariamente un “nosotros”, es decir, la convención siempre hace referencia a una colectividad, y no simplemente a

una población. El concepto de colectividad está relacionado con lo que Gilbert denomina “sujeto plural”, el cual queda definido del siguiente modo:

Recordar que para ser miembro de un sujeto plural, en el sentido que yo le doy, uno debe participar en una acción con algunos otros, donde ‘acción’ incluye tanto acciones en sentido estricto, como creencias, actitudes y otros atributos. Al explicar qué es participar en una acción estrictamente hablando, argumento que cada persona debe manifestar abiertamente un compromiso condicional de su voluntad en relación con un determinado fin, tal que si los otros hacen lo mismo, sus voluntades están comprometidas conjuntamente con tal fin. De este modo, constituyen el sujeto plural de este fin, y están dispuestos a participar en una acción, la persecución de su objetivo conjunto. (Gilbert, 1989: 357)

La noción de *sujeto plural* supone una de las aportaciones fundamentales de la filosofía de Margaret Gilbert, constituyendo el germen de sus discusiones posteriores en torno al concepto de *preferencia colectiva* (Gilbert, 2001). Sin embargo, para hacer evidente la importancia de las “colectividades” en la formulación de la noción de convención no es necesario recurrir al concepto de “sujeto plural” de Gilbert. Recordemos la referencia que hacía Weber en este sentido al “medio específico del actor” como el encargado de aprobar o desaprobado las conductas del propio actor. Dicho esto, volvamos a la noción de principio introducida por Gilbert.

Para evidenciar cuál es el mecanismo que se halla detrás de la noción de principio empleada por Gilbert, pensemos por un momento en un ejemplo que ya he utilizado anteriormente: la convención existente en España de vestir de negro en un funeral. Dicha convención podría ser formulada en los términos empleados por Gilbert del siguiente modo: “dondequiera que un español⁵¹ se encuentre en un funeral, vestirá de negro.” Además, en tanto que en España lo aceptado es vestir de negro en un funeral, yo, como español, me conformo con vestir de negro en tal circunstancia. Aunque esto

⁵¹ Utilizamos el calificativo “español” simplemente para hacer explícito el ámbito de referencia para la acción, sin perjuicio de que pudiera utilizarse cualquier otro ámbito donde exista tal convención.

último puede resultar intuitivo, alguien podría preguntarse cuál es la base para dicha conformidad.

Desde un punto de vista lógico, si vestir de negro en un funeral es mi principio, entonces, desde mi propio punto de vista, tengo una razón para conformarme con el mismo. Ahora bien, esto no quiere decir que, desde el punto de vista de los juicios de valor, yo considere que vestir de negro en un funeral goza de alguna bondad intrínseca. Tanto es así que, aunque pueden existir razones morales para conformarse con una convención, como ya avanzábamos anteriormente, los deberes de la convención no son por sí mismo deberes morales. En palabras de Gilbert (1989: 394), “las convenciones generan razones para actuar por su propia fuerza [...], sin apelar a consideraciones de valor intrínseco⁵².” Por ahora nos conformamos con dejar abierta la puerta a la posible coexistencia entre principios morales y convenciones⁵³.

Recapitulando, la noción de convención de Gilbert está apoyada fundamentalmente en la existencia de un principio en el seno de una colectividad que indica cuál es el comportamiento apropiado en una situación social determinada. Así, en tanto que miembro de dicha colectividad, yo sigo tal principio. Esta caracterización permite, además, la existencia de cierta distancia personal con respecto al principio que uno sigue en la práctica. Incluso no hay que descartar el posible conflicto entre el principio seguido en el seno de la colectividad a la que uno pertenece y su valoración personal de dicho principio. Por ejemplo, uno puede considerar que la excesiva

⁵² Cuando Gilbert emplea aquí el concepto de *valor intrínseco* no está sino haciendo referencia a la noción de valor (*value*) omnipresente en la teoría sociológica.

⁵³ Por ejemplo, uno puede seguir la convención “vestir de negro en un funeral” tanto si considera que dicha convención es buena intrínsecamente, como si piensa todo lo contrario. En el primero de los casos tiene una motivación intrínseca para seguir la convención; en el segundo, puede seguirla simplemente por imitación, anticipando la desaprobación de las personas que le rodean o por cualquier otro motivo.

formalidad del lenguaje académico es contraproducente para el propio avance de la ciencia y, sin embargo, aceptar, como miembro de una institución científica, el conjunto de principios que regulan dicho lenguaje.

Hasta el momento sabemos que el principio al que hace referencia Gilbert es un principio tácito, referido a una colectividad y que no implica necesariamente bondad intrínseca alguna. Sin embargo, aún no hemos dicho nada acerca de la fuerza vinculante de dicha noción de convención. Para entender por qué un principio, tal y como ha sido formulado, tiene un carácter normativo recurriremos a la analogía entre principio y acuerdo utilizada por Gilbert.

3.5. Convenciones y acuerdos: el componente normativo de las *convenciones normativas*

¿Qué implicaciones normativas tienen las convenciones sociales? En el apartado anterior, en el que reconstruíamos la noción de *convención de comportamiento*, concluíamos que lo único que mantenía las convenciones sociales era su carácter individualmente óptimo, su condición de equilibrio. Sin embargo, al tratarse de regularidades arbitrarias e indeterminadas – en el sentido de que siempre existe al menos una regularidad de comportamiento alternativa – necesitábamos algo más para justificar por qué los actores se conformaban con una regularidad y no con otra. Para responder a esta pregunta introducíamos lo que puede considerarse como una “función cognitiva” de las convenciones sociales. Ésta, lejos de suponer una transformación en la función de utilidad de los actores, sólo servía para indicar cuál de las regularidades individualmente óptimas era seguida de forma efectiva por los demás actores,

contribuyendo de este modo a la solución del problema de coordinación a partir del cual la propia convención ha surgido. Esta función cognitiva es totalmente congruente con la representación de las convenciones como el resultado de la “evolución social”. Así, los actores, a través de múltiples procesos de prueba y error, aprenden a coordinarse en aquella regularidad de comportamiento que acaba imponiéndose como la convención establecida en un medio social determinado.

El problema de esta primera aproximación al carácter normativo de las convenciones es que Gilbert (1989: 342) muestra convincentemente que “las convenciones sociales, concebidas intuitivamente, pueden cubrir situaciones de interacción que no suponen problemas [puros]⁵⁴ de coordinación.” En este caso, la relación entre problemas puros de coordinación y convenciones sociales es más una cuestión empírica que teórica. El problema que plantea este hecho es que “si las convenciones sociales no tienen por qué cubrir lógicamente los problemas [puros] de coordinación, su necesario aspecto normativo puede ser difícilmente expuesto en los términos de dicha estructura subyacente” (Gilbert, 1989: 355).

Como hemos visto, para Weber, cuando uno se conforma con una convención no lo hace porque ésta suponga la regularidad de hecho en una situación de interacción social, ni porque esté en su interés hacerlo – no es una “situación de intereses⁵⁵” – sino principalmente porque su medio social aprueba dicha conducta y desaprueba la

⁵⁴ De nuevo tenemos que recalcar que aunque Gilbert se refiere genéricamente a “problemas de coordinación”, se está refiriendo a “problemas puros de coordinación” como los empleados por Lewis y que se corresponden con los juegos presentados en las figuras 1 y 2 del apartado anterior. Esta diferencia es importante, ya que en los siguientes capítulos de la tesis seguiremos planteado las convenciones como soluciones a problemas de coordinación asimétrica, donde se mezclan tanto componentes de coordinación, como de conflicto.

⁵⁵ De acuerdo con Weber (1964[1922]: 23) un “uso está determinado por una *situación de intereses* (“condicionado por el interés”), cuando y en la medida que la existencia empírica de su probabilidad descansa *únicamente* en el hecho de que los individuos orienten racionalmente su acción con arreglo a fines por *expectativas* similares.”

contraria⁵⁶. Este hecho separa radicalmente la versión económica de las convenciones – ligada a los problemas puros de coordinación – de la versión sociológica – construida sobre mecanismos de aprobación y desaprobación social. Además, la confusión entre ambas nociones de convención ha llevado a algunos equívocos que deberíamos evitar. Por ejemplo, el economista Robert Sugden (1989: 95) afirma que “El mecanismo que puede transformar las convenciones en normas es el deseo humano de ser aprobado por otros.” Si bien, nada habría que objetar desde un punto de vista sociológico a la importancia de la “aprobación de los otros” en la emergencia y el mantenimiento de las regularidades de hecho en el comportamiento humano, no es necesario abandonar el concepto de convención por el de norma para captar este hecho – recordemos la definición de convención social de Weber. La noción de *convención normativa* que estamos presentando no necesita llamarse de otra forma para reconocer un hecho evidente: la presencia de sutiles y no tan sutiles presiones sociales cruciales para el mantenimiento de las convenciones. Llegados a este punto, un economista podría preguntarnos qué es exactamente lo que distingue el deber de la convención del deber de la norma moral. Un primer paso para responder dicha pregunta es aclarar qué entendemos por norma, qué tipos de normas podemos encontrar y la línea que separa a éstas de las convenciones sociales.

La distinción entre convención social y norma moral se asienta en el fundamento en la *creencia en valores* de esta última, como el propio Weber nos aclara en el siguiente fragmento:

⁵⁶ En realidad la aprobación o desaprobación del resto no es tan importante cuando nos referimos a problemas puros de coordinación, ya que la solución de éstos, en el lenguaje weberiano, coincide con una “situación de intereses.”

La sociología no tiene problema alguno respecto de las relaciones entre derechos, convención y “ética”. La norma moral se impone a la conducta humana por una determinada *creencia* en valores, pretendiendo aquella conducta el predicado “moralmente buena”, de igual manera que pretende el predicado de lo “bello” la que se mide por patrones estéticos. [...] Toda moral con “validez” efectiva – en el sentido de la sociología – suele estar garantizada ampliamente en forma convencional, o sea por la *probabilidad* de una reprobación de la trasgresión. Por otra parte, no todos los órdenes garantizados convencional o jurídicamente pretenden (o por lo menos: no necesariamente) el carácter de normas morales; [...] Respecto a *si* una determinada representación normativa, dentro de un círculo de hombres, pertenece o no al dominio “moral” (en tal caso “simple” convención o “puro” derecho), es cosa que la *sociología* sólo puede decidir con arreglo a aquel concepto de lo “moral” que haya valido o valga en el círculo de hombres en cuestión. Por eso no caben afirmaciones de carácter más general *sobre este* problema. (Weber, 1964: 29)

Sobre la afirmación de que todo orden moral suele estar sustentado por un orden convencional no diremos nada aquí. Sin embargo, la segunda afirmación por la que tanto la convención como el derecho no responden necesariamente a cuestiones morales es central para nuestro argumento, ya que establece una distinción clara entre convención y norma moral. Pero, volvamos al ejemplo del atuendo propio de los funerales para aclarar este hecho. Nadie pondría en duda que la *convención normativa* en España es vestir de negro en un funeral, aunque cada vez más personas se comportan de forma distinta. También parecen claros los mecanismos de desaprobación que se pueden desencadenar en determinados entornos (si alguien no está convencido que pruebe a vestir de rojo en el próximo funeral al que asista). Sin embargo, no podríamos ser tan taxativos si lo que queremos poner de manifiesto son las motivaciones que cada sujeto individual tiene para conformarse con dicho comportamiento. Así, una persona con fuertes creencias religiosas puede “vestir de negro” porque considera que su religión lo prescribe así. Otra persona puede considerar que “no vestir de negro” supone una falta de respeto a los participantes en el acto, por lo que considera que dicho acto “está mal”, añadiéndole de este modo una consideración moral. La motivación de otra persona puede ser simplemente evitar la desaprobación del resto⁵⁷. Por tanto, las

⁵⁷ Algunos autores incluyen también el “refuerzo de la identidad” como una de las motivaciones para seguir la convención. Refiriéndose a los teóricos de la identidad, Macy afirma: “El cumplimiento con las

motivaciones que se hallan detrás de la conformidad con una convención no pueden ser reducidas a motivaciones morales. Las convenciones no son normas morales.

Más complicada puede ser la distinción entre las convenciones sociales – del modo en el que Weber las conceptualiza – y las normas sociales, en el sentido en que este concepto ha sido utilizado por autores como Elster (1989), Coleman (1990), Hechter y Opp (2001) o Therborn (2002). Dado que ambos conceptos suponen una prescripción acerca de lo que tenemos que hacer, resulta crucial determinar cuál es la fuente de deber propia de las convenciones y que las distingue de las normas sociales. Empleando la terminología de Coleman (1990: 266), lo que se tendría que demostrar es que la “realización efectiva” de las convenciones es diferente a la de las normas⁵⁸. Pues bien, aunque este debate terminológico está lejos de quedar cerrado, y no es nuestro objetivo presente hacerlo, para evitar confusiones nominalistas innecesarias nuestra propuesta se ciñe completamente a la propuesta original de Weber. Éste, como hemos visto, distingue entre convención, derecho y norma moral; sin embargo, no contempla explícitamente lo que con posterioridad ha sido denominado norma social. No lo hace porque para Weber el concepto de convención social incluye todas aquellas normas sociales que no caen necesariamente en el terreno de la moral. Por tanto, en esta tesis, contemplamos que toda norma social que no sea moral por necesidad será calificada

convenciones no está motivado por el cálculo instrumental del riesgo de ser castigado sino por una aportación emocional a la identidad que estas convenciones dibujan” (Macy, 1997: 431)

⁵⁸ Recordemos que se puede establecer una diferencia entre ambas respecto a lo que Coleman (1990: 241) denomina “condición de demanda” de las normas. Mientras que las normas sociales emergen a partir de situaciones denominadas dilemas sociales – donde puede ser identificado algún tipo de bien colectivo –, las convenciones emergen a partir de situaciones asimilables a problemas de coordinación. Estos problemas de coordinación no se reducen a los problemas puros de coordinación planteados por Lewis (1969), sino que en ellos se pueden identificar distintos niveles de coincidencia y conflicto de intereses. Esta distinción entre norma social y convención social de acuerdo a la estructura de interacción que se halla en la base de la emergencia de cada una ha sido adoptada plenamente por la teoría económica, donde las normas sociales resuelven problemas de cooperación, mientras que las convenciones sociales hacen lo propio con los problemas de coordinación. Sobre las diferencias analíticas entre considerar problemas de cooperación y problemas de coordinación, ver Miller (2007).

como convención social. Aclarado esto, volvamos a la pregunta acerca de la fuente de obligatoriedad de las convenciones sociales.

Para aclarar en qué sentido las *convenciones normativas* implican obligatoriedad, recurriremos de nuevo a la propuesta de Gilbert (1989).

Para Gilbert (1989: 367), aunque no son lo mismo, “existe una analogía llamativa entre el ‘deber’ de una convención social [...] y el ‘deber’ de un acuerdo explícito.” Como ya hemos argumentado, en el contexto de las convenciones la noción de deber implica que algo es lo que se supone que uno hace en una determinada situación. Es lo que Gilbert (1989: 369) denomina un “*casi-acuerdo* para participar en una práctica en cuestión.” Por tanto, una primera forma de aproximar el carácter normativo de las convenciones sería caracterizar a éstas como “normas de *casi-acuerdos*.” Como señala Gilbert (1989: 369-370), se pueden detectar al menos cinco virtudes en el hecho de considerar las convenciones sociales como “normas de *casi-acuerdos*”. En primer lugar, dicha caracterización “respeta una intuición básica sobre las convenciones: las partes de una convención deben creer que deben actuar conforme a la convención.” En segundo lugar, “respeta el hecho de que las partes de una convención juzgarán que deben actuar de determinada manera, pudiendo ser sus juicios de deber caracterizados negativamente en el modo en el que han sido especificados.” En tercer lugar, “respeta la observación de que las partes de una convención consideran que estos juicios de deber tenían algún fundamento positivo.” En cuarto lugar, “esta explicación representa un posible desarrollo de la observación de que existe una analogía bastante fuerte entre los acuerdos y las convenciones sociales.” En quinto lugar, “respeta la idea de que las convenciones sociales pueden ser tácitas.”

Una vez que han sido presentadas las virtudes de esta aproximación, ¿cómo define exactamente Gilbert las convenciones como normas de casi-acuerdos? Lo hace del siguiente modo:

A grandes rasgos, existe una convención social en un grupo, de acuerdo con este punto de vista, cuando y sólo cuando es de conocimiento común en el grupo que la mayoría de la gente piensa que cualquier miembro del grupo debe hacer esto y lo otro en un contexto determinado, y que esto es así porque es de conocimiento común en el grupo que la mayoría de los miembros creen que los miembros del grupo deben hacer esto. (Gilbert, 1989: 370)

A pesar de que esta definición proporciona una aproximación bastante plausible al carácter normativo de las convenciones, Gilbert no cree que pueda ser considerada como definitiva, debido a que no garantiza la existencia de una colectividad. Así, esta aproximación viola la característica más importante de la noción de convención propuesta por Gilbert: toda convención debe estar referida a una colectividad. En este sentido, la autora afirma que “las convenciones sociales y los fenómenos relacionados están necesariamente anclados en los sujetos plurales” (Gilbert, 1989: 362).

Aunque la caracterización de las convenciones como normas de casi-acuerdos no sea definitiva, la analogía entre convención y acuerdo puede continuar arrojando luz sobre los mecanismos que se hallan detrás de la noción de convención como principio. Así, “un acuerdo puede ser propiamente visto como una decisión conjunta sobre un determinado curso de acción” (Gilbert, 1989: 380), lo que implica que un acuerdo no puede deshacerse unilateralmente. Por ejemplo, si acuerdo con mi compañero de piso limpiar el piso cada sábado y un determinado sábado decido no hacerlo, de ningún modo puedo aducir que el acuerdo es ahora otro, por lo menos no hasta que ambos acordemos que los términos que regulan nuestro acuerdo han sido cambiados

conjuntamente. En un acuerdo “cada individuo tiene al menos una razón *prima facie* para conformarse” (Gilbert, 1989: 381). Esta razón no tiene que ver con que el curso de acción sea el que maximiza la utilidad individual, ni con que conformarse goce de un valor intrínseco, sino con el mero hecho de que se ha acordado hacerlo.

Gilbert alude a tres similitudes entre los conceptos de convención y acuerdo. En primer lugar, ambos fenómenos implican una pluralidad de sujetos. En segundo lugar, ambos implican obligaciones del mismo tipo. En tercer lugar, se puede establecer una analogía entre un acuerdo como una decisión conjunta y una convención como un principio aceptado conjuntamente. En este sentido concluye que “las convenciones y los acuerdos son fenómenos distintos, aunque análogos” (Gilbert, 1989: 382).

Como se deriva de todo lo anterior, la noción de convención social propuesta por Gilbert tiene que ver con la aceptación conjunta de un principio que prescribe cómo actuar en una situación determinada; un principio que comparte la mayoría de características de un acuerdo. Quizá la diferencia más prominente entre convención como principio y acuerdo es el carácter tácito de la primera y el carácter explícito que generalmente tiene el segundo⁵⁹.

Aunque se puede deducir implícitamente de la propuesta de Gilbert que la aceptación tácita de un principio de acción conjunta deriva obligaciones, esta autora no es muy explícita al respecto. De todos modos, esto no nos crea mayor problema. Si aceptamos seguir un principio que prescribe un comportamiento concreto para los

⁵⁹ Latsis (2005: 718) critica el hecho de que Gilbert no logra establecer una diferencia clara entre las nociones de convención y de acuerdo. Sin entrar más en detalle, en esta tesis, al referirnos a la noción de convención como principio, nos referiremos a un tipo de principio tácito y no explícito, lo que nos aleja de la noción cotidiana de acuerdo.

miembros de un grupo, es evidente que cuando un miembro del grupo se desvía del mismo, los demás miembros desaprueben su conducta, intentando que el comportamiento del primero vuelva a ajustarse al principio en cuestión. Lo que más nos interesa de la propuesta de Gilbert, y lo que la aleja ciertamente de lo que hemos denominado *convención de comportamiento*, es que la existencia de un principio garantiza que los miembros del grupo puedan verbalizarlo, puedan transmitirlo y aprenderlo y puedan pedir y rendir cuentas a propósito del cumplimiento del mismo. El siguiente texto de Biggart y Beamish, muestra cómo la noción sociológica de convención manejada por estos autores está muy cerca de la formulación de la convención como principio normativo llevada a cabo por Gilbert:

Las convenciones - y conceptos relacionados como los hábitos, las costumbres, las rutinas y las prácticas estandarizadas - son nociones, a menudo tácitas pero también conscientes, que organizan y coordinan la acción de un modo predecible. Las convenciones son guías acordadas, aunque flexibles, para la interpretación y la interacción económica. Aunque son utilizadas por los individuos cuando compran, negocian y venden, las convenciones no residen en, y no son reducibles a, los individuos. Los teóricos de las convenciones explican el orden económico como un producto de actores socialmente informados que operan dentro de nociones colectivas acerca de lo que es posible, probable y plausible [...]. Las convenciones son plantillas compartidas para interpretar situaciones y diseñar cursos de acción de forma mutuamente comprensible que implican responsabilidad social, esto es, que proporcionan una base para el juicio acerca de lo apropiado que son los actos propios y ajenos. Así, las convenciones son medios de coordinación económica entre los actores que son esencialmente colectivos, sociales e incluso morales en su naturaleza. (Biggart y Beamish, 2003: 444)

Por tanto, si los juicios que implican la presencia de normas morales responden a la dicotomía entre “lo bueno” y “lo malo”, los juicios que recurren a las convenciones responden a la dicotomía entre lo que es “apropiado” y lo que no lo es. Y, en este sentido, algo es “apropiado” si se ajusta al comportamiento prescrito por una serie de principios aceptados tácitamente por una sociedad o, más concretamente, por los grupos sociales de los que se compone una sociedad.

3.6. Resumen

El objetivo de este capítulo ha sido reconstruir el tipo ideal de *convención normativa* propio de la teoría sociológica. Al contrario que al reconstruir la noción de *convención de comportamiento*, cuando hablamos de *convención normativa* nos referimos a un principio normativo que prescribe una acción concreta en una situación de interacción social determinada. Este principio suele ir acompañado en la realidad por una regularidad de comportamiento, aunque la coexistencia entre ambos fenómenos no tiene por qué darse siempre. Las características de este principio al que denominamos *convención normativa* son las siguientes:

- (1) Se trata de un *principio de acción aceptado de forma tácita y conjunta por una colectividad*.
- (2) La *fuerza de obligatoriedad* del mismo radica en la existencia de *mecanismos de aprobación* de los comportamientos que se ajustan al principio y de *mecanismos de desaprobación* de los comportamientos contrarios a éste.
- (3) La motivación que tienen los individuos para aprobar o desaprobar una acción de acuerdo con la *convención normativa* existente está relacionada con la equiparación de la convención con un acuerdo tácito y, por tanto, la desviación de la convención desencadena una actitud similar a la resultante de la trasgresión de un acuerdo.

- (4) Mientras que las normas morales definen las conductas como “buenas” o “malas”, las *convenciones normativas* definen las acciones como “apropiadas” o “inapropiadas”.

Una vez que hemos reconstruido idealmente las nociones de *convención de comportamiento* y *convención normativa*, en el último capítulo de esta parte, avanzaremos hacia nuestra propia definición de convención social a partir de elementos tomados de las dos nociones de convención reconstruidas anteriormente.

CAPÍTULO 4

HACIA UNA DEFINICIÓN INTEGRADA DE LA NOCIÓN DE CONVENCION SOCIAL

4.1. Introducción

El objetivo principal de esta parte era desgranar los elementos necesarios para elaborar una definición operativa de convención social que nos fuera útil en el resto de la tesis. Para ello, hemos utilizado una estrategia argumentativa que nos ha conducido a un segundo objetivo colateral: mostrar la existencia de dos corrientes teóricas en el estudio de las convenciones sociales que, a grandes rasgos, pueden ser calificadas de corriente económica y corriente sociológica. La primera bebe del concepto de convención introducido por David Hume, mientras que la segunda se basa en la definición de convención social de Max Weber. Antes de ocuparnos del objetivo principal, diremos algunas palabras acerca de este objetivo secundario.

En una tesis elaborada recientemente (Latsis, 2004), así como en un artículo derivado de la misma (Latsis, 2005), se ha afirmado que las dos corrientes presentadas anteriormente se derivan de la noción de convención presente en Hume. Tanto es así que al autor define a ambas como tradiciones *neo-humeanas* en el estudio de las convenciones sociales⁶⁰. El motivo que le lleva a considerar la propuesta de *convención como principio normativo* de Gilbert como una propuesta *neo-humeana* es una cita ya introducida del propio Hume en la que éste afirma:

⁶⁰ Latsis (2005) considera a Lewis el máximo exponente de la primera, mientras que a Gilbert como la máxima representante de la segunda.

Una vez que este común sentimiento de interés ha sido mutuamente expresado y nos resulta conocido a ambos, produce la resolución y conducta correspondiente. Y esto es lo que puede ser denominado con bastante propiedad convención o mutuo acuerdo, aun cuando no exista la mediación de una promesa, dado que las acciones de cada uno de nosotros tienen referencia a las del otro y son realizadas en el supuesto de que hay que realizar algo a favor de la otra parte (Hume, 1978 [1740]: III.ii.2).

La referencia que hace Hume a que el “sentimiento de interés común” es “mutuamente expresado” le basta a Latsis para derivar de la misma la propuesta de Gilbert. Sin embargo, como se ha argumentado anteriormente, existen numerosos elementos que llevan a pensar que dicha interpretación es errónea. En primer lugar, la propia Gilbert rechaza hacer una interpretación de la noción de convención en Hume en la nota 4 del capítulo VI de *On Social Facts* (Gilbert, 1989: 478). En segundo lugar, en la primera página de dicho capítulo sexto, Gilbert remite al lector a la definición de convención de Weber. En tercer lugar, entre las características que anteriormente atribuíamos a la definición humeana de convención no se encuentra ninguna que pueda dar cabida a la interpretación de la convención como un principio normativamente vinculante, ya que el propio Hume critica explícitamente a los teóricos del contrato y, frente a la propuesta teórica de estos últimos, propone su noción de convención. En cuarto y último lugar, refiriéndonos al texto de Hume que Latsis cita para sostener su argumento, si uno sigue leyendo se da cuenta de que Hume no está pensando en la aceptación conjunta de un principio normativo, sino más bien en la necesaria existencia de una regla de reciprocidad, como ya argumentábamos anteriormente.

Una vez que hemos puesto de manifiesto, de nuevo, algunas de las diferencias entre ambas corrientes, a continuación nos ocuparemos del objetivo principal de este capítulo: aportar una definición operativa del concepto de convención social.

4.2. Definición de convención social

Al inicio del capítulo anterior se introducían dos citas – una de Weber (1964 [1922]: 264) y otra de Kliemt (1986: 169) – para poner de manifiesto el peligro de confundir las regularidades de hecho que observamos en la conducta humana con el conjunto de principios que contribuyen a la regulación normativa de dicha conducta. En otras palabras, no podemos olvidar el carácter “bifronte” de las convenciones sociales. Es por ello que cualquier definición del concepto de convención social debe responder a dos preguntas: (1) ¿Qué características definen el tipo de regularidad de hecho al que denominamos convención social? (2) ¿Por qué los sujetos individuales ajustan su comportamiento a dicha regularidad? Si la definición que presentamos a continuación responde satisfactoriamente a ambas preguntas, podremos decir que disponemos de una definición provisional que da cuenta de lo que intuitivamente denominamos convención social. Sin más, definimos convención social como:

(a) Una *regularidad R* empíricamente observable en el comportamiento humano que cumple las siguientes características:

(a.1.) se trata de una regularidad *R arbitraria e indeterminada*, en tanto que podría ser identificada una regularidad alternativa *R'* que respondería de igual forma a la demanda originada por la estructura de la situación de interacción social a partir de la cual ha surgido *R*;

(a.2.) se trata de una regularidad *estable*, cuya estabilidad puede deberse a dos motivos: bien los individuos no tienen incentivos para desviarse de la misma

(problemas puros de coordinación) o bien cualquier desviación de la convención está penalizada por la desaprobación de aquellos que participan en ella, ya que para éstos dicha desviación no supone sólo comportarse de forma diferente a la regularidad *R*, sino que supone desviarse de un principio normativo. Ambos motivos no son excluyentes.

(b) Un *principio*⁶¹ *normativo* que indica cuál es la acción prescrita entre el conjunto de acciones posibles en una situación de interacción social concreta. Dicho principio tiene las siguientes características:

(b.1.) se trata de un principio *tácito*: no es necesario que sea expresado de forma explícita para que sea aceptado por los participantes en la convención;

(b.2.) se trata de un principio de acción *conjunto*: remite a la existencia de una colectividad, en el seno de la cual el principio cobra sentido;

(b.3.) la *validez* de dicho principio depende de los juicios potenciales respecto de a lo *desapropiada* que resultaría la conducta desviada.

Aunque analíticamente resulte necesario separarlos, hay que decir que *regularidad* y *principio* no son entidades independientes sino que forman parte de la misma cosa⁶².

⁶¹ Mantenemos aquí la noción de principio ya que es la utilizada por Gilbert (1989) y cuyo significado preciso ha sido aclarado en el apartado anterior. S. Miller (1992) utiliza la noción de “procedimiento” (*procedure*) con un significado bastante similar.

⁶² Esta “doble naturaleza” de las convenciones ha sido puesta de manifiesto, empleando términos distintos, por otros autores, especialmente algunos de los integrantes de la Escuela Francesa de la Economía de las Convenciones. Por ejemplo, Favereau (1986) habla de “Convención₁” para referirse a un

Tras hacer explícita nuestra definición de convención social, nos preguntamos si somos capaces de responder a las dos preguntas lanzadas más arriba – la primera acerca de las características que definen el tipo de regularidad de hecho al que denominamos convención social y la segunda acerca de las razones que llevan a los sujetos a ajustar su comportamiento a dicha regularidad. La respuesta sería un sí, al menos un sí provisional. En primer lugar, a través de las características de *arbitrariedad*, *indeterminación* y *estabilidad* somos capaces de distinguir a las convenciones sociales de otras regularidades empíricamente observables en el comportamiento humano. Como hemos visto anteriormente, las convenciones no pueden ser clasificadas como *costumbres*, ya que en estas últimas falta la aprobación y desaprobación del resto. Las convenciones son distintas a lo que denominamos *moda*, en tanto que esta última es efímera por definición y no estable como las convenciones. Las convenciones pueden coincidir con las *situaciones de intereses*, aunque sólo en un escenario de condiciones muy concreto: cuando nos hallamos ante problemas puros de coordinación. Por tanto, estamos en condiciones de afirmar que hemos respondido satisfactoriamente a la primera de las preguntas planteadas. En cuanto a la segunda pregunta, relativa al carácter normativo de la convención, hemos identificado dos mecanismos que actúan como explicaciones plausibles del *porqué* de la aceptación de las convenciones. El mecanismo más general señala la *aprobación* del resto como principal estímulo de la conformidad con la convención. En situaciones muy concretas (problemas puros de coordinación) la función cognitiva de las convenciones sociales es suficiente.

tipo de representación social o colectiva, y de “Convención₂” para referirse a una pauta de comportamiento interindividual; Batifoulier *et al.* (2001) distinguen entre “modos convencionales de evaluación”, para referirse a “convenciones de pensamiento”, y “reglas convencionales”, para referirse a “convenciones de comportamiento”; Orléan (1999) utiliza la noción de “convención de interpretación” para referirse al componente evaluativo o normativo de las convenciones.

Como última defensa de nuestra definición dual de la noción de convención social, presentamos la siguiente tabla (Tabla 1), en la que se puede observar cómo el mayor o menor peso de la característica de *regularidad* y de la de *principio* en la convención nos permite derivar una tipología de convenciones sociales cuyo poder explicativo es mayor del que se derivaría de una definición unidimensional de la noción de convención social.

Tabla 1: Tipología de convenciones

	Principio difuso	Principio bien definido
Regularidad muy extendida	Convenciones emergentes	Convenciones establecidas
Regularidad poco extendida	Convenciones difusas	Convenciones decadentes

Así, denominamos *convenciones emergentes* a aquellas que responden a situaciones donde se empieza a extender una regularidad en el seno de una población aunque todavía no está claro el principio normativo, es decir, no está clara cuál es la conducta prescrita en una situación recurrente de interacción social. No es extraño que el concepto de *convención emergente* haya sido utilizado por los teóricos de juegos para denominar a aquellas situaciones donde *emerge* una convención de forma tácita y paulatina (ver, entre otros, Rankin, Van Huyck y Battalio, 1997). Podemos encontrar multitud de ejemplos cotidianos en situaciones novedosas, como las relacionadas con el uso de nuevas tecnologías. Por ejemplo, consideremos las siguientes *convenciones emergentes* relacionadas con el uso del correo electrónico:

El correo electrónico emplea la mayoría de convenciones del lenguaje escrito estándar, excepto que puede variar ampliamente el grado de formalismo. Las convenciones relativas a las aperturas y los cierres son diferentes a las de las cartas tradicionales.

Las listas de correo son similares a los correos electrónicos, pero dado que van dirigidas a un número más amplio de receptores, el lenguaje tiende a ser considerablemente más formal.

La comunicación escrita en tiempo real [*Messenger, chats*] a través del ordenador [...] conduce a un conjunto de usos especiales del lenguaje muy distintivos. [Aunque] es distinto de otros géneros (correo electrónico, listas de correo, grupos de noticias, foros), algunos elementos originados en el *chat* se vierten en estos géneros. (Cumming, 2003)

Todo este conjunto de convenciones asociadas al uso de las nuevas tecnologías pueden ser calificadas de emergentes, debido a la ausencia de principios reguladores claros, a pesar de la extensión de su práctica.

Las *convenciones establecidas* son aquellas en las que podemos identificar el principio que guía la acción de forma precisa y donde la regularidad de comportamiento definida por tal principio está suficientemente extendida. Una característica interesante de las *convenciones establecidas* es que, a menudo, las personas tienden a olvidar el carácter arbitrario de las mismas y las ven como totalmente necesarias. Con respecto a la convención que prescribe conducir por la izquierda en el Reino Unido, Binmore (1994: 290) afirma que “algunos creen que ‘conducir por la derecha’ más que arbitrario es perverso. Los británicos, por ejemplo, dicen que es razonable conducir por la izquierda porque los diestros pueden mantener la mano y el brazo que dominan más en el volante mientras usan la zurda para cambiar de marcha.”

Las *convenciones difusas* representan normalmente situaciones de transición donde ni está claro el principio predominante en un momento determinado, ni se observa una regularidad que se impone sobre el resto. Por ejemplo, un colega inglés nos señalaba que, una vez que la convención que supone “estrechar la mano” de modo

formal ha caído en desuso entre los jóvenes británicos, y no habiéndose incorporado una práctica distinta, éstos no saben muy bien qué fórmula emplear al saludar a otra persona. Un caso parecido ocurre cuando el encuentro se produce entre dos personas de nacionalidad distinta, donde existen principios que prescriben comportamientos distintos en lo que al saludo se refiere. Un tercer ejemplo de *convención difusa* lo encontramos en sociedades multilingües en las que cualquier conversación comienza con un problema de coordinación previo para determinar qué idioma va a ser empleado en la misma.

Por último, las *convenciones decadentes* responden a situaciones donde, aunque está bien definido el principio que guía la conducta, dicha práctica está en claro proceso de transformación o desaparición. En países como España éste es el caso de las convenciones que implican comportamientos diferenciados por género, tales como la convención que prescribe “servir primero a las mujeres en la mesa”, “ceder el paso a una mujer al atravesar una puerta”, etc.

La breve alusión realizada a estos cuatro tipos de convenciones sólo pretende poner de manifiesto la necesidad de considerar las “dos caras” de las convenciones sociales. Un análisis más profundo de los mismos se escapa a los objetivos centrales de esta tesis.

4. 3. Conclusiones

El objetivo principal de esta primera parte de la tesis ha sido establecer una definición provisional del concepto de convención social con la que podamos avanzar en los

aspectos teóricos y empíricos presentes en las partes posteriores. En el proceso de construcción de tal definición hemos planteado otras cuestiones, tales como: la diferencia entre la noción de convención empleada por los teóricos económicos y aquella manejada por los sociólogos; la diferencia entre la noción de convención y otros conceptos, como los de *derecho*, *norma moral*, *costumbre*, *moda* o *situación de intereses*; o el establecimiento de una tipología de convenciones en relación con la mayor o menor presencia de los componentes de la definición de convención propuesta.

Todas estas cuestiones, aunque pueden ser valoradas independientemente, sólo cumplen aquí la función de preparar el terreno conceptualmente para el resto de la tesis. De este modo, nuestra definición será puesta a prueba teóricamente (en la parte II) y experimentalmente (en la parte III). Una vez que haya sido puesta a prueba nos preguntaremos si la definición aportada es adecuada o si merece ser modificada en alguno de sus componentes.

PARTE II

APROXIMACIÓN TEÓRICA A LA NOCIÓN DE CONVENCION SOCIAL

Como se ha puesto de manifiesto en la primera parte de la tesis, la investigación teórica contemporánea acerca de la noción de convención social no responde ni campos ni a enfoques homogéneos. A grandes rasgos, podemos distinguir entre aquellos autores que se dedican al estudio de las dinámicas evolutivas que soportan las convenciones y aquellos otros que están más preocupados por el componente normativo de las convenciones sociales.

En cierto modo, si uno sigue de cerca los avances en ambas líneas de investigación, se da cuenta de que hasta cierto punto parecen planteamientos irreconciliables. Así, los planteamientos evolutivos de autores como Boyer y Orléan (1992), Sugden (1986; 1989; 1995) o Young (1993; 1996; 1998a), por nombrar sólo a algunos de los más destacados, apenas han tenido eco en la teoría sociológica. Algunas de las razones que pueden ser esgrimidas para ello son: la excesiva formalización matemática de los argumentos empleados por estos autores; la explicación puramente conductual, basada en disposiciones de comportamiento, de las convenciones sociales; o la ausencia de toda referencia al componente normativo de las convenciones sociales. Por otra parte, tampoco parece que los argumentos sociológicos contemporáneos – como, por ejemplo, los desarrollados por los miembros de la escuela francesa de la Economía de las Convenciones – acerca de la noción de convención social estén presentes en las discusiones de los teóricos evolutivos. Para encontrar las razones de este hecho sólo necesitamos darles la vuelta a las que esgrimíamos unas líneas más

arriba, a saber: la poca o nula formalización de los argumentos sociológicos; la recurrencia de los aspectos normativos que, en la forma en la que son presentados por numerosos sociólogos, difícilmente pueden ser reducidos a creencias y preferencias individuales; o la dificultad para describir mecanismos que permitan una interpretación de las convenciones en el plano individual.

Dado este panorama teórico, el objetivo de esta segunda parte de la tesis es el desarrollo de un marco teórico formal en el que tengan cabida las dos nociones de convención descritas en la primera parte de la misma, a saber: las *convenciones de comportamiento* y las *convenciones normativas*. En concreto, nuestra propuesta teórica tendrá que ver con una posible incorporación del componente normativo de las convenciones, tal y como ha sido presentado en la parte primera, en un marco evolutivo general de las convenciones sociales. Dicha propuesta se desarrollará en un único capítulo (capítulo 5) en el que, tras una breve discusión acerca de la pertinencia de la formalización de los argumentos desarrollados hasta el momento, se presentará nuestro marco teórico formal, una interpretación evolutiva del mismo y una discusión informal de las posibles implicaciones de esta propuesta teórica.

CAPÍTULO 5

LA ESTABILIDAD DE LAS CONVENCIONES SOCIALES: UNA PROPUESTA TEÓRICA EVOLUTIVA

5.1. Introducción

Si nuestro objetivo es desarrollar un modelo explicativo a propósito de la evolución de las convenciones sociales que recoja las principales aportaciones tanto de la teoría sociológica, como de la teoría económica, tendremos que hacer frente a un primer problema: ambas disciplinas, en su estado actual, hablan un lenguaje distinto. En otras palabras, el conjunto de convenciones que rigen la investigación económica actual difieren de tal modo de las que imperan en el discurso sociológico que en contadas ocasiones encontramos nexos teóricos de unión que nos permitan establecer proposiciones teóricas que resulten útiles a ambos campos de conocimiento⁶³. En este capítulo, nuestro objetivo será utilizar una herramienta ampliamente aceptada en la teoría económica actual: la teoría de juegos evolutiva, para captar un argumento puramente sociológico: el carácter normativo de las convenciones sociales.

⁶³ Las convenciones que regulan lo que es considerado “apropiado” o “no apropiado” en cada una de estas disciplinas imponen un control tan férreo sobre el comportamiento de los investigadores adscritos a cada campo que la trasgresión de las mismas suele conducir a distintos grados de lo que Weber denominaba “boicot declarado” por parte de los miembros de cada uno de estos “compartimentos estanco” académicos. Sirvan como ejemplo dos actitudes – una por cada bando – que, a modo de anécdota, pueden ilustrar esta reflexión. Por un lado, ¿qué sociólogo de la elección racional no ha sido acusado alguna vez de contribuir a la perpetuación del capitalismo más despiadado sólo por utilizar unas simples ecuaciones matemáticas que sólo pretenden hacer un argumento inteligible? Por otro, ¿qué sociólogo no ha sido acusado de mantener una postura pre-científica cuando ha intentado exponer sus argumentos ante un grupo de teóricos microeconómicos contemporáneos? Tanto una actitud, como la otra, han dificultado sobremanera la comunicación entre ambas disciplinas académicas.

Nótese que nuestra aplicación de una herramienta matemática-formal, como es la teoría de juegos evolutiva (TJE, en adelante), no pretende fines predictivos, sino más bien explicativos. Por tanto, se sitúa en la línea de la propuesta de “reconstrucción racional” de las normas sociales de Ullmann-Margalit (1977) o la tradición de los “tipos ideales” de Weber (1964 [1922]). Dicho esto, y antes de adentrarnos de lleno en nuestra propuesta de formalización, creemos necesario responder a algunas cuestiones relacionadas con la pertinencia de la formalización matemática de los argumentos que estamos desarrollando en esta tesis.

La primera pregunta que merece la pena hacerse es la siguiente: ¿qué aporta la formalización matemática al trabajo realizado en la primera parte de esta tesis? En primer lugar, permítasenos utilizar un argumento un tanto obvio: si buena parte de la literatura empleada en la misma está expresada mediante modelos formales, algo tendremos que decir acerca de las ventajas y desventajas de éstos y, para ello, debemos comprender el lenguaje, a menudo críptico, empleado por los mismos. Una segunda respuesta más profunda coincide con la desarrollada en un texto reciente por Noguera (2006), en la que éste esgrime las razones de “inteligibilidad” y “consistencia lógica” para defender el uso de la formalización lógica y el uso de modelos formales en sociología. Noguera afirma que ambos procedimientos buscan:

promover la claridad y la inteligibilidad del discurso, la consistencia lógica y una precisa definición de los términos y conceptos empleados. Por lo tanto, busca evitar el impresionismo, el intuicionismo, las pretensiones literarias, las ‘sutilezas’, la ambigüedad, la vaguedad, los dobles significados y los juegos de palabras, la ausencia de sistematicidad en los argumentos y el ámbito de la ‘sugestión’ y la ‘evocación’.

No abundaremos más sobre esta característica de inteligibilidad, ya que nos es el objetivo de esta tesis una defensa del uso de modelos formales en la sociología⁶⁴, más allá de justificar el empleo que se da a éstos en este trabajo.

Del mismo modo, modelar formalmente una situación social implica un esfuerzo por captar las características definitorias de la misma. Así, el conjunto de sentencias que describen verbalmente cuál es el comportamiento esperado de una serie de variables puede ser transformado en un modelo lógico o matemático en el que se especifiquen las relaciones exactas entre las distintas variables en cuestión. Desde luego, no hay que olvidar que modelar una situación tiene como contrapartida una pérdida de algunos de los matices que podemos encontrar en la compleja realidad social. Pero, ¿es que existe alguna otra forma de hacer ciencia? En todo caso, los supuestos realizados en este capítulo pueden ser cuestionados a partir de su mayor o menor “realismo” respecto de otros supuestos alternativos que pudieran ser empleados para reconstruir los argumentos teóricos desarrollados hasta el momento.

El uso de modelos formales también puede ser tachado de un ejercicio puramente especulativo, debido a que algunos de los supuestos que se postulan pueden ser difícilmente contrastados en situaciones empíricas. Ante esta objeción se pueden ofrecer dos respuestas distintas. En primer lugar, no es cierto que la pura especulación teórica sea totalmente inútil para las ciencias sociales. Así, la teoría de juegos evolutiva ha proporcionado explicaciones plausibles acerca de multitud de fenómenos sociales, como la confianza (Güth y Kliemt, 1998) o el altruismo (Güth y Bester, 1998), que,

⁶⁴ Noguera (2006) realiza una argumentación contundente acerca de las ventajas del uso de modelos formales en las ciencias sociales. Sobre la característica de inteligibilidad de las explicaciones científicas ver también Francisco (1997: 113 y ss.). Para una exposición acerca de algunos de los ámbitos de aplicación de los modelos formales propios de la teoría de la elección racional en sociología ver Aguiar, Criado y Herreros (2003).

escapándose a toda lógica de mercado, no eran captados apropiadamente por otras propuestas teóricas, como la de la teoría de la elección racional más estándar. Por tanto, si bien en multitud de ocasiones los resultados de este tipo de investigaciones no deben ser interpretados como afirmaciones acerca de regularidades empíricas concretas, sí que pueden ofrecernos una interpretación posible de las condiciones de emergencia y de mantenimiento de dichas regularidades. Justo esta capacidad interpretativa es lo que se persigue en esta segunda parte de la tesis. En segundo lugar, como se tratará de mostrar en la tercera parte de la misma, en las últimas tres décadas se ha producido un importante florecimiento de la aplicación del método experimental para poner a prueba teorías que, con anterioridad, quedaban al margen de la comprobación empírica.

Por último, es importante que distingamos el tipo de análisis que se va a llevar a cabo en este capítulo, de otros análisis que se le asemejan pero que, de ningún modo, son idénticos.

En primer lugar, diremos que nuestro modelo no es un modelo de elección racional, al menos en la versión estándar que domina la teoría económica neoclásica. Intuitivamente en la primera parte hemos argumentado por qué la teoría de la elección racional (TER, a partir de ahora) no nos ayudaba demasiado a resolver problemas de coordinación como los representados en las Figuras 1, 2 ó 3. Además, aceptamos el reto planteado por Güth y Kliemt (1998), intentando superar algunos de los aspectos más problemáticos de la TER. En concreto, estos autores plantean cuatro “limitaciones” de la TER, así como algunas “recetas” para superarlas empleando una aproximación evolutiva. Éstas son:

1. “En primer lugar, en la teoría de la elección racional estándar, las preferencias individuales se toman como dadas [...] pero más bien deberíamos buscar las formas y los medios para integrar la formación de preferencias en nuestros modelos.”
2. “En segundo lugar, el supuesto de la elección racional de que todas las decisiones son tomadas intencionalmente y estrictamente de forma oportunista [...] parece fuera de toda evidencia empírica. Los individuos eligen de acuerdo con reglas o disposiciones de naturaleza general más que caso por caso.”
3. “En tercer lugar, [...] el comportamiento humano no está exclusivamente determinado por las expectativas de consecuencias futuras de las alternativas de elección [...] de algún modo debemos incorporar el hecho de que el comportamiento no es determinado exclusivamente por ‘delante’ sino también empujado por ‘detrás’.”
4. “Finalmente [...] los decisores reales, al contrario que los ideales, no ‘saltan’ instantáneamente a la ‘solución’ de un problema de elección. Éstos llegan a la ‘solución’, si es que llegan, de una forma gradual típicamente caracterizada por [un proceso] de prueba y error.” (Güth y Kliemt, 1998: 378).

¿Cuántos de los postulados propuestos por Güth y Kliemt cumple el modelo que presentaremos a continuación? Pues bien, excepto el primero (no se aporta una teoría precisa acerca de la formación de preferencias), los demás son alcanzados con solvencia. En nuestro modelo, los individuos siguen principios generales de

comportamiento⁶⁵; su elección está completamente determinada por dichos principios, por lo que no se trata de maximizadores consecuencialistas⁶⁶; y las soluciones del mismo son alcanzadas a través de un proceso de prueba y error⁶⁷.

En segundo lugar, nuestro modelo se ajusta fielmente a la propuesta original de TJE de autores como Maynard Smith y Price (1973) o Maynard Smith (1974; 1982). Esto quiere decir que utiliza como criterio para estudiar la estabilidad de los comportamientos (en nuestro caso principios) humanos el concepto de estrategia evolutivamente estable (EEE), que ya hemos introducido brevemente en la parte I. Así, los argumentos aquí presentados nada tienen que ver con otras teorías que utilizan el calificativo de “evolutivas” como son: la denominada “economía evolutiva”, asociada tradicionalmente a autores como Schumpeter; los estudios sobre “evolución de las preferencias”, como ya hemos dejado claro más arriba; o los estudios sobre la “supervivencia del mercado”⁶⁸.

Tras discutir la pertinencia del estudio formal de la noción de convención social, tiene sentido plantear una pregunta que, en cierto modo, ya ha sido respondida en el apartado anterior, a saber: ¿por qué los modelos evolutivos suponen la aproximación más adecuada para estudiar las propiedades formales de las convenciones sociales? La respuesta puede ser derivada a partir de las dos características que definen el aspecto de

⁶⁵ De este modo, el modelo se ajusta a la propuesta de Macy (1997), en la que el sociólogo norteamericano aboga por un tipo de racionalidad emergente caracterizada por la evolución de reglas generales de comportamiento.

⁶⁶ Aquí nos estamos refiriendo a que los sujetos no serían consecuencialistas en el sentido de consecuencialistas en sus actos. Quizá podría argumentarse que se trata de un tipo de los denominados “consecuencialistas de las reglas”. De esto nos ocuparemos en la tercera parte de la tesis, a la luz de los resultados experimentales obtenidos.

⁶⁷ Situando la “interacción”, más que la decisión racional, en el centro del modelo explicativo. Sobre este tema, ver Giner (1997: 94 y ss.).

⁶⁸ Una discusión sobre las principales diferencias entre estas líneas de investigación puede ser encontrada en Weibull (1998).

regularidad de comportamiento de las convenciones sociales. Como hemos definido anteriormente, tales regularidades son al mismo tiempo arbitrarias, indeterminadas y estables. Las características de arbitrariedad e indeterminación pueden ser representadas adecuadamente a través de los juegos de coordinación, tanto si se trata de juegos puros de coordinación (como los representados en las Figuras 1 y 2), como si empleamos juegos de coordinación asimétrica (como el de la Figura 3). La característica de estabilidad supone equiparar las convenciones a equilibrios estables en este tipo de juegos. Recordemos que para que un equilibrio sea estable debe cumplir dos características: debe ser un equilibrio de Nash y al mismo tiempo cumplir una segunda condición de estabilidad, tal como fue definida en el capítulo anterior. Por tanto, intuitivamente no tenemos ningún problema a la hora de modelar el aspecto de regularidad social de las convenciones. Para estudiar el aspecto de principio normativo tendremos que introducir un elemento que no se encuentra en la literatura precedente: la existencia de un conjunto de principios normativos que prescriben cómo jugar una determinada clase de juegos de coordinación. Por último, estos principios o *convenciones normativas* serán analizados evolutivamente asumiendo que los individuos están comprometidos a ejecutar el comportamiento sugerido por un determinado principio al tomar parte en la clase de juegos ante la que se van a enfrentar.

En resumen, nuestro análisis se compone de tres elementos: una subclase de juegos de coordinación asimétrica, una serie de principios generales que prescriben cómo comportarse en cada momento de interacción dentro de la secuencia de juegos que se ajustan a la subclase definida y un mecanismo de selección que determinará bajo

qué condiciones un determinado principio es evolutivamente estable. Cada uno de estos elementos será discutido ampliamente en los siguientes apartados.

5.2. Clase de juegos: “*The Game of Life*”

Decíamos en la introducción de este capítulo que uno de los aspectos sobre los que un modelo formal debe rendir cuentas es acerca del realismo de sus supuestos. En aras de dicho realismo, a partir de ahora dejaremos de utilizar los juegos simples que se han empleado en la primera parte con un objetivo meramente ilustrativo, para aproximarnos a lo que Güth y Napel (2002) han llamado metafóricamente “*the game of life*”. La intuición que se halla detrás de esta idea es que para estudiar cómo los individuos se enfrentan a situaciones reales asimilables a las matrices empleadas convencionalmente por los teóricos de juegos, “uno tiene que estudiar la evolución del comportamiento o de los determinantes del comportamiento no sólo para un juego sino para el ‘juego de la vida’, el cual comprende una variedad de tipos de juegos diferentes que los agentes pueden o no pueden diferenciar” (Güth y Napel, 2002: 3)⁶⁹. Desde luego, la aplicación de esta sugerencia tiene consecuencias nefastas para la armonía del modelo presentado, pero no es armonía lo que buscamos con nuestro modelo, sino una forma clara e intuitiva de representar formalmente nuestras ideas. En general, como veremos, los comportamientos que el modelo predice dependen en gran medida del contexto, es decir de los parámetros concretos que configuren cada uno de los juegos; algo que no debe resultar extraño a los sociólogos, pero que gusta tan poco a los economistas.

⁶⁹ Aquí realizamos un uso un tanto libre de la metáfora del “juego de la vida”. En el sentido empleado por Güth y Napel (2002), este término se refiere a situaciones en las que los jugadores se enfrentan a distintos juegos de forma sucesiva, a estructuras de interacción diferentes, y no sólo a diversas versiones de la misma clase de juegos básica. Sin embargo, como se comprobará en la tercera parte de la tesis, las distintas concreciones numéricas que puede tomar nuestra clase de juegos hace que, desde un punto de vista psicológico, los jugadores crean estar afrontando estructuras de interacción sustancialmente distintas. Por ello, creemos que el uso libre que aquí se da a dicha metáfora no queda tan lejos del sentido original de la misma.

Dicho lo anterior, la forma a partir de la que en este capítulo nos aproximamos a la metáfora del “juego de la vida” es a través de la definición de una clase de juegos que, dentro de unos límites preestablecidos, pueden tomar tantos valores como seamos capaces de imaginar.

Así, consideraremos una subclase de juegos bimatrixiales (2x2) en los que interactúan dos jugadores (A y B), cada uno de ellos enfrentándose a dos posibles acciones diferentes (X , Y). La estructura de preferencias de ambos jugadores queda definida por cuatro parámetros, a y b para el jugador B y c y d para el jugador A , los cuales satisfacen las dos inecuaciones $1 > a > b > 0$ y $1 > c > 0 > d$, así como determinan la siguiente matriz de pagos, representada en la Figura 4:

Figura 4: Clase de juegos básica

		<i>B</i>	
		X_B	Y_B
<i>A</i>	X_A	1, a	d, 0
	Y_A	0, b	c, 1

Los pagos deben ser interpretados en el orden convencional, es decir, de izquierda a derecha, el primer pago corresponde al jugador A mientras que el segundo pago corresponde al jugador B . Además de los dos equilibrios estrictos $X = (X_A, X_B)$ e $Y = (Y_A, Y_B)$, existe un equilibrio en estrategias mixtas (p^*, q^*) , donde $p^* = \Pr\{X_A\}$, $q^* = \Pr\{X_B\}$ y cuyos pagos son $U_i(p^*, q^*)$.

$$p^* = \frac{1-b}{1-b+a} \quad , \quad q^* = \frac{c-d}{1+c-d}$$

$$U_A(p^*, q^*) = \frac{c}{1+c-d} \quad , \quad U_B(p^*, q^*) = \frac{a}{1-b+a}$$

No es difícil deducir observando los parámetros introducidos en la matriz que, dependiendo del valor de estos, podríamos imaginar infinitos juegos. Sin embargo, dadas las restricciones introducidas, cada uno de dichos juegos tendría tres equilibrios de Nash, dos en estrategias puras y uno en estrategias mixtas, tal y como han sido definidos. Por tanto, seguimos teniendo el mismo problema de “selección de equilibrios” propio de los problemas de coordinación que se hallan en la base de la emergencia de las convenciones sociales.

La literatura que ha explorado problemas de selección de equilibrios en entornos donde existe una multiplicidad de éstos, como es nuestro caso, recomienda comenzar considerando los dos equilibrios estrictos como candidatos a la solución del problema de coordinación planteado (ver Harsanyi y Selten, 1988). Sin embargo, debido a la coexistencia de X e Y como equilibrios estrictos, esta recomendación por sí sola no resuelve la incertidumbre estratégica presente en el juego⁷⁰. Así, incluso esta idea más refinada de equilibrio no ayuda a los jugadores a resolver su problema de coordinación (ver Van Damme, 1987). El problema se incrementa cuando el atractivo de los equilibrios estrictos no es comúnmente conocido y aceptado. A partir de este hecho, otros criterios de selección de equilibrios pueden volverse relevantes. Es éste el punto en el que recurrimos a principios de carácter normativo para interpretar cómo los jugadores consiguen coordinarse en uno de los equilibrios del juego. Pero antes de

⁷⁰ Recordemos que, por definición, esta incertidumbre estratégica se halla siempre presente en la base de la emergencia de las convenciones. De hecho, existe un interés creciente en el estudio de las convenciones sociales como fenómenos “generadores de predictibilidad”, o lo que es lo mismo, “reductores de incertidumbre”. Sobre esta función de las convenciones sociales, ver Engel (2005).

pasar a describir el conjunto de principios que vamos a considerar, nos detendremos en proporcionar algunas intuiciones que pueden ser derivadas de la clase de juegos que acabamos de presentar.

En primer lugar, no queremos dejar de resaltar, de nuevo, la diferencia entre un juego simple y una clase de juegos. Si comparamos el juego simple presentado en la Figura 5 y la clase de juegos introducida en la Figura 4, observamos que el primero es sólo uno de los juegos que responden a la estructura de preferencias establecida por la segunda. De hecho pueden existir tantos juegos distintos como valores puedan tomar los parámetros a , b , c , d de acuerdo con los límites establecidos con anterioridad ($1 > a > b > 0$ y $1 > c > 0 > d$). Esta clarificación es importante porque, como veíamos en la parte primera, las convenciones no tienen por qué responder a una estructura de situación única y perfectamente definida, sino que más bien dan respuesta a un conjunto de situaciones análogas. Así, por ejemplo, la convención que asigna prioridad en el uso de un recurso escaso – asientos en el autobús, en el cine o en el teatro – no puede verse afectada por pequeños cambios en la estructura de la situación, sino que más bien se aplica a todo el conjunto de situaciones que se ajusten a una determinada estructura de preferencias.

Figura 5: Juego de coordinación asimétrica

		<i>B</i>	
		X_B	Y_B
<i>A</i>	X_A	1, 0,5	-0,5, 0
	Y_A	0, 0,2	0,4, 1

Además, definir una clase de juegos supone que también podemos estudiar si se producen alteraciones en el comportamiento de los jugadores cuando los valores de los

parámetros son extremos. Por ejemplo, si el parámetro a es lo suficientemente grande y/o el parámetro c es lo suficientemente pequeño podemos encontrar una inclinación de ambos jugadores a seleccionar conjuntamente el equilibrio X ⁷¹.

En segundo lugar, merece la pena argumentar por qué hemos seleccionado una clase de juegos de coordinación asimétrica. A estas alturas, no es necesario abundar más sobre por qué juegos de coordinación, sin embargo sí que merece la pena explicar el porqué de la conveniencia de introducir asimetrías en nuestro análisis. La gran mayoría de las investigaciones económicas sobre las convenciones, tanto teóricas como empírico-experimentales, analizan situaciones completamente simétricas⁷², tales como los juegos puros de coordinación representados en las Figuras 1 y 2, o el juego de la seguridad – un juego que tiene dos equilibrios simétricos para los jugadores, pero asimétricos desde el punto de vista de los pagos. Este supuesto de partida no debe resultar extraño teniendo en cuenta las reticencias de la ciencia económica a estudiar situaciones donde existe heterogeneidad de comportamientos individuales. Sin embargo, la sociología ha mostrado hasta la saciedad que esto no es así. Recordemos las convenciones de género a las que nos referíamos en el capítulo anterior o la equiparación de la convención a la *costumbre estamental* que realizaba el propio Weber. Por tanto, consideramos que el estudio de las convenciones en entornos asimétricos hace mayor justicia a nuestra intuición cotidiana de las convenciones sociales.

⁷¹ Éste es un razonamiento que se escapa a la teoría de juegos estándar, aunque es algo bastante usual en la “teoría de juegos del comportamiento”. Como esta última postula, aquí estamos más interesados en analizar cómo los individuos juegan realmente los juegos, que en las respuestas derivadas de los axiomas de la elección racional.

⁷² Una interesante excepción es Schotter y Sopher (2003).

Una justificación alternativa sobre el porqué de analizar situaciones asimétricas puede ser encontrada en el texto de Ullmann-Margalit (1977) introducido en el capítulo anterior. En el capítulo cuarto de este trabajo, la autora define cada uno de los equilibrios de una situación similar a la Figura 6 como una “situación paradigmática de desigualdad”. Una situación que, como la propia autora indica, supone “un *status quo* que es un equilibrio en los términos de la teoría de juegos, aunque es inestable” (Ullmann-Margalit, 1977: 162). Dicho *status quo* “implica dos partes (personas o conjunto de personas) de tal modo que una parte está más favorecida que la otra” (Ullmann-Margalit, 1977: 134). La definición de este *status quo* de desigualdad permite a la autora, en primer lugar, discutir el concepto de desigualdad presente en los trabajos de Dahrendorf y otros teóricos de la desigualdad. En segundo lugar, le permite derivar su concepto de *norma de desigualdad*, una norma cuya función es mantener un determinado *status quo* que favorece a una de las partes en cuestión.

Figura 6: *Status quo* de desigualdad

		<i>B</i>	
		X_B	Y_B
<i>A</i>	X_A	2, 1	0, 0
	Y_A	0, 0	1, 2

Como Sugden (1986: 168) ha puesto de manifiesto, aunque la propuesta de Ullmann-Margalit representa un interesante intento de proveer de microfundamentos al análisis teórico de la desigualdad, no logra sus objetivos en tanto que en su análisis no explota los componentes estratégicos contenidos en dicha situación. Al contrario de sus análisis a propósito de las *normas de dilema del prisionero* y de las *normas de coordinación*, la aproximación a las *normas de desigualdad* no escapan al argumento funcionalista *ad hoc* que la propia Ullmann-Margalit pretende evitar en su trabajo.

Nuestro análisis de las convenciones emergentes a partir de la clase de juegos representada en la Figura 4 puede contribuir a dotar de análisis estratégico a la situación paradigmática de desigualdad definida en el trabajo de Ullmann-Margalit.

Pues bien, una vez aclarados estos dos puntos acerca de la clase de juegos seleccionada para el análisis, presentaremos la segunda parte de nuestro argumento formal: la presencia de principios normativos que prescriben soluciones a los problemas de coordinación derivados de nuestra clase de juegos.

5.3. Principios normativos⁷³

Hasta el momento hemos representado formalmente la mitad de nuestro problema teórico. En concreto, hemos planteado una clase de juegos de coordinación asimétrica en la se pueden distinguir dos equilibrios estrictos, o *convenciones de comportamiento* (X e Y). La mitad restante supone dar cuenta del componente normativo de las convenciones sociales. Recordemos que en el capítulo anterior equiparábamos este componente normativo a “un *principio normativo* que indica cuál es la acción prescrita entre el conjunto de acciones posibles en una situación de interacción social concreta.” Para dar cuenta del mismo, en nuestro modelo definiremos un número limitado, pero suficientemente amplio, de principios que prescriben comportamientos distintos en la clase de juegos analizada.

En este sentido, un principio normativo resolvería el problema de coordinación sobre cómo jugar cada uno de los posibles juegos contenidos en nuestra clase de juegos

⁷³ En las partes segunda y tercera de la tesis utilizaremos los términos *convención normativa* y *principio* (o *principio normativo*) de manera indiferenciada.

básica, en el sentido de recomendar un equilibrio (estricto) o, con carácter más general, uno de los cuatro posibles vectores de estrategias puros. Dicha coordinación requiere, por supuesto, que ambos jugadores sean comúnmente conscientes de qué principio normativo debería ser aplicado⁷⁴. A continuación presentamos los ocho principios que van a ser introducidos en nuestro análisis (ver Tabla 2 para un resumen).

A partir de los modelos formales de evolución de las convenciones (Young, 1993; 1996; 1998a), sabemos que el concepto más convincente a la hora de determinar qué convención será la seleccionada es el de *risk dominance* (R)⁷⁵ que, para la clase de juegos considerada, está caracterizado por los tres axiomas de *best reply invariance*, *isomorphic invariance* y *monotonicity* (Harsanyi y Selten, 1988)⁷⁶. Como consecuencia de la *best reply invariance*, nuestra matriz original puede ser transformada en la siguiente:

		B	
		X_B	Y_B
A	X_A	1, a	0, 0
	Y_A	0, b	c-d, 1-b

⁷⁴ Este aspecto constituirá uno de los elementos fundamentales de los diseños de los experimentos utilizados en la tercera parte de esta tesis.

⁷⁵ Para no crear una duplicidad terminológica innecesaria se mantendrá el nombre de los principios y los axiomas matemáticos empleados en inglés. La razón es que así es como aparecen en la mayoría de textos, incluso en las traducciones al español, y traducirlos podría llevar a confusión.

⁷⁶ El concepto de *risk dominance*, el cual fue desarrollado originariamente por Harsanyi y Selten (1988), pretende representar matemáticamente lo que intuitivamente caracterizaríamos como el componente de riesgo presente en cualquier combinación de estrategias en equilibrio. En última instancia nos muestra cómo la elección de determinadas estrategias conducentes a una situación de equilibrio es más arriesgada que otras elecciones posibles. La existencia de equilibrios más arriesgados que otros es un factor importante cuando los jugadores deben aprender a jugar un juego a partir de repeticiones sucesivas del mismo. En estos casos, las penalizaciones que van recibiendo los jugadores al elegir una estrategia u otra de forma repetida hacen que sea difícil la convergencia hacia equilibrios con un alto componente de riesgo, aunque éstos sean más eficientes, es decir, aunque proporcionen mayores beneficios al conjunto de los jugadores.

Esta matriz, puede ser transformada, aplicando el axioma de *isomorphic invariance* (en realidad, a través de lo que se conoce como *affine utility transformation*,) en:

		<i>B</i>	
		X_B	Y_B
<i>A</i>	X_A	$1, \frac{a}{1-b}$	$0, 0$
	Y_A	$0, 0$	$c-d, 1$

Por último, debido al axioma de *monotonicity*, la solución es:

- $X = (X_A, X_B)$ si $a > (c-d)(1-b)$ y
- $X = (Y_A, Y_B)$ si $a < (c-d)(1-b)$;

donde garantizaremos, a través de nuestra elección de parámetros en los experimentos (ver diseño de los experimentos en la parte III), que se cumple que:

$$a \neq (c-d)(1-b)$$

A continuación, consideraremos el principio *loss avoidance* (*L*), que excluye la elección de la estrategia X_A por parte del jugador *A* ya que, en conjunción con la estrategia Y_B supone una pérdida para el jugador *A*. El principio *loss avoidance*, unido al “comportamiento en equilibrio” sugiere, por tanto, la solución (Y_A, Y_B) siempre⁷⁷.

⁷⁷ Sobre resultados experimentales previos a propósito del principio *loss avoidance* como un principio de selección de equilibrios, ver Cachon y Camerer (1996), así como Rydval y Ortmann (2005).

El principio *unbiasedness* (U), por otro lado, asume que el otro jugador utiliza las dos estrategias con la misma probabilidad y que uno reacciona óptimamente a dicha expectativa. Consecuentemente:

- el jugador A elige $\begin{cases} X_A & \text{si } 1+d > c \\ Y_A & \text{si } 1+d < c \end{cases}$
- el jugador B elige $\begin{cases} X_B & \text{si } a+b > 1 \\ Y_B & \text{si } a+b < 1 \end{cases}$

De nuevo, nuestra selección de parámetros garantizará que $1+d \neq c$ y $a+b \neq 1$.

El principio *minimizing risk* (M) sugiere la estrategia que garantiza el mejor entre los peores resultados posibles⁷⁸. De este modo el jugador A debería seleccionar Y_A y el jugador B la estrategia X_B . Nótese que la conjunción de ambas estrategias no produce un conjunto de expectativas que se refuercen a sí mismas, es decir, no conduce a una situación estable de equilibrio.

El principio *sum maximization* (S) simplemente recomienda la constelación de estrategias que producen la mayor suma de ambos pagos⁷⁹, es decir:

- (X_A, X_B) si $a > c$ y
- (Y_A, Y_B) si $a < c$

⁷⁸ En este sentido es similar al criterio maximin propuesto como criterio de justicia por Rawls (1971).

⁷⁹ Este principio coincide con el principio *payoff dominance* propuesto por Harsanyi y Selten (1988) y supone el principio que garantiza una solución eficiente.

donde, de nuevo, excluimos el caso degenerado $a = c$.

El principio *generosity* (G) sugiere a los jugadores elegir lo que es mejor para el otro cuando el otro usa ambas estrategias con la misma probabilidad, como ocurría en el caso de (U). Así, el jugador A debería seleccionar Y_A y el jugador B la estrategia X_B , al igual que ocurría con el principio (M). Este hecho ilustra que el mismo principio puede ser justificado de diversas formas o, lo que es lo mismo, como una misma *convención de comportamiento*, puede estar apoyada en varias *convenciones normativas*.

El principio de *indifference induction* (I) sugiere jugar de forma aleatoria de tal modo que el otro jugador le sea indiferente la opción que elegir. Esto sugiere jugar de acuerdo con el equilibrio en estrategias mixtas (p^*, q^*) .

Por último, el principio *alternating* (A) recomienda alternar la elección de ambas estrategias del siguiente modo: elegir X_i en las rondas pares e Y_i en las impares.

La selección de principios (ver Tabla 2), que puede ser considerada arbitraria a primera vista, está basada en resultados empíricos previos (la mayoría experimentales). Así, sólo nombrando algunos ejemplos recientes, encontramos evidencia experimental de estos principios en los siguientes trabajos: *risk dominance* y *sum maximization* (Haruvy y Stahl, 2003), *loss avoidance* (Rydval y Ortmann, 2005), *alternating* (Schotter y Sopher, 2003), entre otros. La principal diferencia entre estos trabajos y el nuestro es que los trabajos citados, así como bastantes trabajos experimentales que serán revisados en la tercera parte de la tesis, obtienen dichos principios de forma

inductiva, es decir, derivan los mismos del comportamiento de los sujetos en las sesiones experimentales. En cierto modo, y de una forma que nos recuerda bastante a la propuesta humeana de convención, este conjunto de trabajos plantean que los individuos emplean dichos principios de forma inconsciente o, dicho de una forma más correcta, que las acciones de los sujetos acaban reflejando el comportamiento prescrito por estos principios, sin que se produzca una aplicación deliberada de los mismos. Por el contrario, un elemento central en nuestra propuesta teórica es que los individuos reconocen los principios, los evalúan hasta cierto punto y actúan más o menos consecuentemente con ellos. En este sentido, nuestro diseño de los principios de actuación no puede estar más cerca de la formulación práctica de la noción de convención como principio propuesta por Margaret Gilbert. Sin embargo, llegados a este punto, nuestra propuesta se separa de la de Gilbert, o al menos explora un terreno no explorado por esta filósofa analítica, a saber: explora las condiciones bajo las cuáles cada uno de estos principios es evolutivamente estable.

Un principio puede ser estable por muchos motivos. Así, puede ser estable, por ejemplo, porque es óptimo desde un punto de vista individual y cada sujeto concreto no tiene incentivos para desviarse unilateralmente del mismo. También puede ser estable porque ha sido internalizado por una población, adquiriendo de este modo cierto contenido moral y, por tanto, estando sostenido por sanciones, tanto externas como internas, que penalizan la desviación del mismo. Sin embargo, en la primera parte de la tesis negábamos que los principios o *convenciones normativas* a los que nos estamos refiriendo respondieran de forma exclusiva a cualquiera de estas dos explicaciones; aunque ambas, de hecho, suelen desempeñar un papel importante en la estabilidad de los mismos. Nuestra aproximación teórica al estudio de la estabilidad de los principios

será, por el contrario, una aproximación evolutiva. Lo que significa que un principio estable tendrá que ser lo suficientemente robusto para ser escogido en una situación de interacción donde otros principios también están disponibles. En el siguiente apartado presentaremos nuestra caracterización técnica de dicho concepto de estabilidad. Tras esta caracterización, procederemos a discutir informalmente, a través de ejemplos hipotéticos, nuestra interpretación evolutiva.

Tabla 2: Tabla de principios

Abreviatura	Nombre ⁸⁰	Comportamiento prescrito
<i>R</i>	<i>Risk dominance</i>	(X _A , X _B) si $a > (c-d)(1-b)$ (Y _A , Y _B) si $a < (c-d)(1-b)$
<i>L</i>	<i>Loss avoidance</i>	(Y _A , Y _B) siempre
<i>U</i>	<i>Unbiasedness</i>	jugador A elige X _A si $1+d > c$ y Y _A si $1+d < c$ jugador B elige X _B si $a+b > 1$ y Y _B si $a+b < 1$
<i>M</i>	<i>Minimizing risk</i>	jugador A debería seleccionar Y _A jugador B debería seleccionar X _B
<i>S</i>	<i>Sum maximization</i>	(X ₁ , X ₂) si $a > c$ (Y ₁ , Y ₂) si $a < c$
<i>G</i>	<i>Generosity</i>	jugador A debería seleccionar Y _A jugador B debería seleccionar X _B
<i>I</i>	<i>Indifference induction</i>	Jugar de forma aleatoria empleando la estrategia mixta
<i>A</i>	<i>Alternating</i>	Alternar ambas estrategias, es decir elegir X _i en las rondas pares y Y _i en las impares

5.4. Análisis evolutivo

El objetivo de este apartado es el estudio de la estabilidad evolutiva de las *convenciones normativas*, o principios, que han sido introducidas en el apartado anterior. De forma intuitiva, obtendremos una aproximación a la idoneidad de cada uno de estos principios

⁸⁰ Alguien podría preguntarse por qué hemos seleccionado estos nombres para principios que ya tienen una denominación en el terreno de la filosofía (como el maximin), o en el de la teoría de juegos (como el de *payoff dominance*). La razón no es otra que intentar captar la intuición que se halla detrás de estos principios, más que su denominación científica. El nombre empleado para cada principio tiene más que ver con la intuición que podría tener un individuo particular al emplearlo.

para resolver de forma recurrente nuestra clase de juegos básica. Es muy importante que quede claro que aquí no estamos interesados en el estudio de la evolución de las *convenciones de comportamiento*, sino en la evolución de los principios normativos que recomiendan estas *convenciones de comportamiento*. En otras palabras, nuestro objetivo no es estudiar las dinámicas que pueden conducir a una población P de individuos a coordinarse en el comportamiento convencional X o Y , sino estudiar las condiciones que conducen a la estabilidad del uso de un determinado principio en una población donde existe una heterogeneidad de éstos.

Antes de entrar de lleno en el análisis no queremos dejar de insistir en algunos elementos de la teoría de juegos evolutiva que ya han sido comentados tanto en la primera, como en esta segunda parte de la tesis. Como se comprobará a continuación, el lenguaje de la teoría de juegos evolutiva es un lenguaje importado directamente de la biología evolutiva por autores como Maynard Smith (1982). No debe pasar desapercibido que se trata de un intento de recuperar los planteamientos evolutivos en sociología; unos planteamientos que, como han afirmado Güth y Kliemt (1998), cayeron totalmente en desuso después del “ocaso” de lo que se conocía en su momento como *darwinismo social*. Con respecto a este hecho Güth y Kliemt (1998) afirman lo siguiente:

Después del ocaso del Darwinismo Social, las explicaciones evolutivas de los fenómenos sociales se convirtieron en algo cada vez más impopular. Aunque en las ciencias económicas conceptos como el de competición y selección evolutiva aparecieron ocasionalmente, incluso después de la segunda guerra mundial, en otras ciencias sociales tales conceptos fueron rechazados, a menudo con buenas razones. Sin embargo, a finales de siglo las ventajas de la moderna teoría de juegos evolutiva sugieren una reevaluación de las aproximaciones evolutivas a los fenómenos sociales. (p. 377)

Como nuestro propio análisis pondrá de manifiesto, lo que hace diferente a la teoría de juegos evolutiva de otros intentos de aplicar los planteamientos evolutivos a las explicaciones sociológicas es el hecho de quién constituye el sujeto de la explicación evolutiva. Siguiendo a Macy (1997), en nuestro caso, son los principios generales de comportamiento (reglas en el caso de Macy) los que se ven sujetos a las presiones evolutivas. Así, no son las personas las que compiten entre sí produciendo una suerte de selección natural de individuos, sino los principios los que compiten (metafóricamente) por ser utilizados por los individuos. A continuación aplicaremos uno de los conceptos más simples de estabilidad evolutiva para estudiar dicha propiedad en los principios introducidos anteriormente.

Para poder llevar a cabo nuestro análisis, asumimos que los individuos tienen tipos correspondientes a las convenciones normativas elaboradas en el apartado anterior, a saber: R , U , L , M , S , G , I y A . Así, un tipo M , o individuo M , elegirá siempre Y_A como jugador A y X_B como jugador B, mientras que el comportamiento de un individuo U dependerá de los parámetros del juego: como jugador A el individuo U utilizará la estrategia X_A si $1+d > c$ e Y_A si $1+d < c$, y como jugador B utilizará la estrategia X_B si $a+b > 1$ e Y_B si $a+b < 1$. Asumiendo que cada individuo participa en el juego desempeñando ambos roles (modelo de una población), podemos analizar qué poblaciones “monomórficas” – donde todos los individuos son del mismo tipo – pueden ser evolutivamente estables⁸¹.

⁸¹ “Un estado de la población es monomórfico si cada miembro [de ésta] emplea la misma estrategia [principio en nuestro caso] y polimórfico si más de una estrategia está presente” (Hammerstein y Selten, 1994: 932). Nuestro análisis se ciñe a casos en los que toda la población sigue el mismo principio. Como veremos en la parte III, este hecho dota de un importante regusto rawlsiano a nuestro análisis. Así, aunque, como hemos visto anteriormente, nuestra clase de juegos básica presenta diferentes asimetrías entre los jugadores, nuestro análisis evolutivo obliga a éstos a elegir un principio válido para toda la población. En los experimentos presentados en la tercera parte interpretaremos este hecho a través de la metáfora del “velo de ignorancia” (Rawls, 1971).

La respuesta dependerá, en general, de los parámetros del juego. Diremos que un monomorfismo T , en el que $T \in \{R, U, L, M, S, G, I, A\}$, supone una convención normativa evolutivamente estable si existe una región genérica de parámetros, en nuestra clase de juegos básica, donde T constituye la única mejor respuesta a sí misma. Este requisito eliminaría inmediatamente M y G como monomorfismos estables, ya que ambas convenciones normativas implican la misma convención de comportamiento. Por lo tanto, para evitar dicho resultado trivial, combinamos M y G en M/G y nos referimos a M/G como un único principio normativo.

Para analizar la estabilidad evolutiva de cada uno de los principios tenemos que derivar, para cada tipo $T \in \{R, U, L, M/G, S, I, A\}$, el éxito de un tipo $\tilde{T} \in \{R, U, L, M/G, S, I, A\}$ al entrar como mutante \tilde{T} en una población monomórfica T ⁸². Por ejemplo, para $T = U$ y $\tilde{T} = U$ (tanto el “mutante” como la población receptora siguen la convención normativa U) el éxito⁸³ obtenido por éste es el siguiente:

- **1+a** si $1+d > c$ y $a+b > 1$ dado que los individuos de tipo U juegan X_i independientemente de su rol ($i = A, B$),
- **d** si $1+d > c$ y $a+b < 1$ dado que los individuos de tipo U juegan X_A como jugador A y Y_B como jugador B,

⁸² Esto no quiere decir otra cosa que el pago que obtendría un individuo que sigue una determinada *convención normativa* al entrar en contacto con una población cuya totalidad de miembros utiliza la misma *convención normativa*, coincida o no ésta con la del individuo al que nos estamos refiriendo (el “mutante”).

⁸³ Aunque en castellano el término *éxito* tiene una connotación muy positiva, el significado técnico que se le da en esta tesis es bastante más neutro. Así, lo que denominamos “éxito evolutivo” (*evolutionary success*) es el resultado obtenido en el largo plazo por un individuo que selecciona una determinada estrategia, o principio que prescribe qué estrategias seleccionar, como ocurre en nuestro caso. Dicho resultado puede ser incluso negativo, por lo que no siempre el concepto de *éxito evolutivo* tiene que ver con la obtención de un resultado positivo.

- **b** si $1+d < c$ y $a+b > 1$ dado que los individuos de tipo U juegan Y_A como jugador A y X_B como jugador B,
- **c+1** si $1+d < c$ y $a+b < 1$ dado que los individuos de tipo U juegan Y_i independientemente de su rol ($i = A, B$).

Del mismo modo, estudiamos ahora el éxito de un individuo “mutante” $\tilde{T} = M/G$ al interactuar con una población $T = U$. Procediendo del mismo modo que lo hemos hecho anteriormente, obtenemos que el éxito de este “mutante” $\tilde{T} = M/G$ es:

- **a** si $1+d > c$ y $a+b > 1$ dado que los individuos de tipo U juegan X_i independientemente de su rol ($i = A, B$) y los individuos de tipo M/G juegan Y_A como jugador A y X_B como jugador B,
- **c+a** si $1+d > c$ y $a+b < 1$ dado que los individuos de tipo U juegan X_A como jugador A y Y_B como jugador B y los individuos de tipo M/G juegan Y_A como jugador A y X_B como jugador B,
- **b** si $1+d < c$ y $a+b > 1$ dado que los individuos de tipo U juegan Y_A como jugador A y X_B como jugador B y los individuos de tipo M/G juegan Y_A como jugador A y X_B como jugador B,
- **c+b** si $1+d < c$ y $a+b < 1$ dado que los individuos de tipo U juegan Y_i independientemente de su rol ($i = A, B$) y los individuos de tipo M/G juegan Y_A como jugador A y X_B como jugador B.

Al proceder de la misma forma para todas las posibles constelaciones de $T, \tilde{T} \in \{R, U, L, M / G, S, I, A\}$ obtenemos la siguiente tabla de éxitos:

Tabla 3: Tabla de éxitos

R (Ī, T)		T						
		U	M/G	S	R	L	I	A
Ī	U	1+a for $1+d>c$ & $a+b>1$	b+1 for $1+d>c$ & $a+b>1$	a+c for $1+d>c$ & $a+b>1$ & $a>c$	a+1 for $1+d>c$ & $a+b>1$ & $a>(c-d)(1-b)$	d+b for $1+d>c$ & $a+b>1$	(c/1)+(c-d) + (a/1-b+a)	(1+a+b+d)/2 for $1+d>c$ & $a+b>1$
				b+d for $1+d>c$ & $a+b>1$ & $a<c$	d for $1+d>c$ & $a+b>1$ & $a<(c-d)(1-b)$			
		d for $1+d>c$ & $a+b<1$	2 for $1+d>c$ & $a+b<1$	1 for $1+d>c$ & $a+b<1$ & $a>c$	a+d for $1+d>c$ & $a+b<1$ & $a>(c-d)(1-b)$	1+d for $1+d>c$ & $a+b<1$		(2+d)/2 for $1+d>c$ & $a+b<1$
				1+d for $1+d>c$ & $a+b<1$ & $a<c$	c for $1+d>c$ & $a+b<1$ & $a<(c-d)(1-b)$			
		b for $1+d<c$ & $a+b>1$	b for $1+d<c$ & $a+b>1$	a for $1+d<c$ & $a+b>1$ & $a>c$	1+b for $1+d<c$ & $a+b>1$ & $a>(c-d)(1-b)$	b+c for $1+d<c$ & $a+b>1$		(a+b+c)/2 for $1+d<c$ & $a+b>1$
				c+b for $1+d<c$ & $a+b>1$ & $a<c$	1 for $1+d<c$ & $a+b>1$ & $a<(c-d)(1-b)$			
	c+1 for $1+d<c$ & $a+b<1$	1 for $1+d<c$ & $a+b<1$	0 for $1+d<c$ & $a+b<1$ & $a>c$	b+d for $1+d<c$ & $a+b<1$ & $a>(c-d)(1-b)$	b+c for $1+d<c$ & $a+b<1$	(1+c)/2 for $1+d<c$ & $a+b<1$		
			c+1 for $1+d<c$ & $a+b<1$ & $a<c$	c+1 for $1+d<c$ & $a+b<1$ & $a<(c-d)(1-b)$				
	M/G	a for $1+d>c$ & $a+b>1$	b		a for $a>(c-d)(1-b)$	c+b		(a+b+c)/2
		c+a for $1+d>c$ & $a+b<1$		a for $a>c$				

	b for $1+d < c$ & $a+b > 1$		$b+c$ for $a < c$	$b+c$ for $a < (c-d)$ $(1-b)$	
	$c+b$ for $1+d < c$ & $a+b < 1$				
S	$1+a$ for $1+d > c$ & $a+b > 1$ & $a > c$	$b+1$ for $a > c$	$a+1$ for $a > c$	$a+1$ for $a > (c-d)$ $(1-b)$ & $a > c$	$d+b$ for $a > c$
	0 for $1+d > c$ & $a+b > 1$ & $a < c$			0 for $a > (c-d)$ $(1-b)$ & $a < c$	
	$a+d$ for $1+d > c$ & $a+b < 1$ & $a > c$				
	c for $1+d > c$ & $a+b < 1$ & $a < c$				
	$b+1$ for $1+d < c$ & $a+b > 1$ & $a > c$	1 for $a < c$	$c+1$ for $a < c$	$b+d$ for $a < (c-d)$ $(1-b)$ & $a > c$	0 for $a < c$
	1 for $1+d < c$ & $a+b > 1$ & $a < c$			$c+1$ for $a < (c-d)$ $(1-b)$ & $a < c$	
	$b+d$ for $1+d < c$ & $a+b < 1$ & $a > c$				
	$c+1$ for $1+d < c$ & $a+b < 1$ & $a < c$				
R	$1+a$ for $1+d > c$ & $a+b > 1$ & $a > (c-d)$ $(1-b)$	$1+b$ for $a > (c-d)$ $(1-b)$	$a+1$ for $a > (c-d)$ $(1-b)$ & $a > c$	$1+a$ for $a > (c-d)$ $(1-b)$	$d+b$ for $a > (c-d)$ $(1-b)$
	d for $1+d > c$ &				
					$(1+a+b+d)/2$ for $a > c$
					$(1+c)/2$ for $a < c$
					$(1+a+b+d)/2$ for $a > (c-d)$ $(1-b)$

	$a+b>1$ $\& a < (c-d) (1-b)$					
	1 for $1+d>c$ & $a+b<1$ $\& a > (c-d) (1-b)$		$d+b$ for $a > (c-d)$ $(1-b)$ $\& a < c$			
	$d+1$ for $1+d>c$ & $a+b<1$ $\& a < (c-d) (1-b)$					
	a for $1+d<c$ & $a+b>1$ $\& a > (c-d) (1-b)$	1 for $a < (c-d) (1-b)$	0 for $a < (c-d) (1-b)$ $\& a > c$	$c+1$ for $a < (c-d) (1-b)$	$1+c$ for $a < (c-d) (1-b)$	$(1+c)/2$ for $a < (c-d) (1-b)$
	$c+b$ for $1+d<c$ & $a+b>1$ $\& a < (c-d) (1-b)$					
	0 for $1+d<c$ & $a+b<1$ $\& a > (c-d) (1-b)$					
	$c+1$ for $1+d<c$ & $a+b<1$ $\& a < (c-d) (1-b)$		$c+1$ for $a < (c-d) (1-b)$ $\& a < c$			
L	0 for $1+d>c$ $\& a+b>1$	1	0 for $a > c$	0 for $a > (c-d) (1-b)$	$1+c$	$(1+c)/2$
	c for $1+d>c$ $\& a+b<1$					
	1 for $1+d<c$ $\& a+b>1$					
	1 for $1+d<c$ $\& a+b<1$		$1+c$ for $a < c$	$1+c$ for $a < (c-d) (1-b)$		
I	$(a(c-d)/1+c-d) +$ $(1-b/1-b+a)$	$(b(c-d)+1/1+c-d)+$ $(1-b/1-b+a)$	$(a(c-d)/1+c-d) +$ $(1-b/1-b+a)$	$(a(c-d)/1+c-d) +$ $(1-b/1-b+a)$	$(b(c-d)+1/1+c-d)+$ $(d(1-b)+ca/1-b+a)$	$((a+b)(c-d)+1)/$ $2(1+c-d) +$

	for $1+d>c$ & $a+b>1$ $(a(c-d)/1+c-d) +$ $(d(1-b)+ca/1-b+a)$ for $1+d>c$ & $a+b<1$ $(b(c-d)+1/1+c-d)+$ $(1-b/1-b+a)$ for $1+d<c$ & $a+b>1$ $(b(c-d)+1/1+c-d)+$ $(d(1-b)+ca/1-b+a)$ for $1+d<c$ & $a+b<1$		for $a>c$ $(b(c-d)+1/1+c-d)+$ $(d(1-b)+ca/1-b+a)$ for $a<c$	for $a>(c-d) (1-b)$ $(b(c-d)+1/1+c-d)+$ $(d(1-b)+ca/1-b+a)$ for $a<(c-d) (1-b)$		$((1-b)/(1-b+a))$
A	$(1+a)/2$ for $1+d>c$ & $a+b>1$ $(a+c+d)/2$ for $1+d>c$ & $a+b<1$ $(2+b)/2$ for $1+d<c$ & $a+b>1$ $(1+b+d)/2$ for $1+d<c$ & $a+b<1$	$(2+b)/2$	$(1+a)/2$ for $a>c$ $(1+b+c+d)/2$ for $a<c$	$(1+a)/2$ for $a>(c-d) (1-b)$ $(1+b+c+d)/2$ for $a<(c-d) (1-b)$	$(1+b+c+d)/2$	$(2+a+c)/2$

Una vez que hemos calculado el pago que obtendría cada uno de los tipos $\tilde{T} \in \{R, U, L, M/G, S, I, A\}$ al interactuar con una población $T \in \{R, U, L, M/G, S, I, A\}$ podemos determinar qué convenciones normativas son evolutivamente estables aplicando el requisito de estabilidad evolutiva al que ya nos referíamos en el capítulo segundo. Así, para calificar a U como una *convención normativa evolutivamente estable* tendríamos que encontrar una región de parámetros genérica donde:

- $R(U, U) > R(T, U)$ para todo $T \neq U$ y $T \in \{R, U, L, M/G, S, I, A\}$ ⁸⁴,

Por ejemplo, en el caso de $1+d > c$ y $a+b > 1$ encontramos que $R(U, U) = 1+a > a = R(M/G, U)$, lo que significa que una población monomórfica U no puede ser invadida satisfactoriamente por mutantes M/G si sólo se juegan juegos en la subregión genérica $1+d > c$ y $a+b > 1$.

Procediendo de esta forma para el resto de la Tabla 3, obtenemos los siguientes resultados:

Resultado 1. Podemos descartar I^* y M/G^* como convenciones normativas. Estos principios no constituyen estrategias evolutivamente estables (EEE) debido a que no suponen la mejor respuesta a sí mismas en ninguna región de la tabla.

⁸⁴ Nótese que, en este caso, estamos empleando un concepto de estabilidad evolutiva más estricto que el empleado en nuestro análisis del juego de la Figura 3 de la primera parte. Aquí, el resultado obtenido al emplear el principio U deber ser estrictamente mayor ($>$) al que se obtendría de la aplicación de cualquier otro principio.

Resultado 2. Existen varias convenciones normativas que son EEE (ver Tabla 4).

Resultado 3. Existe una multiplicidad genérica de convenciones normativas. Por ejemplo, para $(c-b)(1-d)(1-b) < c$ y [no: $1+d > c$ y $a+b > 1$], R^* y L^* son estrategias evolutivamente estables (EEE). Además, A^* es una EEE para todas las subregiones de parámetros.

Tabla 4: Convenciones normativas evolutivamente estables

EEE	Subregiones de parámetros
U^*	$1+d > c$ y $a+b > 1$ y $a < c$ y $a < (c-d)(1-b)$
S^*	$(c-d)(1-b) > a > c$
R^*	$(c-d)(1-d)(1-b) < a < c$ y [no: $1+d > c$ y $a+b > 1$]
L^*	$c > a > (c-d)(1-b)$
A^*	Para todas las subregiones de parámetros

5.5. Interpretación de los resultados formales

Los resultados obtenidos hasta el momento nos permiten refinar los argumentos desarrollados en la primera parte de la tesis. A continuación, discutimos brevemente las implicaciones de cada uno de los tres resultados.

Resultado 1: Convenciones normativas no evolutivamente estables

El resultado 1 demuestra formalmente dos intuiciones básicas: la inestabilidad evolutiva del equilibrio en estrategias mixtas y la inestabilidad evolutiva de comportamientos “generosos” en un modelo de población monomórfica.

La inestabilidad del equilibrio en estrategias mixtas (nuestro principio *I*) demuestra que resulta siempre más beneficioso el empleo de un principio normativo evolutivamente estable que el empleo de ambas estrategias de forma aleatoria. Esto sustenta la intuición de que en una población donde no existan principios claros sobre cómo coordinarse la emergencia de convenciones de comportamiento resulta bastante complicada.

La inestabilidad de nuestras convenciones normativas *G* y *M* resulta también bastante intuitiva. En situaciones de interacción similar a la de nuestra clase de juegos básica el comportamiento “generoso” por parte de ambos jugadores conduce a resultados tan desastrosos como un comportamiento “mutuamente egoísta”. En ambas situaciones, los jugadores no pueden aprovecharse de las ventajas que implica la “división del trabajo” en este tipo de entornos de coordinación asimétrica⁸⁵.

No hay que olvidar que todos estos comentarios se refieren a un modelo en el que todos los miembros de la población emplean el mismo principio. Dada nuestra clase de juegos, un análisis donde se consideraran dos poblaciones distintas (dos tipos de individuos) interactuando podría conducir a una eventual estabilidad evolutiva de todos los principios empleados (quizá excepto el principio *I*). Dicho análisis se escapa a los objetivos de la tesis.

⁸⁵ Nótese que en nuestro modelo el “egoísmo” (elegir siempre la estrategia deseada individualmente) no constituye ningún principio, ya que parece poco lógico esgrimir tal razón si lo que se pretende alcanzar es la coordinación entre las acciones de ambos jugadores.

Resultado 2: Multiplicidad de convenciones normativas

El modelo presentado muestra cómo un pequeño cambio en los parámetros del juego puede conducir a la adopción de un principio normativo distinto, como se observa en la Tabla 4. Esto implica que la estabilidad y evolución de las convenciones es *path-dependence*, es decir depende críticamente del punto de partida de una sociedad en concreto, así como de los cambios producidos en los parámetros que definen la situación de interacción. Esta influencia de la historia concreta nos lleva a afirmar que no sólo las *convenciones de comportamientos* son arbitrarias e indeterminadas por definición, sino que las *convenciones normativas* también lo son. Por tanto, deberíamos añadir las características de arbitrariedad e indeterminación a la parte relativa a las convenciones normativas de la definición propuesta en el capítulo 4.

Resultado 3: Coexistencia de convenciones normativas

Este resultado implica que una misma regularidad de comportamiento puede estar apoyada en varias convenciones normativas. Esto refuerza una de nuestras intuiciones centrales, en tanto que hace más evidente la distinción conceptual entre *convenciones normativas* y *convenciones de comportamiento*. Si un comportamiento concreto puede basarse en varios principios normativos, esto quiere decir que no está unido necesariamente a ninguno de éstos. Así, la implementación de una convención concreta no tiene por qué implicar el establecimiento de un principio único. En la tercera parte de la tesis discutiremos ampliamente este hecho a través de los resultados obtenidos en los experimentos.

5.6. Interpretación informal de los resultados

Llegados a este punto vamos a abandonar, por el momento, el lenguaje sumamente abstracto de la teoría de juegos para proporcionar algunos ejemplos que puedan ayudarnos a captar la intuición que se halla detrás de nuestra propuesta teórica.

Uno de los ejemplos que nos llevó a considerar una situación de coordinación asimétrica, así como la existencia de una multiplicidad de principios normativos, fue el problema de la coordinación entre distintos cuerpos de policía. Así, por ejemplo, en España, cuando unidades de diferentes cuerpos de policía (principalmente Policía Nacional y Guardia Civil) tienen que actuar en casos de terrorismo o crimen organizado, ambos cuerpos necesitan trabajar de forma coordinada, no sólo por una cuestión de efectividad, sino también de eficiencia en la distribución de recursos y de esfuerzos. Por tanto, ambos grupos prefieren actuar de forma conjunta. Sin embargo, para poder desarrollar su trabajo de forma conjunta necesitan ponerse de acuerdo en el principio que va a regular la interacción entre ambos, y ambos pueden tener (y de hecho tienen) distintas preferencias en relación con el principio de coordinación⁸⁶.

Al ejemplo de la coordinación entre los distintos cuerpos de policía en España⁸⁷ podemos unir otros como: la coordinación entre cuerpos de policía a escala

⁸⁶ En este caso, el único principio que la ley prescribe (en una situación de igualdad jerárquica entre los miembros de cada organización) es el principio de prioridad en la llegada a un determinado caso. El principio de prioridad (el conocido “*first come, first serve*” anglosajón) es un principio normativo que contribuye a la resolución de problemas de coordinación en situaciones asimétricas, es decir, situaciones en las que los actores tienen preferencias diversas sobre el resultado de la coordinación.

⁸⁷ El problema de la coordinación entre cuerpos de policía está despertando un gran interés entre los estudiosos de la organización y el comportamiento de los cuerpos de policía. Ver, por ejemplo, el monográfico sobre el mismo tema coordinado por Linde y Pérez (2003). En un capítulo de este monográfico, Martín (2003: 223) define la coordinación policial como “necesaria en la ejecución o prestación de servicios en materias comunes o de prestación indiferenciada, buscando armonizar esfuerzos y actividades para conseguir mayor eficiencia y evitar solapamientos y duplicidades.”

internacional; la coordinación entre ejércitos de diferentes países a la hora de intervenir en situaciones de conflicto; o la coordinación entre diversos actores en situaciones de emergencia, como es el caso de la interacción entre cuerpos de policías, bomberos, ambulancias, etc. a través del Sistema de Emergencias 112⁸⁸. No resulta difícil pensar en todos estos ejemplos como situaciones de coordinación asimétrica donde, aún compartiendo un objetivo común, los actores tienen preferencias diversas en relación con la forma de alcanzar dicho objetivo y donde el uso de principios generales de coordinación puede contribuir a garantizar la coordinación efectiva de las acciones.

Hasta aquí hemos hecho referencia a algunos casos reales donde puede merecer la pena prestarle atención tanto a la emergencia espontánea, como al establecimiento deliberado de principios similares a los que hemos estudiado en este capítulo. A continuación, a través de un ejemplo hipotético sobre la interacción entre organizaciones ecologistas, nos intentaremos acercar al significado preciso de algunos de estos principios.

Pensemos de nuevo en nuestra clase de juegos básica (Figura 4) y supongamos que la interacción entre ambos jugadores es repetida. Dicho juego podría representar, por ejemplo, la interacción entre dos organizaciones ecologistas acostumbradas a cooperar conjuntamente para exigir al gobierno que cumpla sus compromisos medioambientales. Sin embargo, el hecho de que ambas estén dispuestas a cooperar no significa que hayan resuelto todos sus problemas de acción colectiva. Por ejemplo, una organización (jugador A) puede preferir la estrategia X_A , consistente en la realización de acciones puntuales con una gran repercusión mediática, mientras que la otra (jugador

⁸⁸ Por ejemplo, entre los objetivos del Sistema de Emergencias 112 en Andalucía encontramos el siguiente: “Favorecer la coordinación entre los efectivos que intervienen en las emergencias” (Junta de Andalucía, 2003).

B) puede preferir la estrategia Y_B , consistente en desarrollar un programa continuado de acciones destinadas a influir en la gestión gubernamental. Además, cada acción puede ser más efectiva, y de este modo reportar unos mayores beneficios conjuntos, en un momento concreto de la interacción, y las organizaciones pueden estar dispuestas a optar por aquella acción que en cada momento consideran más adecuada (principio S). En nuestro ejemplo, elegirían el equilibrio (X_A, X_B) cuando $a > c$ y el equilibrio (Y_A, Y_B) cuando $a < c$.

Otro principio que podría regular la interacción entre ambas organizaciones es el de alternarse en la elaboración de las campañas para presionar el gobierno, desarrollando alternativamente acciones de choque y acciones más sutiles pero al mismo tiempo más constantes (principio A). Así, en la ronda t ambas organizaciones elegirían la estrategia Y , en la ronda $t + 1$ elegirían X , en $t + 2$ de nuevo Y , y así sucesivamente.

Los principios que pueden regular la interacción entre ambas organizaciones son múltiples, aunque limitados. Otro podría conducir a una situación de desigualdad, en la que una organización elige siempre su forma de actuar preferida y la otra se suma a la misma (esta segunda estaría utilizando el principio G o el principio M). Por último, también podemos pensar en principios que eliminen determinadas estrategias debido a los costes que éstas suponen. Así, puede que una de las organizaciones (jugador A) decidiera no elegir nunca la estrategia X_A (acción mediática) ya que le puede generar un coste d que no está dispuesta a asumir. Ante tal escenario la otra organización (jugador B) se vería obligada a descartar también dicha estrategia si quiere actuar coordinadamente con la primera, produciéndose, de este modo, una coordinación

duradera en el equilibrio (Y_A, Y_B) (programa continuado de acciones). En este último caso ambas organizaciones estarían empleando el principio L para coordinarse⁸⁹.

Los ejemplos presentados hacen referencia a situaciones micro-sociales donde los principios descritos pueden contribuir a la coordinación entre las partes. Además, también podemos reflexionar sobre la utilidad de los mismos en contextos macro-sociales. En un artículo reciente, Berninghaus, Güth y Kliemt (2005) han señalado la pertinencia de la comprensión por parte de los encargados de diseñar las políticas económicas de los mecanismos que garantizan la estabilidad de las convenciones, como se puede observar en el siguiente texto:

Bajo nuestro punto de vista, la explicación precedente de los juegos de coordinación con múltiples equilibrios estrictos sugiere que “la gestión de las relaciones públicas”, “las interpelaciones al público”, “la persuasión moral” y otras actuaciones similares pueden ser entendidas perfectamente como medidas respetables de política, o incluso terapia, económica. Mientras que normalmente la política económica está bien dirigida a aspectos institucionales (en el sentido de perseguir proyectos de diseño de mecanismos o diseños institucionales) o bien a cambiar los parámetros [del juego] (políticas monetaria y fiscal o de gasto público), la “terapia o política de selección de equilibrios” no persigue cambiar las reglas, sino (re)coordinar las expectativas de tal modo que la sociedad cambie de un equilibrio estricto preferido a otro. [...] Ya que la evolución y la adaptación del comportamiento pueden conducir a convenciones ineficientes, queda espacio para la “política o terapia de selección de equilibrios.” [...] Si la “persuasión moral” puede desempeñar un papel en las políticas más allá de las intervenciones directas, sean éstas institucionales o de otro tipo, es necesario analizar ésta más cuidadosamente y debería ser incorporada eventualmente en la explicación científico-social de la política económica. El objetivo no debe ser “el economista como un predicador”, sino más bien una especie de “economía del predicar.”

Por tanto, otra de las ventajas de analizar los principios que eventualmente pueden conducir a una sociedad a coordinarse en una convención concreta es que éstos pueden ser empleados por los gobiernos para inducir una determinada dinámica de comportamiento en la población. Ni que decir tiene que este razonamiento también

⁸⁹ Desde luego, aquí estamos forzando el argumento con el objetivo de acercarnos intuitivamente a lo que significan los principios como mecanismos de coordinación. El conjunto de principios que hemos introducido en este capítulo responden a un planteamiento teórico, incluso matemático (como en el caso del principio R), que no tiene por qué trasladarse directamente a situaciones empíricas.

puede ser aplicado a organizaciones intermedias, tales como empresas, organizaciones sociales, etc⁹⁰.

5.7. Conclusiones

El propósito general de esta segunda parte de la tesis ha sido desarrollar un marco teórico general que responda, en primer lugar, a la “doble naturaleza” teórica de las convenciones sociales presentada en la primera parte y que sirva de base, en segundo lugar, para la investigación empírico-experimental llevada a cabo en la tercera parte de la misma. Para ello, tras una breve justificación acerca de nuestra propuesta de formalización, hemos elaborado un modelo formal en el que se pueden distinguir dos partes, a saber: (1) una clase de juegos básica, que permite derivar posibles *convenciones de comportamiento* que pueden emerger de los problemas de coordinación contenidos en dicha clase de juegos y (2) una serie de *principios o convenciones normativas* que prescriben comportamientos concretos en los juegos contenidos en dicha clase de juegos básica.

⁹⁰ El análisis acerca de cómo las expectativas colectivas pueden ser (re)coordinadas implícita o explícitamente no se circunscribe sólo al ejemplo económico en la actualidad. Un precursor de dicho análisis fue John Maynard Keynes (1973) quien a través de su conocido ejemplo del “concurso de belleza” explicó cómo se producen los macro procesos de coordinación y re-coordinación de expectativas en los mercados de capitales. En el ámbito de las teorías acerca del cumplimiento de la ley, McAdams y Nadler (2005) también ha recurrido al análisis de cómo las expectativas de los ciudadanos pueden ser intencionalmente coordinadas para propiciar un mayor cumplimiento de la ley. Según la “teoría expresiva” del cumplimiento desarrollada por este autor, la ley puede tener una influencia sobre el comportamiento independientemente de las sanciones que la acompañan. Dicha influencia no tiene por qué recurrir a argumentos de carácter normativo. Específicamente, la ley proporciona un “punto focal” alrededor del cual los individuos pueden coordinar su comportamiento. Cuando los individuos tienen un interés común en coordinarse, como ocurre frecuentemente, una regla legal puede guiar el comportamiento simplemente ejerciendo una influencia sobre las expectativas acerca de cómo los demás se comportarán. Un último intento de explicar multitud fenómenos sociales a través del argumento de la coordinación de las expectativas es el libro de Chwe (2001) citado en la primera parte de la tesis. En este caso la coordinación de las expectativas de un buen número de personas le sirve al autor para explicar fenómenos tan dispares como acciones colectivas de carácter revolucionario o pautas masivas de consumo.

Una vez definido el marco teórico general, hemos empleado el concepto de *estrategia evolutivamente estable* (EEE) para analizar la estabilidad evolutiva de los distintos principios propuestos. Dicho análisis nos ha conducido a tres resultados: (1) la posibilidad de descartar ciertos principios por no ser evolutivamente estables, (2) la constatación de que existen múltiples principios que son evolutivamente estables y (3) la demostración de la existencia teórica de “multiplicidad genérica” de convenciones normativas, es decir la coexistencia de convenciones normativas. Como principal aportación a los argumentos desarrollados en la primera parte, debemos destacar la constatación del carácter arbitrario e indeterminado de las convenciones normativas.

Tras el análisis evolutivo, hemos realizado algunas reflexiones de carácter más general dirigidas a una mayor comprensión de nuestro marco teórico. Nuestra propuesta es que el mismo puede ser aplicado tanto a situaciones de interacción micro-sociales, por ejemplo a la solución de problemas de coordinación entre diversas organizaciones, como a situaciones macro-sociales, en el sentido de la propuesta de política económica basada en la “persuasión moral” de Berninghaus, Güth y Kliemt (2005).

Con este conjunto de reflexiones llegamos al ecuador de la tesis. La tercera parte de la misma estará dirigida por completo al estudio de cómo los principios propuestos son realmente utilizados por los individuos para coordinarse.

PARTE III

APROXIMACIÓN EXPERIMENTAL A LA NOCIÓN DE CONVENCION SOCIAL

En el camino recorrido hasta el momento, esta tesis ha ido de una aproximación conceptual general, bastante amplia por necesidad, a una operacionalización muy concreta de la noción de convención social. En la segunda parte, al intentar hacer manejable nuestra propuesta teórica integrada del concepto de convención social, no sólo hemos restringido considerablemente el ámbito teórico de la tesis, sino que, haciendo uso de la teoría de juegos evolutiva, hemos proporcionado una serie de resultados teóricos que nos han ayudado a completar y refinar la propuesta conceptual realizada en la parte primera. Pero los resultados teóricos no dejan de ser construcciones ideales de lo que podría suceder si un conjunto de supuestos, más o menos plausibles, se cumplen. A pesar de que en el capítulo anterior hemos defendido el carácter plausible de nuestros supuestos de partida, esto no quiere decir que por ello tengamos que renunciar a contrastar nuestras proposiciones teóricas con los resultados obtenidos a partir de la aplicación rigurosa de técnicas de recogida y análisis de datos sociales. El análisis empírico de las afirmaciones teóricas realizadas en la primera y, sobre todo, en la segunda parte de la tesis se constituye en el objetivo principal de esta tercera parte.

Sin embargo, como ya se había afirmado en la introducción, ésta no es una tesis empírica, sino fundamentalmente teórica. Por ello, los resultados empíricos que van a ser presentados en esta tercera parte serán utilizados para revisar las proposiciones

teóricas introducidas anteriormente y no para describir nuevas regularidades en el comportamiento social. Dado este objetivo, hemos elegido el método experimental como el más adecuado para calibrar la validez del modelo teórico descrito en la segunda parte. La utilización de experimentos de laboratorio, de este modo, nos permitirá comprobar hasta qué punto los argumentos desarrollados hasta el momento tienen que ver con el comportamiento de los sujetos experimentales en una situación controlada, o al menos una en la que se intentan controlar los factores ajenos a la relación causal que se pretende analizar.

El método experimental no es muy frecuente en la investigación sociológica actual. En este sentido, resulta sorprendente como una ciencia que dedica tantas páginas al problema de los métodos y de las metodologías, enfrascada en el eterno debate entre “lo cuantitativo” y “lo cualitativo”, apenas se ha hecho eco del “renacimiento” de la metodología experimental que actualmente se está dando en otras disciplinas colindantes, como la ciencia política y, sobre todo, la economía. Es por ello que antes de dedicar un capítulo (Capítulo 7) a exponer los resultados de las distintas sesiones experimentales llevadas a cabo para poner a prueba nuestras hipótesis teóricas, dedicaremos un capítulo (Capítulo 6) a discutir la validez de la metodología experimental, especialmente de lo que ha venido en llamarse “experimentos económicos”, en el seno de la sociología.

CAPÍTULO 6

LA VALIDEZ DE LOS EXPERIMENTOS ECONÓMICOS EN SOCIOLOGÍA

6.1. Introducción

El tipo de diseño experimental que va a ser empleado en esta tesis se ajusta a lo que Ernest Greenwood denominaba “experimento puro”⁹¹, un tipo de diseño que supone el “ideal de método experimental, el modelo hacia el que todos los demás tipos deberían aspirar” (Greenwood, 1978: 44). Este capítulo sexto está dedicado a una discusión metodológica en torno a dicho método, sus potencialidades, así como los problemas que entraña su aplicación. Sin embargo, antes de proceder a detenernos sobre el tipo “experimento puro” en particular, nos parece pertinente una presentación del método experimental de forma genérica que nos sirva para situar en un contexto más amplio el tipo de diseño que va a ser empleado⁹².

El método experimental no ha ocupado, ni ocupa, un lugar privilegiado en el arsenal metodológico propio del sociólogo. Su aplicación en esta disciplina queda lejos no sólo de su uso en psicología – la ciencia humana experimental por excelencia – sino incluso lejos de su empleo actual en otras ciencias que estudian el comportamiento humano, como la economía o la ciencia política. Por tanto, no parece descabellado aprovechar la discusión metodológica de una mayor raigambre presente en estas otras

⁹¹ Otras etiquetas, que pueden resultar más intuitivas, han sido empleadas para denominar este tipo de diseños experimentales, tales como: “método de laboratorio”, “experimento directo”, “experimento controlado”, “experimento científico”, “experimento verdadero” o “experimento de laboratorio”.

⁹² En este capítulo no abordaremos las características particulares del diseño y el procedimiento que van a ser utilizados en el capítulo siguiente. En dicho capítulo, y antes de presentar los resultados, se realizará una descripción exhaustiva de los experimentos llevados a cabo, realizando las necesarias aclaraciones metodológicas. Por el momento sólo nos ocuparemos de realizar una presentación del método experimental, en general, y del tipo de “experimento puro” o “experimento de laboratorio”, en particular.

disciplinas para defender el uso de dicho método en sociología. Sin embargo, aunque haremos uso de ella, nuestro propósito es anclar nuestra reflexión en el debate en torno al método experimental que se ha producido en el seno de la propia sociología.

Como veremos, la discusión sociológica a propósito de este método ha girado en torno a dos grandes temas. El primero hace referencia a la delimitación conceptual de aquello a lo que podemos denominar experimento, así como en qué medida el método experimental difiere de otros métodos empleados por la disciplina sociológica. El segundo tema apunta directamente al problema de la validez general de dicho método y a la pertinencia de su aplicación al estudio de fenómenos sociales. Ambos aspectos serán ampliamente abordados en este capítulo.

Empezando por el primero de los temas planteados, y antes de proceder a proporcionar una definición exacta de lo que aquí entendemos por método experimental, nos preguntamos: ¿qué características básicas se encuentran presentes en la mayor parte de las investigaciones que se definen a sí mismas como experimentales? De acuerdo con Greenwood (1978: 29)⁹³, que a su vez se apoya en la seminal obra de John Stuart Mill (1891), el método experimental queda definido por tres características

⁹³ Puede sorprender el presente uso de una obra como la de Greenwood (1978) – cuya primera edición data de 1945 – para aproximarnos a las características básicas del método experimental. De hecho, a lo largo del capítulo se utilizan otra serie de referencias, tanto sociológicas como procedentes de otras ciencias sociales, bastante más actuales. Sin embargo, la obra de Greenwood cumple varios requisitos difíciles de encontrar en obras más recientes. En primer lugar, es uno de las pocas monografías – quizá la única junto con la de Chapin (1947) – dedicadas íntegramente a una discusión metodológica de la aplicación del método experimental en sociología. En segundo lugar, como Oakley (1998) ha puesto de manifiesto, se trata de una de las referencias básicas utilizadas por los sociólogos que, durante las décadas de los sesenta y setenta del pasado siglo, utilizaron el método experimental como herramienta privilegiada para el análisis sociológico, sobre todo en lo que se refiere a programas de evaluación de toda índole. En tercer lugar, el libro está escrito en un momento en el que existía un amplio debate acerca del método experimental en sociología, como pone de manifiesto Paul Lazarsfeld en la introducción del propio libro. Por último, Greenwood no se compromete con la utilización de un diseño experimental u otro, sino que pretende evaluar las luces y sombras de cada tipo particular de diseño, lo que nos puede ser muy útil para situar el tipo de diseño experimental concreto que emplearemos en esta tesis.

básicas, a saber: (1) la presencia de una hipótesis causal, (2) que es estudiada a través de una serie de situaciones de contraste; (3) estas situaciones de contraste han sido controladas. Nuestro planteamiento de partida asume que para que podamos añadir el adjetivo *experimental* a una investigación, estas tres características deben estar en algún grado presentes. Este hecho, a pesar de acotar nuestra definición de lo que es un experimento, todavía permite imaginar diseños de investigación bastante diferentes atendiendo a cómo pongamos en práctica cada una de las características mencionadas. Así, el componente de causalidad puede ser entendido en el sentido cronológico, como una *relación temporal asimétrica*, donde la causa precede al efecto o puede ser entendido como la regularidad en la presencia conjunta de dos hechos que no implica asimetría temporal alguna, como ocurre cuando hacemos referencia al carácter *reflexivo* del comportamiento en sociedad. Por otra parte, las situaciones de contraste pueden proceder de la observación de un fenómeno natural o pueden haber sido creadas por el propio investigador, como en el caso de los experimentos de laboratorio que describiremos a continuación. Por último, también el control ejercido sobre dichas situaciones de contraste puede ser llevado a cabo de muy diversas formas. Por ejemplo, el control puede ser ejercido sobre el desempeño del investigador, como en las situaciones en las que dos o más investigadores observan y codifican el mismo fenómeno independientemente para garantizar un mayor grado de objetividad sobre lo observado. Pero el control también puede ejercerse sobre el propio objeto de estudio, por ejemplo utilizando grupos de contraste que sean similares en todas las características imaginables menos una, que es la que constituye el objeto de investigación. Esta heterogeneidad de los diseños que podemos clasificar como experimentales nos llevará en el próximo apartado a contraponer varios tipos de experimentos antes de adentrarnos en la descripción de nuestro diseño en particular.

Pero antes de esto, y a la luz de las características introducidas más arriba, proporcionaremos algunas definiciones específicas del método experimental.

A partir de una abstracción de las características particulares de cada tipo de diseño experimental, Greenwood (1978) define un experimento del siguiente modo:

Un experimento es la comprobación de una hipótesis que busca relacionar dos factores en una relación causal a través del estudio de situaciones de contraste en las que todos los factores, excepto aquel que interesa [estudiar], son controlados; este factor es bien la causa hipotética o bien el efecto hipotético. (p. 28)

Como se observa, esta definición sólo supone una combinación de las tres características introducidas anteriormente. La misma todavía no excluye ninguno de los tipos de diseños experimentales que presentaremos en el apartado siguiente. Nótese que la definición no incluye ninguna referencia a que el experimento conlleve una manipulación explícita de las variables por parte del investigador. Este elemento de manipulación, sin embargo, constituye la característica fundamental del método experimental para los investigadores sociales contemporáneos. Sirva como ejemplo la definición propuesta por Levin (1999):

[El método experimental] es el método que el investigador elige cuando está principalmente interesado en determinar las causas y los efectos. El método experimental está caracterizado por la manipulación de una o más variables denominadas *variables independientes* y el control de todas las demás. *Manipulación* quiere decir crear o seleccionar niveles discretos de una variable y comparar las respuestas obtenidas en los distintos niveles. *Control* significa mantener una *variable superflua* (una variable distinta a la variable independiente que puede influenciar los resultados de un experimento) constante en los distintos niveles de la variable independiente. Si el investigador consigue controlar todas las variables superfluas, entonces él o ella puede determinar que los cambios en el comportamiento (la *variable*, o variables, *dependiente*) deben haber sido causados por la manipulación de la variable independiente. (p. 5)

Incluso hay quien va más allá y afirma que es precisamente la manipulación de la variable independiente lo que separa el método experimental de los métodos no-

experimentales en la ciencia social, como puede leerse en la siguiente cita correspondiente a la introducción que Michael Lewis-Beck realiza al trabajo de Brown y Melamed (1990):

Los métodos de la ciencia social pueden ser clasificados en dos [tipos], no-experimental y experimental. Un principio clave del trabajo experimental es la *manipulación* de la variable de tratamiento (*X*), seguido de la observación de una variable de respuesta (*Y*). (Brown y Melamed, 1990: v)

Estas citas ponen de relieve una primera diferencia entre lo que denominaremos como *experimentos de orientación empírica* y *experimentos de orientación teórica*. En los siguientes apartados de este capítulo profundizaremos en otros elementos adicionales que nos permiten establecer sendas diferencias entre ambas concepciones⁹⁴.

6.2. Dos concepciones distintas acerca de la experimentación

Resulta obvio afirmar que no todos los experimentos son iguales ni sirven para lo mismo. En la introducción de este capítulo hemos señalado tres características comunes a todo experimento, pero a partir de ahí existen múltiples diseños posibles. En este apartado vamos a tratar de responder una pregunta básica: ¿para qué le sirve al

⁹⁴ No constituye un objetivo central de este capítulo indagar sobre las diferentes etapas en la aplicación del método experimental en sociología. A pesar de ello, por conveniencia, utilizaremos las etiquetas de *experimentos de orientación empírica* y *experimentos de orientación teórica* para referirnos a dos momentos diferentes en el uso dado por los sociólogos a dicho método. Las características de ambos momentos serán desveladas a lo largo de este capítulo, así que no abundaremos más en ellas en este momento. En cuanto a su localización temporal diremos que la *orientación empírica* cubriría un período de unos sesenta años, desde comienzos de los años veinte hasta finales de los años setenta. Dos obras claves en este extenso período son los ensayos metodológicos de Chapin (1947 [1931]) y de Greenwood (1978 [1945]). Según el relato de Oakley (1998), en la década de los ochenta se produce un descrédito de lo que hemos denominado *experimentos de orientación empírica*. La denominada *orientación teórica* es bastante más reciente y, a grandes rasgos, supone la traslación a la sociología de la “revolución del comportamiento” que ha tenido lugar en los últimos veinte años en disciplinas como la economía, en primer lugar, y la ciencia política, en segundo. Aunque ya encontramos trabajos a lo largo de los años ochenta y noventa, una fecha clave es, sin duda, la concesión del premio Nobel de economía al economista experimental Vernon Smith y al psicólogo Daniel Kahneman en el año 2002 por haber aplicado con éxito el método experimental al estudio de fenómenos económicos.

sociólogo el método experimental? Algunos sociólogos dirían que el método experimental proporciona al investigador una forma de descubrir nuevos datos, así como las relaciones que se pueden establecer entre éstos, presentes en la realidad social. En este caso, ante una situación social existente, la aplicación *ex post* del método de investigación permite al investigador reconstruir tales relaciones y en última instancia atribuirles *causalidad*, es decir, determinar las causas y los efectos correspondientes a la situación en cuestión. Otros podrían responder que el método experimental sirve para contrastar las hipótesis teóricas elaboradas *ex ante* por el científico social. En este caso, las hipótesis teóricas, independientemente de cuál sea su procedencia⁹⁵, son sometidas a un riguroso estudio, preferentemente en un entorno que permita la manipulación explícita de las variables que se hallan en la base de dichas hipótesis. Denominaremos la primera de estas respuestas intuitivas *orientación empírica* del método experimental y la segunda *orientación teórica*. La diferencia entre ambas concepciones no es trivial y necesita ser aclarada antes de que procedamos a describir el tipo de diseño experimental empleado en esta tesis.

6.2.1. Experimentos de orientación empírica

La experimentación de *orientación empírica* vivió su momento álgido entre las décadas de los sesenta y de los ochenta en Estados Unidos, aunque se pueden encontrar ejemplos con anterioridad a la Segunda Guerra Mundial⁹⁶. La Tabla 5 muestra el abanico de temas, objetivos, diseños de investigación y resultados obtenidos por distintos sociólogos que utilizaron el método experimental a lo largo de estas décadas.

⁹⁵ Estas hipótesis teóricas pueden proceder tanto de investigaciones empíricas precedentes, como de argumentaciones lógico-deductivas apoyadas en teorías cuyo alcance y condiciones de posibilidad estén bien definidos, como es el caso de la teoría de la elección racional.

⁹⁶ Ver Chapin (1947) para un repaso de experimentos de *orientación empírica* realizados en las primeras décadas del siglo XX.

Como se observa, este método fue utilizado para evaluar la eficacia de numerosos programas de intervención social; desde programas de ayuda a la incorporación al mercado de trabajo, hasta programas relacionados con la reinserción social de ex presidiarios. La metodología empleada por los mismos seguía siempre un patrón parecido, consistente en la comparación de diferentes grupos de individuos a los que, de forma aleatoria, se les sometía a algún tipo de intervención social. Así, por ejemplo, se comparaban los resultados obtenidos tras un período determinado de tiempo de ex presidiarios que habían participado en un programa de reinserción social con otros que no habían tenido dicha oportunidad. El objetivo era siempre determinar el efecto que cada programa de intervención tenía sobre una determinada población.

El elemento central de este tipo de experimentos es el control de las variables superfluas. En otras palabras, lo que importa es determinar, con un alto grado de confianza, el efecto producido por la introducción de una determinada variable independiente (la intervención) sobre la variable dependiente (sea ésta la participación en el mercado de trabajo, un cambio en las condiciones de vida, etc.). Dicho control es alcanzado de forma privilegiada a través de dos mecanismos, a saber: la utilización de grupos de control, que permiten aislar el efecto producido por la variable independiente introducida; y la asignación aleatoria de los sujetos experimentales a los grupos donde se produce la intervención y a los grupos de control. Este último mecanismo fue utilizado profusamente ante la imposibilidad de controlar todas las variables superfluas que podrían distorsionar la medición de los resultados de la intervención en cuestión. Desde luego, la búsqueda de un mayor control de las variables que intervienen en el

experimento llevó a estos científicos sociales a emplear diseños experimentales bastante sofisticados⁹⁷.

En este capítulo no nos interesan tanto los detalles de estos diseños experimentales de *orientación empírica*, como las diferencias de éstos con los experimentos de *orientación teórica* que emplearemos en esta tesis. De los ejemplos señalados anteriormente se deriva que los primeros pretenden responder a una cuestión eminentemente empírica, por ejemplo: ¿hasta qué punto los ex presidiarios que participan en un programa de reinserción social tienen más posibilidades de alcanzar un empleo estable fuera de la cárcel? La respuesta, positiva o negativa, a esta pregunta nos conducirá a la posibilidad de establecer una regularidad empírica del tipo: aquellos ex presidiarios que participan en un programa de reinserción social consiguen, en media, mejores empleos. Chapin (1947) hablaba de experimentos *ex post facto* en este caso. En resumen, se trata de experimentos que emplean muestras representativas, que investigan una posible regularidad empírica en la sociedad y cuya manipulación de la realidad (si es que ésta existe) se limita a controlar y diferenciar mediante técnicas estadísticas la población que se ha visto expuesta a algún tipo de intervención de aquella que no lo ha sido⁹⁸.

⁹⁷ Ver, por ejemplo, el trabajo de Campbell y Stanley (1963) o el capítulo sobre diseños experimentales del libro de Moser y Kalton (1971).

⁹⁸ A partir de la década de los ochenta este tipo de diseño experimental de *orientación empírica* cayó en desuso en los Estados Unidos. Oakley (1998: 1241) ofrece dos razones que influyeron en esta decadencia. Una primera causa fue el desencanto con la aparente ineficacia (a veces los efectos aparentemente contraproducentes) de las intervenciones, tal y como fue observado en algunas de las evaluaciones. Una segunda causa radicaba en que los políticos se mostraban a menudo impacientes con el largo período de tiempo que necesitaban las evaluaciones para aportar resultados concluyentes.

Tabla 5: Ejemplos de experimentos controlados sobre programas sociales llevadas a cabo en los Estados Unidos entre 1968-1991

Experimento	Año	Objetivo	Diseño	Resultado evaluado
Experimentos sobre el sistema de garantía de rentas				
<i>New Jersey-Pennsylvania negative income tax experiment</i>	1968-72	Estudiar los efectos del impuesto negativo sobre la renta sobre los incentivos laborales	Asignación aleatoria de 1216 familias con ingresos bajos a 8 grupos de intervenciones y un grupo de control.	Participación en el mercado de trabajo; gasto de consumo; salud y comportamiento familiar; asistencia a la escuela.
<i>Rural negative income tax experiment</i>	1970-72	Replicar el experimento anterior en áreas rurales pobres con familias desestructuradas, tanto en el caso de un cabeza de familia hombre, como mujer	Asignación aleatoria estratificada de 809 familias con ingresos bajos a 5 grupos de intervención y 1 grupo de control.	Participación en el mercado de trabajo; gasto de consumo; salud y comportamiento familiar; asistencia a la escuela.
<i>Gary income maintenance experiment</i>	1971-74	Estudiar los efectos en la participación en el mercado de trabajo y otros comportamientos familiares de diferentes niveles y formas de complementos de ingresos, ayudas para guarderías e información y servicios de ayudas a la búsqueda de empleo.	Asignación aleatoria estratificada de 1799 hogares con ingresos bajos y sólo un progenitor.	Participación en el mercado de trabajo; gasto de consumo; salud y comportamiento familiar; asistencia a la escuela; actitudes sociales y psicológicas.
<i>Denver-Settle income maintenance experiments</i>	1970-91	Estudiar los efectos en la participación en el mercado de trabajo y otros comportamientos familiares de diferentes niveles de complementos de ingresos, asesoramiento laboral y ayudas para formación.	Asignación aleatoria estratificada de 2042 familias asignadas a 84 "celdas" experimentales con diferentes combinaciones de niveles de apoyo, impuestos fiscales, etc., y un grupo de control.	Participación en el mercado de trabajo; gasto de consumo; salud y comportamiento familiar; asistencia a la escuela.
Ayudas familiares				
<i>Experimental housing allowance program (demand experiment)</i>	1978-80	Estudiar los efectos en el comportamiento del hogar de diferentes formas de subsidios familiares y estimar el coste de su efectividad.	Asignación aleatoria estratificada de 2241 hogares con ingresos bajos a 17 grupos de intervención con diferentes ayudas familiares y dos grupos de control.	Calidad del hogar; consumo del hogar; movilidad.
Programas en apoyo de los trabajadores				
<i>Supported work program</i>	1975-78	Estudiar los efectos y los costes de ambientes de apoyo al trabajo para trabajadores discapacitados.	Asignación aleatoria de 6616 trabajadores discapacitados a una intervención y un grupo de control.	Participación en el mercado de trabajo; horas trabajadas; ingresos totales.

<i>Texas worker adjustment program</i>	1984-5	Estudiar los efectos y el coste de la combinación de la asistencia en la búsqueda de empleo y la formación en competencias profesionales para trabajadores desplazados.	Asignación aleatoria de 2259 individuos difíciles de contratar a través de una tabla de números aleatorios a 2 grupos de intervención y 1 de control en un lugar y a una intervención y 1 grupo de control en 2 lugares.	Ingresos; desempleo; subsidios de desempleo.
Experimentos penales				
<i>Living insurance for ex-prisoners (LIFE)</i>	1971-74	Estudiar los efectos en las reincidencias y la participación en la población activa a partir de diferentes niveles del pago posterior a la puesta en libertad, así como esquemas de asistencia.	Asignación aleatoria estratificada de 432 prisioneros puestos en libertad a 3 grupos de intervención (sólo pago, asesoramiento y colocación, ambos combinados) y 1 grupo de control.	Arrestos y condenas por tipo de delito; participación en el mercado de trabajo; salud y condiciones de vida.
<i>Transitional and research Project (TARP)</i>	1975-77	Estudiar los efectos en las reincidencias y la participación en la población activa a partir de diferentes niveles del pago posterior a la puesta en libertad, así como esquemas de asistencia.	Asignación aleatoria estratificada de 3982 prisioneros puestos en libertad a 4 grupos de intervención mediante combinaciones de diferentes períodos de pago y tasas de impuestos (3 grupos) y servicios de colocación laboral (1 grupo), más 2 grupos de control.	Arrestos y condenas por tipo de delito; participación en el mercado de trabajo; salud y condiciones de vida.

Fuente: traducción y adaptación de la tabla-resumen elaborada por Oakley (1998: 121)⁹⁹.

⁹⁹ Con respecto a la tabla original presentada por Oakley, hemos respetado el inglés original de los nombres de los experimentos. Además hemos eliminado las referencias exactas de cada uno de los experimentos ya que el objetivo aquí sólo es mostrar, aprovechando el trabajo de Oakley, las características básicas del tipo de experimentos de *orientación empírica* a los que nos hemos referido más arriba.

6.2.2. Experimentos de orientación teórica

Los experimentos de *orientación teórica*, por el contrario, no analizan directamente una situación social, sino que buscan contrastar la validez de una determinada teoría, hipótesis teórica o mecanismo explicativo. Al tratarse de un método importado de la economía, en general esta teoría suele ser la teoría de la elección racional (TER); aunque el desarrollo experimentado por este tipo de método experimental de *orientación teórica* hace que ya se cuestionen las propias teorías que han surgido como respuestas a las limitaciones de la TER. Entre los trabajos pioneros de esta corriente se encuentran los trabajos en el campo de la teoría de la decisión llevados a cabo a lo largo de la década de los cincuenta, cuyo principal exponente es el trabajo de Maurice Allais (1953), aunque, sin lugar a dudas, el pistoletazo de salida de la revisión experimental de los principales supuestos de la TER fue dado por la publicación del trabajo Kahneman y Tversky (1979)¹⁰⁰. Pero antes de detenernos en las características fundamentales de este tipo de diseños experimentales, volvamos a la relación entre realidad, teoría y experimento a la que nos referíamos con anterioridad. Kanazawa (1999) insiste en la *orientación teórica* de los experimentos de laboratorio – los experimentos puros de los que hablaba Greenwood – del siguiente modo:

uno *nunca* generaliza los hallazgos empíricos de los experimentos de laboratorio a una población mayor en la que está interesado. Como han argumentado muchos en el pasado [...], el propósito de llevar a cabo experimentos de laboratorio en cambio es *someter a prueba* una teoría abstracta y general *cuyas condiciones de posibilidad están enunciadas claramente*, y, una vez contrastada ésta, es la teoría, no los datos experimentales, la que es generalizada a situaciones naturales de interés que se ajustan a sus condiciones de posibilidad. (p. 452)

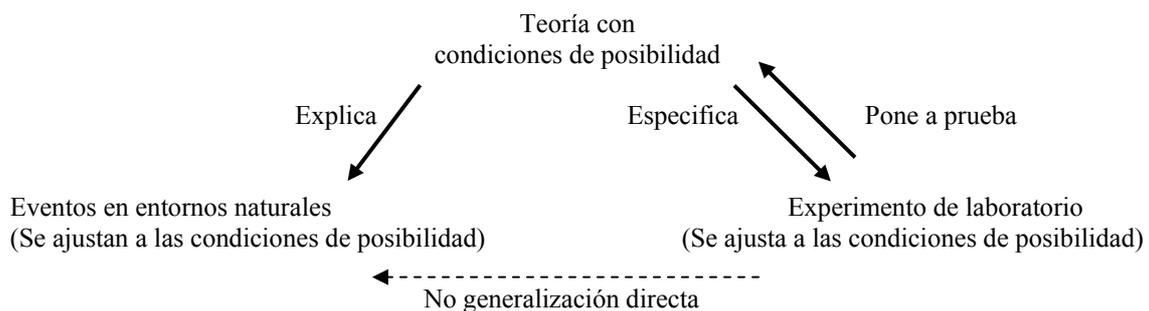
Más adelante afirma:

¹⁰⁰ Para un repaso de algunos de los experimentos simples que han contribuido a poner en duda la validez descriptiva de la TER, incluyendo el propio trabajo de Allais, ver Aguiar (2004).

un experimento de laboratorio cuidadosamente diseñado para ajustarse a todas las condiciones de posibilidad de la teoría no es más que un caso empírico de los que la teoría explica. Sin embargo, tampoco es menos.

Por tanto, al contrario que los experimentos de *orientación empírica*, en los que un caso empírico real era analizado, los experimentos de orientación teórica se proponen reconstruir en el laboratorio una determinada teoría y, de este modo, estudiar el efecto producido por la manipulación explícita de los distintos componentes de la misma. La Figura 7 nos puede ayudar a visualizar los argumentos introducidos por Kanazawa.

Figura 7: Diseño apropiado del experimento



Fuente: traducido de Kanazawa (1999: 452).

El primer paso para la realización de un experimento de laboratorio (de *orientación teórica*) es disponer de una teoría cuyo alcance esté bien definido. Una vez que disponemos de ésta nuestro objetivo será intentar que la especificación que se realiza de la misma en el experimento se ajuste en la mayor medida posible al alcance y condiciones de posibilidad de dicha teoría. Este ajuste es lo que denominaremos *validez interna* del experimento, que se preocupa de “la cuestión de si mediante el experimento se obtiene una medida verdadera del efecto de la variable independiente [en la variable

dependiente]” (Moser y Kalton: 1971: 220). Si se diseña un experimento para poner a prueba una teoría preexistente, la validez *interna* de los resultados obtenidos, ya vayan éstos en la dirección de la confirmación (parcial) o refutación (parcial) de la teoría, dependerá de lo capaces que seamos de recrear los distintos aspectos de la teoría en la situación experimental concreta. Este argumento quizá pueda quedar más claro con un ejemplo concreto.

Pensemos en la conocida proposición teórica de Mancur Olson con la que este economista norteamericano predecía la imposibilidad de la acción colectiva en grupos lo suficientemente grandes y en ausencia de incentivos selectivos de carácter individual. En una afirmación que ya se ha hecho célebre, Olson afirma que “los individuos racionales y auto-interesados no actuarán para alcanzar sus intereses comunes” (Olson, 1965: 2). Algunos años más tarde, la estructura estratégica de la lógica de la acción colectiva fue representada formalmente a través del juego del dilema del prisionero de n personas (Hardin, 1971; 1982). Una generalización del juego del dilema del prisionero a n personas nos conduce a un juego de bien público (Kollock, 1998). Tanto el juego del dilema del prisionero como el juego de bien público han sido las dos situaciones estratégicas que han recibido una mayor atención por parte de la economía experimental. En la actualidad podemos hablar de miles de experimentos en los que ambas estructuras han sido estudiadas, haciendo cada uno de ellos hincapié en aspectos diferentes de ambas¹⁰¹. ¿Quiere esto decir que después de tal cantidad de experimentos sabemos más acerca del problema que planteaba Olson? La respuesta correcta a esta pregunta es que actualmente sabemos mucho más acerca de cómo las personas se comportan ante un dilema del prisionero o un juego de bien público y esto, a su vez, nos

¹⁰¹ Para un repaso de los diversos tipos de experimentos diseñados para estudiar estos juegos ver Ledyard (1995). Para un resumen más recientes de las aportaciones de la economía experimental al estudio de la cooperación humana, ver Ostrom (2000).

permite discutir los fundamentos teóricos que llevaron a Olson a desarrollar su hipótesis teórica básica. Pero nunca podríamos decir que la obtención de evidencia experimental sobre la cooperación de los individuos en este tipo de situaciones experimentales refrenda o invalida la propuesta olsoniana. Así, el experimentador será el responsable de que la situación experimental se ajuste a las características exactas del dilema del prisionero, el teórico de juegos deberá rendir cuentas de que dicho juego capta la esencia del problema de la ausencia de cooperación en el comportamiento colectivo y el teórico social, finalmente, será el encargado de defender que la proposición teórica que afirma que los individuos tienen una estrategia dominante no cooperativa bajo las circunstancias descritas por Olson y Hardin goza de suficiente soporte empírico¹⁰².

Sugden (2005) ha resumido esta triangulación entre el experimento, la teoría y el mundo real del siguiente modo:

un experimento es una comprobación de una teoría preexistente. Dicha teoría pretende ser una representación de ciertas características del mundo. Al poner a prueba la teoría, el experimentador construye una situación de laboratorio que en algunos aspectos se parece a la teoría (y así representa el mundo en tanto que la teoría lo hace), mientras en otros aspectos representa el mundo de forma diferente, al reemplazar ciertos presupuestos de la teoría por el comportamiento de sujetos humanos autónomos. (p. 180)

Por tanto, aunque “la investigación experimental ha descubierto regularidades empíricas para las que, por el momento, no existen explicaciones teóricas bien desarrolladas” (Sugden, 2005: 182), éste no es su objetivo fundamental, sino el de poner

¹⁰² El problema de la cooperación en la acción colectiva es uno de los temas estrella en los experimentos sociológicos de *orientación teórica*. Otros temas sobre los que se han realizado un número considerable de experimentos teóricos en los últimos años son: “análisis de organizaciones” (Kanazawa, 1999a, 1999b), “confianza” (Kollock, 1994; Hayashi y Yosano, 2005; Yamagishi *et al.*, 2005), “redes sociales” (Smatka *et al.*, 1999; Willer, Troyer y Lovaglia, 2005), “creencias sociales” (Markovsky y Thye, 2001) o “diferencias de género” (Shelly y Munroe, 1999).

a prueba teorías preexistentes. Terminaremos este apartado con otra cita de Kanazawa (1999a) que deja meridianamente claro el argumento que estamos desarrollando:

Los experimentos de laboratorio no son métodos de descubrimiento. Sin una teoría general y abstracta con condiciones de posibilidad claramente establecidas, éstos no pueden decirnos nada más allá de cómo algunos “estudiantes universitarios” se comportan en un entorno dado altamente artificial; no tienen ninguna utilidad. Sin embargo, *con* una teoría, los experimentos son inigualables como un medio de evaluar y mejorar la teoría. (p. 455)

Aunque se puede deducir de lo comentado hasta el momento, en el siguiente apartado nos ocuparemos de uno de los aspectos más polémicos, aunque también uno de los que provocan más equívocos, de los experimentos de laboratorio: su validez externa.

6.3. La validez externa de los experimentos de *orientación teórica*

La validez externa de una investigación tiene que ver con la posibilidad de generalizar los resultados obtenidos en la misma a una población más amplia, otros entornos de investigación, otras variables de tratamiento u otras variables de medición (ver, entre otros, Moser y Kalton, 1971: 220; Campbell y Stanley, 1963: 5). Intuitivamente se trata de una problemática que asociamos a las encuestas, sondeos de opinión y otros tipos de técnicas de investigación que pretenden, a partir de una determinada muestra, determinar las opiniones, actitudes y comportamientos de una población mayor. Sin embargo, este tipo de requisito técnico no solamente ha preocupado a los profesionales de las encuestas, sino, como veíamos anteriormente, también a otro tipo de investigadores sociales como aquellos que han empleado diseños experimentales. Ahora bien, y esta es una de las tesis centrales de este capítulo, la preocupación por la validez externa no tiene sentido cuando nos referimos a los experimentos de *orientación*

teórica. La siguiente afirmación de Kanazawa (1999) no puede ser más clara al respecto:

Una de las críticas más comunes a los experimentos de laboratorio es la ausencia de validez externa. Los críticos a menudo acusan a los experimentos de laboratorio de tener una validez externa débil. Esto es falso; los experimentos de laboratorio *no* tienen validez externa alguna. (p. 452)

Esta afirmación puede rechinar en los oídos del sociólogo contemporáneo, tan preocupado por determinar hasta qué punto sus resultados son generalizables, pero recordemos que los experimentos de *orientación teórica* no suponen un método de descubrimiento, sino un método mediante el cual poner a prueba determinadas teorías. Volvamos por un momento al ejemplo del problema de la cooperación en la acción colectiva introducido en el apartado anterior. Según numerosos estudios experimentales (ver Ostrom, 2000: 140), aquellos sujetos que creen que otros van a cooperar en una situación de acción colectiva determinada tienen una mayor probabilidad de cooperar en la misma. En otras palabras, la expectativa de un comportamiento cooperativo del resto supone una fuerza motivadora del comportamiento cooperativo propio, lo que nos puede llevar a concluir que se está activando un mecanismo de “cooperación condicional” o, en términos generales, algún tipo de comportamiento recíproco. Este resultado experimental, habiendo sido replicado en innumerables ocasiones, se ha convertido en uno de los resultados más robustos que ha aportado la economía experimental al estudio del problema de la cooperación humana. Sin embargo, este resultado no nos dice nada acerca de si es más probable observar pautas de reciprocidad en España o en Alemania, entre jóvenes o entre personas mayores, entre votantes de izquierda o votantes de derecha. Para abordar estas últimas cuestiones sí que sería necesario plantear el problema de la validez externa. Lo que en realidad aporta el mecanismo descrito anteriormente es que, cuando nos dispongamos a construir nuestros

modelos teóricos y modelos matemáticos, tendremos que tener en cuenta que las “expectativas de reciprocidad” pueden tener cabida en la explicación de por qué un determinado grupo de personas podría cooperar para la obtención de un bien colectivo¹⁰³.

A pesar de lo expuesto anteriormente, la cuestión de la validez externa sigue siendo una cuestión disputada en la economía experimental. Entre otros motivos porque esta subdisciplina en el campo de la economía ha descubierto regularidades empíricas para las que, por el momento, no existen explicaciones teóricas bien elaboradas (Sugden, 2005: 182). Siendo esto así, el problema que se plantea es si estas regularidades empíricas, extraídas del estudio sistemático de situaciones de laboratorio, nos dicen algo acerca de cómo funcionan realmente las instituciones económicas y sociales. Para responder a esta cuestión plantearemos la pregunta acerca de la validez externa de la investigación de forma inversa a como normalmente es planteada. No nos preguntaremos acerca de hasta qué punto los datos derivados de la investigación experimental responden a las pautas generales observadas en el mundo real, sino que nos preguntaremos por los posibles factores que podrían hacer que ambos ámbitos fueran diferentes de un modo sustancial.

Las diversas reflexiones metodológicas a propósito de la especificidad de los experimentos de laboratorio se han centrado en el problema de la *artificialidad*. Ésta, siendo una característica definitoria de los experimentos de laboratorio – el propio

¹⁰³ Alguien podría afirmar que los trabajos de Axelrod (1981; 1984) sobre la cooperación en juegos repetidos ya habían adelantado tal mecanismo teórico. En este caso, la experimentación puede servir para estudiar con “sujetos reales” cómo funciona en la práctica dicho mecanismo, permitiendo plantear preguntas del tipo: ¿tiene el conocimiento del horizonte temporal una influencia sobre las expectativas de reciprocidad?, ¿influye el tamaño del grupo en el comportamiento recíproco?, ¿supone la reciprocidad una preferencia intrínseca o un comportamiento estratégico por parte de egoístas maximizadores?

investigador ha creado *ex ante* la situación que pretende estudiar – plantea algunos interrogantes metodológicos que apuntan directamente a la validez externa de la investigación. Por ejemplo, en determinados casos, “la abstracción de algunos diseños experimentales puede provocar que algunas de las variables que afectan de forma importante las decisiones en el mundo real queden fuera del diseño del laboratorio y por ello no se puedan obtener conclusiones aplicables al exterior” (Rey, 2006: 8). Por su parte, Sugden (2005: 180) ha señalado que en multitud de ocasiones se ha criticado a los diseños experimentales por la ausencia de determinadas características supuestamente necesarias en toda investigación. Entre estas características destacan las siguientes:

- los sujetos no son representativos de la población de agentes cuyo comportamiento económico la teoría está diseñada para explicar;
- éstos no tienen la oportunidad adecuada de aprender por ensayo y error cómo actuar racionalmente, del mismo modo que los agentes económicos lo hacen en situaciones reales;
- los incentivos que han de actuar de forma óptima son menores en el laboratorio que fuera del mismo;
- la naturaleza abstracta de las tareas experimentales elimina pistas que, en el campo, ayudarían a las personas a orientarse en problemas de decisión;
- las normas sociales que guían la interacción en el campo no son reproducidas adecuadamente en el laboratorio.

Tratar cada uno de estos retos metodológicos supone responder a la pregunta previa acerca de la relación entre la situación de laboratorio y el mundo real. Si esta relación es directa, sin la mediación de ninguna teoría, todos los problemas señalados anteriormente cobran una especial relevancia. Por el contrario, si el experimento está construido a partir de una teoría sólida, con unas condiciones de posibilidad bien definidas, un buen diseño *interno* de la investigación puede conducirnos a no necesitar plantearnos el problema de la validez externa, como ya señalaba Kanazawa en el extracto reproducido más arriba.

Una vez que hemos discutido las dos cuestiones metodológicas planteadas en la introducción – qué entendemos por método experimental y cuáles son sus fuentes de validez – vamos a concluir este capítulo haciendo referencia a dos cuestiones clave, y a la vez controvertidas, del tipo de experimentos predominantes en la economía experimental¹⁰⁴.

6.4. La especificidad de los experimentos económicos

El diseño experimental que va a ser empleado en esta tesis, además de ceñirse al tipo de experimento de *orientación teórica* que ha sido definido en el apartado anterior, se ajusta al tipo de diseño predominante en la economía experimental contemporánea¹⁰⁵.

Ello implica dos elecciones metodológicas que no tienen por qué haber estado presentes

¹⁰⁴ Aquí se establece una correspondencia entre lo que denominamos experimentos de *orientación teórica* y el tipo de experimento predominante en la economía experimental contemporánea. En realidad no todas las investigaciones que emplean el método experimental en economía persiguen un objetivo puramente teórico, aunque éste sí es el predominante. No es éste el lugar para profundizar en la heterogeneidad de los objetivos, diseños y procedimientos propios de la economía experimental.

¹⁰⁵ Para un panorama bastante completo de los logros y las limitaciones de la economía experimental, ver la compilación de Kagel y Roth (1995) y el texto de Camerer (2003).

en otros diseños experimentales en el seno de disciplinas como la psicología social o la propia sociología. En primer lugar, todas las decisiones que los agentes toman a lo largo de una sesión experimental tienen que estar relacionadas con el sistema de incentivos materiales empleado en el experimento. En segundo lugar, toda la información ofrecida a los participantes tiene que ser pública y verdadera. Dado que se trata de dos criterios que pueden resultar hasta cierto punto arbitrarios, en este apartado nos detendremos a argumentar brevemente el porqué de su uso.

6.4.1. Uso de Incentivos monetarios

En un tratado reciente sobre metodología de la economía experimental, Francesco Guala (2005: 25) afirma que “El hábito de compensar [económicamente] a los sujetos distingue a la economía experimental de otras disciplinas vecinas, como la psicología experimental.” Aunque este requisito de *pagar* a los participantes por las tareas realizadas en el experimento pueda resultar innecesario para otras tradiciones experimentales, no debe sorprender que el método experimental aplicado a una ciencia – la economía – que pone el énfasis en los incentivos que tienen los agentes para tomar decisiones, derive en una importante preocupación por el mejor modo de recrear dichos incentivos en el laboratorio. Dicho esto, ¿qué requisitos deben cumplir los sistemas de incentivos empleados en los diseños experimentales?

Rey (2006: 3) señala que “para que los incentivos del experimento permitan cierto control sobre las preferencias de los sujetos es necesario que se cumplan tres condiciones:

1. *Contingencia*: La cantidad de incentivo recibido debe depender al menos de la decisión tomada por el propio sujeto.
2. *Dominancia*: Los cambios en la satisfacción del sujeto con el experimento se tienen que deber fundamentalmente a cambios en las cantidades de dinero recibido.
3. *Monotonicidad*: Un mayor pago del incentivo debe ser siempre preferido a un pago menor, y los sujetos no deben llegar a estar saciados.”

Cuando se cumplen estas tres condiciones el experimentador puede asumir que los comportamientos observados a lo largo del experimento dependen directamente de los incentivos económicos empleados en el mismo. Pero esto, por supuesto, no quiere decir que todo comportamiento que no se ajuste a la maximización de los incentivos materiales deba ser rechazado. Muy al contrario, el uso explícito de incentivos materiales ha permitido el estudio controlado de motivaciones no materiales, tales como el altruismo, la generosidad o el comportamiento guiado por normas morales y convenciones sociales¹⁰⁶. Nos detendremos en un tipo de experimento bastante popular en la economía experimental para ilustrar este argumento: el *juego del dictador*.

¹⁰⁶ Mecanismos experimentales como el denominado “juego del ultimátum” (Güth, Schmittberger y Schwarse, 1982; Güth y Tietz, 1990) o el “juego del dictador” (Kahneman, Knetsch y Thaler, 1986) han contribuido a impulsar el estudio y análisis de motivaciones no materiales presentes en la toma de decisiones económicas y sociales. A su vez este tipo de estudios experimentales han contribuido de forma muy importante al desarrollo de la denominada “economía del comportamiento”, una rama de la teoría económica en la que se ponen sistemáticamente a pruebas los principales fundamentos teóricos de la aplicación de la teoría de la elección racional al campo económico. Para un exhaustivo trabajo que vincula los resultados experimentales con el avance de la “economía del comportamiento”, ver Camerer (2003).

El juego del dictador (*dictator game*) es uno de los juegos más simples empleados por los economistas experimentales para el estudio de las motivaciones humanas. En éste participan dos jugadores, el “proponente” (*proposer*) y el “receptor” (*receiver*). El proponente determina una asignación (división) de alguna dotación (por ejemplo, una cantidad económica obtenida en un premio). El receptor en este caso simplemente recibe el remanente de la dotación no asignada a sí mismo por el proponente. El receptor desempeña un papel completamente pasivo: no tiene ninguna influencia estratégica sobre el resultado del juego¹⁰⁷. Este juego ha sido utilizado para poner a prueba el modelo del *homo oeconomicus*; así, si los individuos estuvieran preocupados sólo por su propio bienestar, los proponentes se asignarían a sí mismos la totalidad de la “tarta” que se ha de repartir y no le darían nada a los receptores. Sin embargo, numerosos estudios, incluyendo un amplio estudio comparativo trans-cultural (ver Henrich *et al.*, 2004), han demostrado que los proponentes asignan cantidades de la dotación por encima de la predicción teórica (“dar cero”). Estos resultados pueden hacernos conjeturar dos tipos de explicaciones: bien los proponentes no son capaces de maximizar su utilidad esperada, o bien la función de utilidad de los proponentes incluye los beneficios incluidos por los otros. Ambas explicaciones plantean, como mínimo, retos serios para la teoría de la elección racional más estándar.

El ejemplo del juego del dictador nos ha servido para tener una primera intuición acerca de cómo son utilizados los incentivos monetarios en la economía experimental, así como acerca de las interpretaciones que podemos hacer a propósito de las motivaciones individuales a partir de los comportamientos observados en el experimento. En otros casos, como sucederá en el siguiente capítulo de esta tesis, los

¹⁰⁷ Al no existir interacción estratégica, numerosos autores niegan que se pueda llamar juego a este tipo de diseño. En aras de la claridad expositiva aquí seguiremos hablando del “juego del dictador.”

incentivos permiten reconstruir el sistema de interacción contenido en una matriz de juegos concreta. En nuestro caso se tratará de definir los incentivos correctamente para poder poner a prueba el modelo teórico presentado en la segunda parte de la tesis. De ello nos ocuparemos más adelante.

6.4.2. Prohibición del engaño (deception)

De acuerdo con las convenciones imperantes en la economía experimental, toda la información que se suministre a los sujetos experimentales debe ser verdadera y el experimento debe ajustarse completamente a ésta. Como señala Rey (2006: n. 15), “la preocupación por este tema es tal que existe un acuerdo tácito por el que no se acepta la publicación de ningún artículo en el que se sospeche que ha habido ‘engaño’”¹⁰⁸. La prohibición del engaño en economía experimental incluye: el engaño acerca del propósito del experimento, acerca de los pagos que el sujeto puede obtener o acerca de la caracterización de los sujetos con los que el participante está interactuando (Croson, 2002: 940).

Se han señalado dos razones en defensa de esta prohibición (Croson, 2002). En primer lugar, la validez de los experimentos económicos está basada fundamentalmente en la conexión entre el comportamiento de los sujetos experimentales y el pago esperado. Si el sujeto experimental pone en duda dicha conexión el propio experimento pierde todo su sentido. En segundo lugar, en economía experimental, la confianza que los participantes depositan en el experimentador es considerada un bien público. Por

¹⁰⁸ En este sentido, la economía experimental se sitúa bastante lejos de las posiciones mantenidas por los psicólogos experimentales. Recordemos sin ir más lejos los famosos experimentos llevados a cabo por Stanley Milgram sobre el acatamiento a la autoridad. En éstos se hacía creer a los sujetos experimentales que estaban infringiendo un castigo físico real a otros participantes, siendo éste sólo el más llamativo de los engaños a los que estaban sometidos los mismos.

ejemplo, si un sujeto participa en repetidas ocasiones en un experimento e independientemente de su implicación en el mismo siempre recibe una suma constante de dinero, puede tender a desconfiar en la palabra del experimentador y, de este modo, comportarse de una manera que resulte completamente impredecible al no depender de los incentivos propuestos.

Al igual que la cuestión del empleo de incentivos monetarios, la prohibición de utilizar el engaño para estudiar ciertas pautas de comportamiento humano, aun siendo una cuestión susceptible de debate, no está sujeta a discusión en el estado actual de la economía experimental y es aceptada como un precepto que hay que cumplir. En la actualidad existen numerosos debates metodológicos abiertos acerca de éstas y otras cuestiones particulares de la economía experimental; prueba de ello son trabajos como el de Guala (2005) o el monográfico de la *Journal of Economic Methodology* coordinado por Robert Sugden en 2005. Nuestra aproximación al tema se limita a las breves pinceladas introducidas hasta el momento y las explicaciones concretas aportadas en el capítulo 7. Un tratamiento más profundo del mismo se escapa a los objetivos centrales de esta tesis.

6. 5. Conclusiones

Este capítulo ha pretendido ser una introducción al método experimental, en general, y al uso que se va a dar al mismo en esta tesis, en particular. Para ello hemos planteado primero la cuestión de qué entendemos por método experimental – un método que pretender establecer relaciones causales a partir del estudio sistemático y controlado de diferentes situaciones de contraste –, así como qué tipos ideales de diseños podemos

encontrar. En concreto, hemos distinguido entre experimentos de *orientación empírica* y experimentos de *orientación teórica*. Dado que el tipo de diseño que vamos a emplear puede ser encuadrado en esta última categoría, hemos planteado los problemas de validez a los que dicho tipo de experimento tiene que hacer frente.

La cuestión determinante a la hora de evaluar un experimento de *orientación teórica* es su validez interna, es decir, hasta qué punto nuestro diseño mide lo que pretendemos medir en teoría. La validez interna de este tipo de experimentos depende críticamente de una adecuada triangulación de tres elementos: mundo real, teoría y diseño experimental. Para obtener una validez interna alta debemos disponer de una teoría que explique adecuadamente la realidad y que, además, esté lo suficientemente bien definida como para que pueda ser reproducida en el laboratorio de forma satisfactoria.

Si la validez interna es lo que nos permite discriminar entre buenos y malos experimentos de *orientación teórica*, otro tipo de fuente de validez, la validez externa, desempeña un papel secundario. Si nuestro objetivo es analizar el comportamiento de una teoría dentro de su alcance y sus condiciones de posibilidad, la pregunta a propósito de la generalización de los resultados de un experimento concreto no tiene cabida. En todo caso, deberíamos plantearnos en qué medida la situación recreada en el laboratorio es sustancialmente distinta al tipo de mecanismo social observado en la realidad y que pretendemos estudiar. Varios autores coinciden en resaltar el problema de la *artificialidad* propia de las situaciones de laboratorio como una de las amenazas más importantes para la validez externa de los experimentos. Sin embargo, en el caso de los experimentos de *orientación teórica*, esto nos conduce de nuevo al problema de la

especificación del experimento, es decir, su validez interna; volviendo a constatar la intrascendencia de la preocupación por la validez externa cuando nos referimos a este tipo de experimento.

Por último, nos hemos detenido en dos características claves de los diseños experimentales en el campo de la economía experimental: el uso de incentivos materiales y la prohibición del engaño. Dado que en el siguiente capítulo ambos preceptos serán respetados, hemos creído conveniente aportar algunas aclaraciones metodológicas con relación a su empleo.

La tercera parte de esta tesis tiene como objetivo poner a prueba los argumentos teóricos desarrollados en las dos primeras partes. Como preludeo a tal tarea, en la segunda, hemos elaborado una propuesta teórica cuyo alcance y condiciones de posibilidad han quedado suficientemente establecidos. Esta labor nos permitirá aplicar a continuación el método experimental, un método sin parangón cuando lo que pretendemos es poner a prueba hipótesis teóricas bien definidas.

CAPÍTULO 7

LA DOBLE NATURALEZA DE LAS CONVENCIONES SOCIALES: UN ANÁLISIS EXPERIMENTAL

7.1. Introducción

Una vez que han sido presentadas las características básicas del método que va a ser empleado en esta tercera parte de la tesis, volvemos a retomar la línea central de la misma: el estudio teórico de la noción de convención social. En esta parte tercera desarrollaremos el último de los estadios definidos anteriormente para acometer dicho estudio; tras el planteamiento conceptual de la primera parte y la reconstrucción teórico-formal de la segunda, en esta tercera emplearemos el método experimental para poner a prueba las hipótesis derivadas de las dos anteriores.

En el capítulo 6 afirmábamos que el uso de experimentos de *orientación teórica* dependía críticamente, en primer lugar, de disponer de una teoría cuyo alcance y condiciones de posibilidad estén perfectamente definidos. Definir dicho alcance y condiciones de posibilidad ha sido el objeto de nuestro trabajo en el capítulo 5. Pero tener a nuestra disposición una teoría perfectamente definida no nos garantiza una aplicación satisfactoria del método, sino que también es necesario que el diseño concreto del experimento, o los experimentos, se ajuste en la mayor medida posible a las condiciones teóricas especificadas. Por tanto, el objetivo de este capítulo es doble: por un lado, nos preguntaremos hasta qué punto es posible el estudio empírico de los argumentos teóricos desarrollados en capítulos precedentes; en segundo lugar, y

siempre que la respuesta a la pregunta anterior sea positiva, nos preguntaremos por la base empírico-experimental de los argumentos teóricos expuestos hasta el momento.

En la última década, han sido exploradas varias líneas de investigación experimental sobre la emergencia y la evolución de las convenciones sociales, sobre todo como un modo de abordar empíricamente los desarrollos teóricos derivados de la aplicación de la teoría de juegos evolutiva al estudio de las mismas¹⁰⁹. Van Huyck, Battalio y Rankin (1997), por ejemplo, abordan el problema de la emergencia de las convenciones sociales a partir del estudio experimental de juegos puros de coordinación. Schotter y Sopher (2003), por su parte, plantean un problema de investigación – la transmisión de convenciones entre distintas generaciones de participantes en un experimento – en un entorno bastante más cercano al que aquí nos ocupa: un juego de coordinación asimétrica – en concreto, el juego de la batalla de los sexos. Una tercera línea de investigación a la que vamos a hacer referencia es el estudio de cómo las convenciones emergen en “estructuras de interacción local” (Berninghaus, Ehrhart y Keser, 2002), es decir en grupos reducidos de personas que interaccionan directamente. Estas tres referencias sólo cumplen aquí la función de mostrar cómo el estudio experimental de las convenciones sociales está tomando rumbos diversos. Sin embargo, hasta donde conocemos, no se ha llevado a cabo ningún tipo de estudio centrado en el aspecto normativo inherente a toda convención social¹¹⁰.

¹⁰⁹ En el capítulo 5 hacíamos referencia a algunas de las propuestas evolutivas más destacadas en el estudio de las convenciones sociales.

¹¹⁰ Este hecho tampoco debe extrañarnos ya que todo este conjunto de investigaciones han sido llevadas a cabo por teóricos económicos y ya apuntábamos en el capítulo 2 la ausencia de componentes normativos en la noción de convención social como *convención de comportamiento* presente en la teoría económica.

Para estudiar este componente normativo de las convenciones hemos llevado a cabo dos experimentos¹¹¹. La estructura de ambos es casi idéntica (excepto en la manipulación de las variables de tratamiento) e intenta reflejar de forma cristalina la estructura del modelo teórico presentado en el capítulo 5. Así, en ambos experimentos, los participantes toman parte de forma repetida en una serie de juegos que se ajustan a la estructura de nuestra clase de juegos básica. Además, si en la segunda parte diseñábamos una serie de principios que contribuían a resolver la incertidumbre contenida en la clase de juegos básica, en esta tercera los principios son puestos a disposición de los participantes en los experimentos con el objetivo de estudiar cómo son utilizados para incrementar los niveles de coordinación. Por tanto, en resumen, cada experimento se compone básicamente de dos elementos: una serie de juegos que suponen concreciones numéricas de la estructura de preferencias de nuestra clase de juegos básica y una serie de principios que prescriben cómo tomar decisiones al participar en dichos juegos¹¹².

Como se observará, el primero de los experimentos es más complejo ya que introduce distintos niveles de incertidumbre con respecto a la información que los participantes tienen acerca del juego. El segundo, sin embargo, fue explícitamente

¹¹¹ La parte empírica de esta tesis, cuya materialización son los dos experimentos que se presentan a continuación, es el resultado de una colaboración de más de dos años entre el autor de esta tesis y el *Strategic Interaction Group* del *Max Planck Institute of Economics* de Jena (Alemania) – materializada en tres estancias de investigación de un mes cada una aproximadamente. En el diseño y ejecución de los experimentos han desempeñado un papel fundamental los miembros de dicho grupo, en especial su director Werner Güth. Este hecho ha permitido que la tesis cuente con una parte empírica que, tanto por cuestiones técnicas como económicas, no podría haberse llevado a cabo de otro modo. Sin embargo, dicho trabajo conjunto ha hecho que los experimentos que se presentan no supongan una puesta a prueba exacta del marco teórico presentado – aunque éste inspirara los mismos – y podamos encontrar tanto aspectos de dicho marco que quedan sin estudiar, como resultados no derivados directamente del mismo. Aún así, consideramos, en primer lugar, que las pruebas empíricas realizadas suponen, en sí mismas, una aportación relevante para este trabajo. En segundo lugar, hay que decir que ambos experimentos sólo son la primera “piedra de toque” para nuestro marco teórico ya que el programa de colaboración en esta materia con el *Strategic Interaction Group* continúa abierto.

¹¹² En los apartados 7.2.1 y 7.3.1. se realizan sendas descripciones pormenorizadas del diseño y el procedimiento empleados en cada uno de los experimentos.

diseñado de forma más simple para confirmar algunos de los resultados obtenidos en el primero. Así, la estructura general del capítulo es la siguiente. Ambos experimentos¹¹³ son presentados de forma independiente; primero se dedica un subapartado a exponer las especificidades del diseño concreto de cada experimento, después se detallan los resultados del mismo y, por último, se exponen las conclusiones particulares de cada uno. El capítulo se cierra con unas conclusiones generales que, a partir de la comparación de ambos experimentos, pretenden vincular los resultados obtenidos con posibles líneas futuras de investigación.

7.2. Primer experimento

7.2.1. Diseño y conducción del primer experimento

Este primer experimento fue realizado en el laboratorio del Max Planck Institute of Economics en Jena (Alemania). Para la realización del mismo, se reclutaron 160 estudiantes de licenciatura (procedentes de diferentes disciplinas) de la Universidad de Jena, mediante el sistema del reclutamiento *online* para experimentos económicos ORSEE 2.0 (Greiner, 2004). Se llevaron a cabo cinco sesiones experimentales: cada sesión correspondía a un tratamiento distinto. En cada sesión experimental participaron 32 sujetos. La traducción de las instrucciones de este experimento (originalmente en alemán) puede ser encontrada en el Apéndice 1 al final del capítulo. Para la conducción informática del experimento se utilizó el programa z-Tree (Fischbacher, 1999). Los sujetos recibieron las instrucciones por escrito y, además, estas fueron leídas en voz alta

¹¹³ Aunque el período de tiempo transcurrido entre ambos experimentos fue de casi un año, los resultados obtenidos en los dos experimentos, como veremos a continuación, son llamativamente similares. El primer experimento tuvo lugar el 27 de abril del 2005, mientras que el segundo fue llevado a cabo el 13 de marzo del 2006.

para garantizar la existencia de conocimiento común acerca de las mismas. Ningún tipo de comunicación más allá del uso de los principios (como explicaremos más adelante) fue permitida. Además, los sujetos no podían identificar los miembros del grupo (*matching group*) en el que estaban encuadrados con los que estaban interactuando realmente. Al final de cada sesión experimental, se pagó a los sujetos en metálico de acuerdo con los pagos obtenidos en el juego (además de una suma fija de 2€ sólo por haberse presentado).

A cada sujeto le fue asignado un rol (A o B) y todos interactuaban en un grupo de cuatro sujetos a la hora de tomar parte en los 20 juegos simples en los que tenían que participar. En total había 10 matrices de pagos (10 juegos¹¹⁴) distintas y los sujetos tenían que jugar cada juego dos veces. Durante los primeros diez juegos (1-10), cada sujeto interactuaba con un miembro de su grupo y durante los segundos (11-20) con otro miembro distinto del grupo. Cada tratamiento era diferente en cuanto a la manipulación de tres variables, a saber: (i) el conocimiento de los parámetros numéricos del juego, (ii) el conocimiento del rol propio y (iii) el uso estratégico de los principios propuestos. La Tabla 6 presenta un resumen de las principales características del experimento.

¹¹⁴ Los diez juegos pueden ser consultados en las instrucciones en el Apéndice 1 de la tesis.

Tabla 6: Características de los tratamientos del primer experimento

	Tr-C	Tr-P	Tr-PC	Tr-N	Tr-NC
Número de sesiones	1	1	1	1	1
Períodos por sesión	20	20	20	20	20
Grupos (<i>matching groups</i>)	8	8	8	8	8
Participantes	32	32	32	32	32
Conocimiento de parámetros	No	No	Sí	Sí	Sí
Conocimiento de roles	No	Sí	No	Sí	Sí
Uso estratégico de principios	Sí	Sí	Sí	Sí	No
Coordinación alcanzada	69,38%	53,75%	78,13%	74,38%	66,56%
Ganancia media por sujeto	7,5€	6,8€	8,5€	8,2€	7,6€

En cada tratamiento los participantes tenían que desarrollar dos tareas. En primer lugar, tenían que ordenar los ocho principios descritos en la segunda parte de la tesis¹¹⁵ y, segundo, tenían que participar en 10 juegos de forma sucesiva bajo distintos niveles de información. La primera de estas tareas demandaba a los participantes que rellenaran una tabla con los principios propuestos (ver Figura 8). No tenían que sugerir un orden completo; el único requisito era proporcionar al menos un principio. Una vez que cada sujeto había proporcionado su ranking, era emparejado con uno de los miembros de su grupo e interactuaba durante las diez primeras rondas de la sesión con esta persona. Antes de jugar los juegos, los participantes conocían tanto su propio ranking, como el ranking de la persona con la que estaban interactuando (ver Figura 9). Sólo en el grupo de control no existía este conocimiento acerca del ranking de la persona con la que uno interactuaba. Después de cada ronda los sujetos no recibían ningún tipo de información. Al final de los 20 períodos cada participante fue informado acerca de sus ganancias y se le pagó en metálico (ver las ganancias medias obtenidas en la Tabla 6).

¹¹⁵ No volveremos a hacer aquí una presentación exhaustiva de los ocho principios, tampoco de nuestra clase de juegos básica; dicha tarea ya fue realizada en la parte segunda de la tesis. En todo caso, para ver cómo fue administrada toda esta información a los participantes, ver las instrucciones del experimento en el Apéndice 1.

Figura 8: Pantalla del ranking en el primer experimento

Periode 1 von 1

Verbleibende Zeit [sec]: 4

Bitte entscheiden Sie sich für Ihre Rangfolge der Prinzipien!
Verwenden Sie dazu bitte die in den Instruktionen angegebenen Buchstaben!

Erstes Prinzip:

Zweites Prinzip:

Drittes Prinzip:

Viertes Prinzip:

Fünftes Prinzip:

Sechstes Prinzip:

Siebtes Prinzip:

Achstes Prinzip:

OK

Cada sesión seguía el mismo protocolo y los tratamientos sólo se diferenciaban en la manipulación de las tres variables independientes a las que hemos hecho referencia anteriormente (ver Tabla 6). En el tratamiento C, los participantes tomaron las decisiones tras un *velo de ignorancia completo*, dado que no conocían ni los parámetros reales de cada juego, ni su propio rol. En el tratamiento P y en el PC los sujetos interactuaban tras un *velo de ignorancia parcial*, ya que en el tratamiento P conocían su propio rol pero no los parámetros numéricos de cada juego y en el tratamiento PC conocían los parámetros reales del juego pero no conocían su propio rol. El tratamiento N no supone ningún tipo de *velo de ignorancia*, debido a que los participantes conocían tanto los parámetros numéricos del juego, como sus propios roles. Por último, el tratamiento NC fue diseñado como un grupo de control en el que los sujetos establecían su ranking pero no lo comunicaban a la persona con la que estaban interactuando.

Figura 9: Pantalla de toma de decisiones en el primer experimento

Periode: 1 von 1 Verbleibende Zeit [sec]: 25

Ihre Rangfolge:

- Erstes Prinzip: A
- Zweites Prinzip: L
- Drittes Prinzip:
- Viertes Prinzip:
- Fünftes Prinzip:
- Sechstes Prinzip:
- Siebtes Prinzip:
- Achtes Prinzip:

Die Rangfolge des Anderen:

- Erstes Prinzip: S
- Zweites Prinzip: A
- Drittes Prinzip:
- Viertes Prinzip:
- Fünftes Prinzip:
- Sechstes Prinzip:
- Siebtes Prinzip:
- Achtes Prinzip:

Sie haben die Rolle A.

Strategie von A	Strategie von B	Gewinn von A	Gewinn von B
X	X	1	0.5
X	Y	-0.5	0
Y	X	0	0.2
Y	Y	0.4	1

Bitte entscheiden Sie sich nun für X oder Y:

7.2.2. Resultados del primer experimento

Con el objetivo de presentar los resultados de la forma más clara posible, en lo que sigue vamos a tratar de responder a las cuatro cuestiones siguientes:

- A. En primer lugar, analizaremos qué principios fueron seleccionados con más frecuencia. Para ello, presentaremos primero los datos correspondientes a los principios que fueron seleccionados en primer lugar y, después, los datos correspondientes a la totalidad de los rankings.

- B. En segundo lugar, comprobaremos la consistencia que mostraron los sujetos con respecto al principio ordenado en primer lugar, es decir, si el comportamiento real de éstos se ajustó al comportamiento prescrito por los primeros principios.

- C. En tercer lugar, comprobaremos si el uso de principios normativos incrementa los niveles de coordinación. En este sentido, compararemos nuestros resultados con resultados obtenidos anteriormente por otros investigadores.
- D. En cuarto lugar, explicaremos las diferencias en los niveles de coordinación alcanzados en los tratamientos a partir de las dos variables de tratamiento que hemos manipulado, a saber: el conocimiento del propio rol y/o de los parámetros reales del juego.

A. *¿Qué principios fueron seleccionados?*

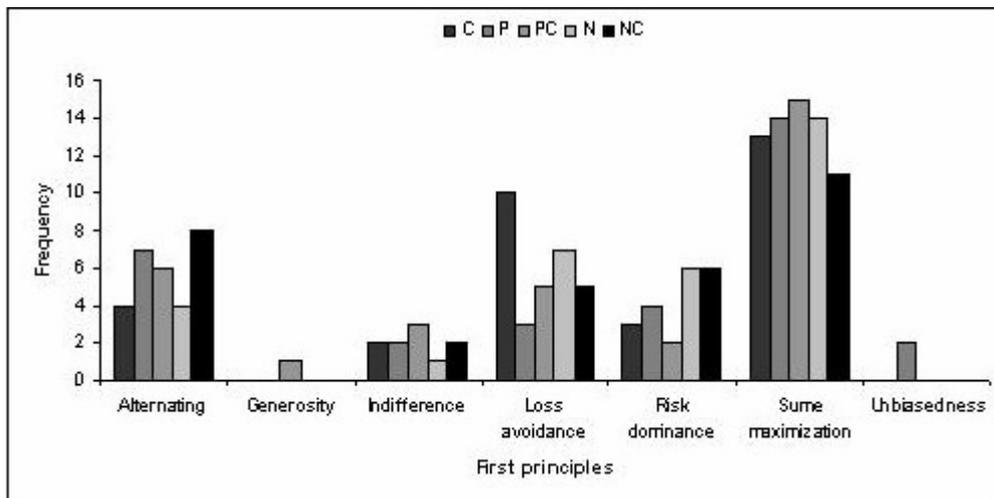
La Figura 10 muestra la frecuencia con que cada principio fue ordenado en primer lugar. Un resultado sorprendente es que los principios que no cumplen con el requisito de ser evolutivamente estables (EEE) – tal como quedó definido en la parte segunda de la tesis – apenas fueron seleccionados en primer lugar. Así, ningún sujeto situó *Minimizing risk* como su primer principio, un sujeto hizo lo propio con *Generosity* y 10 sujetos eligieron en primer lugar *Indifference induction*. En conjunto los principios, o *convenciones normativas*, no evolutivamente estables fueron seleccionados en primer lugar en menos del 7% de los casos.

El resto de principios fueron ordenados en primer lugar por un número mayor de sujetos. La única excepción parece ser el principio *U* (*Unbiasedness*) que fue seleccionado en primer lugar sólo por dos participantes. Una posible explicación es que coordinarse en el principio *U* no genera directamente una *convención de comportamiento* (un equilibrio del juego) dado que *U* implica reaccionar de forma

óptima al comportamiento del otro jugador empleando ambas estrategias con la misma probabilidad. En otras palabras, el principio *U* no proporciona “un sistema de expectativas mutuas concordantes” (Lewis, 1969: 25).

Entre los cuatro principios restantes, el principio más frecuente fue *Sum maximization (S)*, seguido de *Loss Avoidance (L)*, *Alternating (A)* y *Risk dominance (R)*. Todos estos principios recomendaban una determinada *convención de comportamiento*. No podemos afirmar que existan diferencias entre los tratamientos en relación con qué principio fue ordenado en primer lugar (Kruskal-Wallis test; $X^2 = 1,735$, $d.f. = 4$, $p = 0,784$)¹¹⁶.

Figura 10: Frecuencia de primeros principios

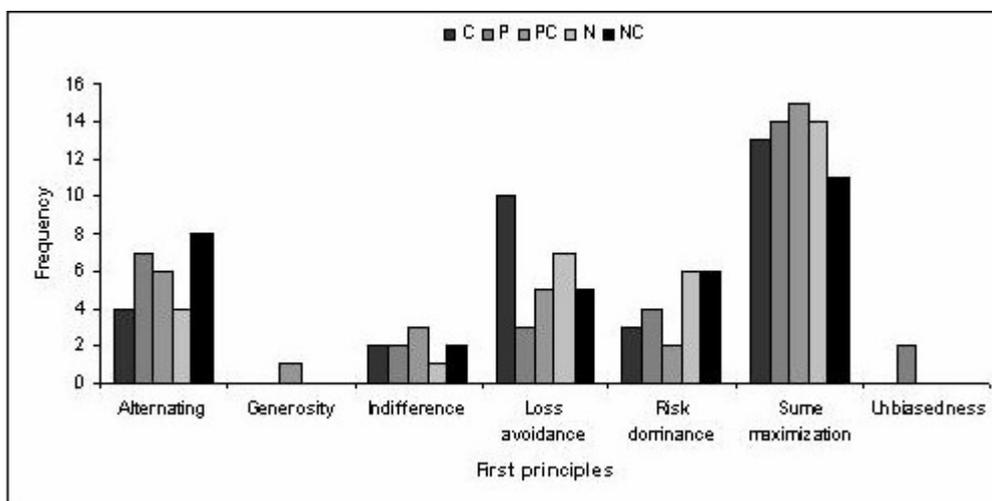


La Figura 11, en vez de considerar sólo los principios seleccionados en primer lugar, tiene en cuenta los rankings completos. Para ello, hemos asignado a cada principio un valor inverso a la posición que ocupa en el ranking (8 puntos al principio

¹¹⁶ Tanto en éste, como en el segundo experimento se utilizarán pruebas estadísticas no paramétricas para determinar las diferencias estadísticas entre los distintos tratamientos aplicados. Los contrastes no paramétricos son especialmente aconsejables cuando realizamos análisis estadísticos con muestras pequeñas, como es nuestro caso. Entre otras ventajas, se sitúa la de que no es necesario realizar supuestos fuertes sobre la distribución de los datos para que dichos contrastes puedan ser aplicables.

seleccionado en primer lugar, 7 al segundo, etc., y así hasta asignar un punto al octavo principio). Los resultados son bastante similares a los que podíamos observar en la Figura 4: los principios que no proporcionan un conjunto de “expectativas mutuas concordantes” fueron seleccionados menos y, entre los que sí las proporcionan, el orden fue como sigue: *S*, *L*, *R*, *A*. De nuevo, no existen diferencias estadísticamente significativas entre los tratamientos en relación con la forma en que los principios fueron ordenados (Kruskal-Wallis test: $X^2 = 4,858$, $d.f. = 4$, $p = 0,784$, para el principio *A*; $X^2 = 5,722$, $d.f. = 4$, $p = 0,221$, para el principio *G*; $X^2 = 2,074$, $d.f. = 4$, $p = 0,722$, para el principio *I*; $X^2 = 4,759$, $d.f. = 4$, $p = 0,313$, para el principio *L*; $X^2 = 5,367$, $d.f. = 4$, $p = 0,252$, para el principio *M*; $X^2 = 4,614$, $d.f. = 4$, $p = 0,329$, para el principio *R*; $X^2 = 1,258$, $d.f. = 4$, $p = 0,869$, para el principio *S*; $X^2 = 5,228$, $d.f. = 4$, $p = 0,265$, para el principio *U*).

Figura 11: Frecuencia de todos los principios utilizados



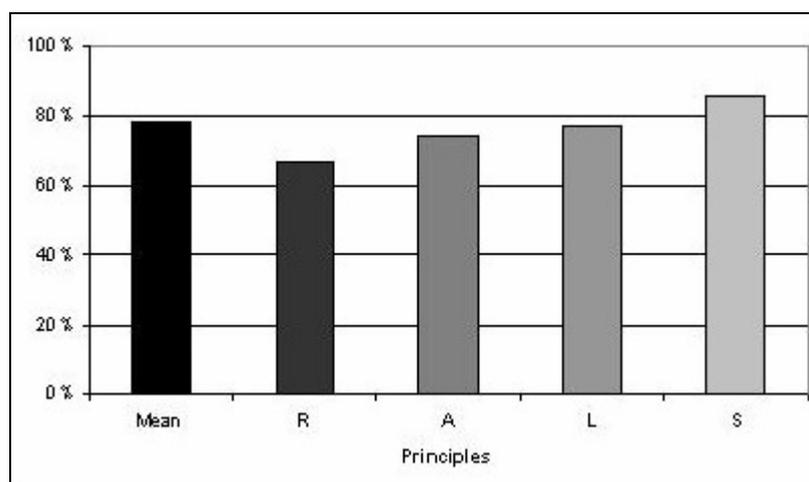
B. ¿Es el comportamiento de los sujetos consistente con los primeros principios?

La Figura 12 muestra el porcentaje de consistencia con el primer principio, es decir, hasta qué punto los participantes seleccionaron la estrategia (*convención de*

comportamiento) prescrita por su primer principio (*convención normativa*). Hemos considerado sólo los casos en los que el principio normativo recomienda una única estrategia. Por tanto, al excluir los sujetos que eligieron los principios *G, I, M, U*, el análisis se refiere únicamente a 113 casos¹¹⁷. El nivel de consistencia con los principios va desde el 66,79% para el principio *R* (*Risk dominance*) hasta el 85,5% para el principio *S* (*Sum maximization*). La consistencia media con los principios fue bastante alta (78,19%). No existen diferencias estadísticamente significativas entre los tratamientos en el nivel de consistencia con los principios (Kruskal-Wallis test; $X^2 = 4,144, d.f. = 4, p = 0,387$).

De los datos presentados se puede deducir que los sujetos no utilizaron estratégicamente los principios para lanzar señales erróneas a la persona con la que estaban interactuando para obtener un mayor pago individual. Esto es así aunque los participantes no estaban comprometidos con la realización del comportamiento prescrito por los primeros principios.

Figura 12: Consistencia con la acción prescrita por los primeros principios



¹¹⁷ Recordemos que tres de estos cuatro principios no cumplen la condición de ser evolutivamente estables y que ninguno de los cuatro logra prescribir una situación de equilibrio, una *convención de comportamiento*.

Bajo nuestro punto de vista, estos datos muestran que los sujetos emplearon los principios normativos como “dispositivos cognitivos” (*cognitive devices*) para alcanzar niveles altos de coordinación evitando, así, un largo proceso de razonamiento que les obligara a tomar una decisión concreta en cada juego particular al que se enfrentaban. Esto refuerza la hipótesis de la función cognitiva de las convenciones sociales a la que hacíamos referencia en el apartado 2.6.¹¹⁸

Aunque el nivel de consistencia con los principios es bastante alto en general, no todos desempeñan el mismo papel en nuestro experimento. En particular, el nivel de consistencia con las estrategias prescritas por el principio *S* (*Sum maximization*) es estadísticamente diferente de la consistencia con los demás principios para cualquier comparación binaria entre el principio *S* y los otros tres principios (U de Mann-Whitney test: $z = -2,496$, $p = 0,013$, *2-tailed*, para la comparación entre los principios *A* y *S*; $z = -2,414$, $p = 0,016$, *2-tailed*, para la comparación entre los principios *L* y *S*; $z = -3,888$, $p = 0,000$, *2-tailed*, para la comparación entre los principios *R* y *S*). Este resultado es, hasta cierto punto, sorprendente ya que *S* supone un criterio de eficiencia y varios estudios experimentales no han encontrado dicha preocupación por la eficiencia en los juegos de coordinación (ver Van Huyck, Battalio y Beil, 1990; 1991; Van Huyck, Cook y Battalio, 1997)¹¹⁹. De acuerdo con nuestros resultados, la eficiencia puede ser un

¹¹⁸ Recordemos que en dicho apartado equiparábamos las convenciones a “reglas heurísticas de carácter social” que ayudan a coordinar las expectativas de un grupo de personas. Sin embargo, a diferencia de lo expuesto en el apartado 2.6. aquí son los principios normativos los que desempeñan dicho papel heurístico y no las *convenciones de comportamiento* por sí mismas. Sobre la función cognitiva de las convenciones sociales ver Sacconi y Moretti (2002).

¹¹⁹ Existe un elemento que hace que esta comparación no pueda llevarse hasta sus últimas consecuencias: mientras que la mayor parte de los estudios experimentales precedentes sobre coordinación se han centrado en juegos simétricos, nuestra clase de juegos básica, tal y como fue presentada en la parte segunda, tiene varias asimetrías. El hecho de intentar dibujar un escenario más realista, en cuanto que más asimétrico, nos crea dificultades para comparar nuestros resultados directamente con investigaciones anteriores.

objetivo alcanzable cuando los sujetos pueden comunicar *ex ante* que van a emplear dicho principio.

C. *¿Contribuye el uso de principios a la mejora de la coordinación?*

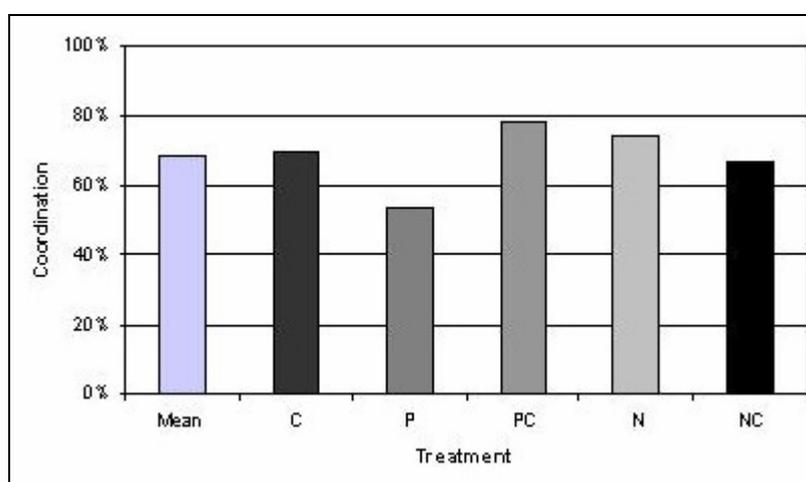
Varios estudios han mostrado que la coordinación alcanzada por los sujetos participantes en experimentos donde el juego base es una “batalla de los sexos” sin comunicación previa se encuentra entre el 40% y el 50% (ver Cooper *et al.*, 1989; 1993; Straub, 1995). Para incrementar este nivel de coordinación se han utilizado diferentes mecanismos, como ofrecer la opción de no participar en el juego a uno de los sujetos, permitir diferentes niveles de comunicación entre éstos o introducir un *timing* particular en el proceso de toma de decisiones (ver Camerer, 2003: cap. 6). Sin embargo, hasta donde llega nuestro conocimiento, no se ha conducido ningún experimento hasta el momento en el que los participantes pudieran elegir un principio normativo concreto antes de participar en un juego de coordinación.

La coordinación media alcanzada en nuestro experimento es del 68,78% y va desde el 53,75% en el tratamiento P, hasta el 78,13% en el tratamiento PC (ver Figura 13). Podemos encontrar diferencias estadísticamente significativas entre los tratamientos (Kruskal-Wallis test; $X^2 = 12,667$, $d.f. = 4$, $p = 0,013$). Al comparar nuestros resultados con los obtenidos por otros investigadores¹²⁰, concluimos que el uso de principios mejora los niveles de coordinación.

¹²⁰ En un experimento cuyo juego base era una batalla de los sexos de 2 jugadores en el que los participantes podían comunicar (*cheap talk*) la estrategia que iban a seleccionar, Cooper *et al.* (1989) obtuvieron un nivel de coordinación del 55%. En una réplica de dicho tratamiento (Cooper *et al.*, 1993) el nivel de coordinación alcanzado fue del 58%.

Diseñamos nuestro grupo de control para poner a prueba esta hipótesis. Como hemos explicado anteriormente, en el tratamiento del grupo de control (NC) los sujetos tenían que ordenar los principios pero no comunicaban dicho ranking a la persona con la que estaban interactuando. Si comparamos los tratamientos N y NC, los cuales difieren sólo en el uso estratégico de los principios, encontramos que el nivel de coordinación alcanzado es un 8% menor en el tratamiento NC (U de Mann-Whitney test; $z = -2,088$, $p = 0,038$, 2-tailed). De este modo, la comunicación de la *convención normativa* propia contribuye a una coordinación efectiva en una *convención de comportamiento* concreta.

Figura 13: Nivel de coordinación alcanzado en cada tratamiento

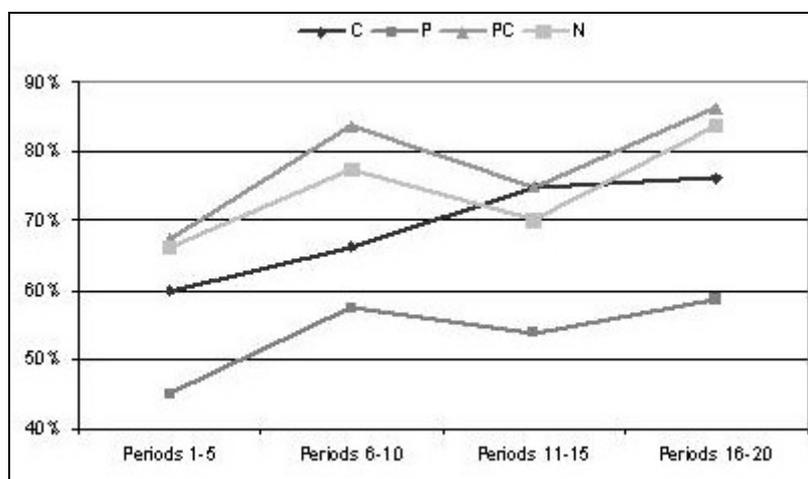


¿Mantuvieron los sujetos los mismos patrones de coordinación durante las veinte rondas de que constaba cada sesión? La Figura 14 muestra los niveles de coordinación agregados por grupos de cinco rondas cada uno (períodos 1-5, períodos 6-10, períodos 11-15 y períodos 16-20). En dicho gráfico podemos observar dos aspectos importantes. En primer lugar, la coordinación en la segunda parte del experimento (períodos 11-20) es significativamente superior a la coordinación en la primera parte (períodos 1-10) (Wilcoxon signed ranks test; $z = -2,858$, $p = 0,004$, 2-tailed)¹²¹. En segundo lugar, en cada tratamiento (excepto en el tratamiento C), el nivel de coordinación es más bajo en el tercer grupo de períodos que en el segundo, lo que se debe al cambio de la persona

¹²¹ A partir de este momento no utilizaremos los datos del tratamiento NC (grupo de control) con el objetivo de mantener constante el posible efecto del uso estratégico de los principios.

con la que se está interactuando y la necesidad de coordinarse en los principios de nuevo¹²². En general, podemos concluir que la coordinación mejora con la experiencia.

Figura 14: Nivel de coordinación alcanzado en distintos grupos de períodos



D. ¿Cómo podemos explicar las diferencias existentes entre los tratamientos?

Al analizar las diferencias entre los tratamientos, encontramos que las dos variables de tratamiento que habíamos manipulado afectan a la coordinación de forma distinta. Mientras el conocimiento acerca de los parámetros numéricos del juego mejora la coordinación, el conocimiento acerca del rol propio hace que los niveles de coordinación disminuyan. Este hecho demuestra claramente la necesidad de estudiar la coordinación en juegos asimétricos, en vez de concentrarse en juegos simétricos, tal y como ha sido realizado en estudios experimentales precedentes.

Cuando los participantes conocen los parámetros numéricos del juego en el que están tomando parte pueden predecir mejor el comportamiento de los demás y, de hecho, esto es lo que ha ocurrido. Como podemos observar en la Tabla 7, la

¹²² En el lenguaje experimental es lo que se suele conocer como *re-start effect*.

coordinación media entre aquellos que conocían los parámetros numéricos del juego es casi un 15% mayor que la media de aquellos que no los conocían (U de Mann-Whitney test; $z = -2,897$, $p = 0,004$, 2-tailed).

El desconocimiento acerca de quién tiene qué rol tiene una influencia distinta. Aunque existe una tendencia a alcanzar unos mayores niveles de coordinación en aquellos tratamientos donde no se conocían los roles, los niveles de coordinación alcanzados por aquellos que conocían su rol no son estadísticamente diferentes de los de aquellos que no los conocían (U de Mann-Whitney test; $z = -1,212$, $p = 0,226$, 2-tailed). Especialmente, cuando se conocía quién tenía cada rol pero todavía no se conocían los parámetros reales del juego, la coordinación parece ser más difícil, incluso cuando los participantes tenían la oportunidad de comunicarse sus *convenciones normativas* para elegir *convenciones de comportamiento*.

Tabla 7: Nivel de coordinación según condiciones informativas

		Conocen su rol	
		Sí	No
Conocen los parámetros reales del juego	Sí	74,38%	78,13%
	No	53,75%	69,38%

7.2.3. Conclusiones de primerl experimento

Tras interpretar una *convención de comportamiento* como la coordinación en un equilibrio estricto de un juego de coordinación, en este experimento estábamos interesados en comprender cómo la comunicación de principios de selección de equilibrios – *convenciones normativas* – puede mejorar la coordinación en una multiplicidad de juegos de coordinación. Al ordenar dichos principios los participantes

fueron más o menos informados sobre su situación específica de acuerdo con un diseño factorial de 2x2. Estos conocían o no conocían su rol y/o los parámetros del juego de coordinación asimétrica. Nuestros alentadores resultados fueron los siguientes:

- Con independencia del velo de ignorancia los participantes sugieren principalmente un principio de eficiencia.
- La mayoría de los participantes actúan obligados por su primer principio.
- La comunicación de los principios mejora la coordinación.
- El conocimiento del rol propio en un juego de coordinación asimétrica, junto al desconocimiento de los parámetros numéricos del juego es perjudicial para la coordinación.

Más allá de los objetivos específicos de esta tesis, estos resultados pueden ser ya muy útiles para el diseño institucional. Sugieren, por ejemplo, decidir sobre los principios de coordinación en el seno de las organizaciones después de establecer la estructura organizacional, incluyendo sus características numéricas, pero antes de asignar a los distintos miembros los roles en la organización. Bajo nuestro punto de vista, no existe duda de que generalmente confiamos en principios generales de coordinación para evitar la necesidad de coordinarnos de nuevo en cada interacción. Pero, dado que actuamos así, lo importante es que aprendamos cómo lo hacemos y hasta qué punto esto contribuye a mejorar la coordinación.

7.3. Segundo experimento

7.3.1. Diseño y conducción del segundo experimento

Este segundo experimento también fue realizado en el laboratorio del Max Planck Institute of Economics en Jena (Alemania). Para la realización del mismo, se reclutaron 96 estudiantes de licenciatura (procedentes de diferentes disciplinas) de la Universidad de Jena, mediante el sistema del reclutamiento *online* para experimentos económicos ORSEE 2.0 (Greiner, 2004). Se llevaron a cabo tres sesiones experimentales: cada sesión correspondía a un tratamiento distinto. En cada sesión experimental participaron 32 sujetos. La traducción de las instrucciones de este experimento (originalmente en alemán) puede ser encontrada en el Apéndice 2 al final del capítulo. Para la conducción informática del experimento se utilizó el programa z-Tree (Fischbacher, 1999). Los sujetos recibieron las instrucciones por escrito y, además, estas fueron leídas en voz alta para garantizar la existencia de conocimiento común acerca de las mismas. Ningún tipo de comunicación más allá del uso de los principios (como explicaremos más adelante) fue permitida. Además, los sujetos no podían identificar los miembros del grupo (*matching group*) en el que estaban encuadrados con los que estaban interactuando realmente. Al final de cada sesión experimental, se pagó a los sujetos en metálico de acuerdo con los pagos obtenidos en el juego (además de una suma fija de 2,5€ sólo por haberse presentado).

A cada sujeto le fue asignado un rol (A o B), pero los participantes sólo fueron informados de su propio rol al final del experimento, después de que todas las decisiones hubieran sido tomadas. Nótese que este procedimiento creaba una especie de

velo de ignorancia en el que todos conocían la estructura de los juegos, pero ninguno conocía los roles de los jugadores en cada juego concreto. Aunque *a priori* se podría pensar que al usar el mecanismo del *velo de ignorancia* los resultados quedarían afectados, el primer experimento comentado anteriormente mostraba que, en la clase de juegos que aquí estamos considerando, el conocimiento del rol de los jugadores no tiene un efecto estadísticamente significativo sobre la probabilidad de coordinar las acciones individuales. Además, como acabamos de ver, en dicho experimento detectamos que el conocimiento de los parámetros reales del juego sí que tiene una fuerte influencia sobre la posibilidad de alcanzar la coordinación. De este modo, en este segundo experimento, el conocimiento de los parámetros del juego es mantenido constante: todo los participantes, independientemente del tratamiento, los conocen.

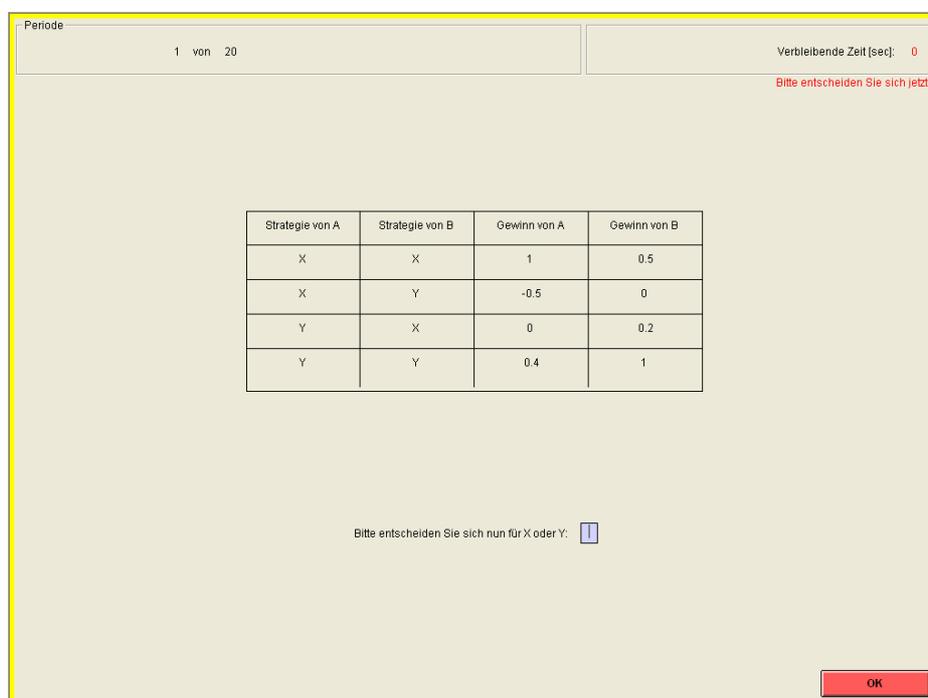
En todos los tratamientos cada participante interactuó en un grupo de cuatro sujetos, tomando parte en un total de 20 juegos simples. En total se habían diseñado 10 juegos distintos (ver el Apéndice 3)¹²³, cada uno de los cuales fue jugado por dos veces. En los diez primeros juegos (1-10) cada participante jugaba con un sujeto de su propio grupo, mientras que jugaba con otro sujeto distinto a lo largo del segundo grupo de diez juegos (11-20). Este procedimiento de “cambio de pareja” fue diseñado para poner a prueba la consistencia de los participantes en lo que al uso de los principios se refiere.

Los tratamientos diferían en términos de la información que los participantes recibían en relación con los principios. En el tratamiento B (tratamiento base), no se proporcionó ninguna información relativa a los principios. Por tanto, los participantes jugaron los 20 juegos solamente con la información referente a los parámetros de los

¹²³ Se trata de los mismos diez juegos utilizados en el primer experimento.

juegos (ver Figura 15). En el tratamiento P (tratamiento con principios), los participantes tenían que llevar a cabo dos tareas. En primer lugar, tenían que ordenar cuatro¹²⁴ principios y, en segundo, tenían que participar sucesivamente en 20 juegos similares a los del tratamiento B. La primera tarea implicaba rellenar una tabla con los principios propuestos (ver Figura 16). No tenían que sugerir un orden completo; el único requisito era proporcionar al menos un principio. Una vez que cada sujeto había proporcionado su ranking, era emparejado con uno de los miembros de su grupo y jugaban las diez primeras rondas de la sesión con esta persona. Después de estas diez primeras rondas, eran emparejados de nuevo con otro sujeto del mismo grupo y jugaban otras diez rondas. A lo largo de los 20 períodos, la única información que conocían, además de los parámetros del juego, era su propio ranking.

Figura 15: Pantalla de toma de decisiones del tratamiento B en el segundo experimento



¹²⁴ Como indicamos en la introducción de este capítulo, en aras de la simplicidad, en este segundo experimento el número de principios fue reducido a 4, habiendo sido eliminados aquéllos que apenas fueron seleccionados en el primer experimento. Los cuatro principios que mantuvimos fueron: *A* (*Alternating*), *L* (*Loss avoidance*), *R* (*Risk dominance*) y *S* (*Sum maximization*).

El tratamiento D (tratamiento con dictador) sólo difería del tratamiento P en un detalle. En este tratamiento, después de que todos los participantes habían elaborado su ranking y habían sido emparejados con un miembro de su grupo, uno de los dos rankings era seleccionado aleatoriamente por el ordenador (bien el ranking del jugador A o el ranking del jugador B). Después de que este ranking fuera mostrado a ambos jugadores (ver la Figura 17), éstos tomaban parte en los diez juegos simples. A continuación, los jugadores eran emparejados de nuevo con otro miembro del grupo, el mecanismo del *dictador aleatorio*¹²⁵ era utilizado de nuevo y se jugaban otros diez juegos.

Figura 16: Pantalla del ranking en el segundo experimento

Periode

1 von 20

Verbleibende Zeit [sec]: 12

Bitte entscheiden Sie sich für Ihre Rangfolge der Prinzipien!
Verwenden Sie dazu bitte die in den Instruktionen angegebenen Buchstaben!

Erstes Prinzip:

Zweites Prinzip:

Drittes Prinzip:

Viertes Prinzip:

OK

¹²⁵ El “dictador aleatorio” (*random dictator*) es un tipo de diseño empleado frecuentemente por los economistas experimentales. Una de sus principales ventajas es que supone un “incentivo compatible” para que los participantes proporcionen, en nuestro caso, su ranking verdadero, debido a que existe una probabilidad del 50% de que el ranking propio sea el seleccionado.

Figura 17: Pantalla de toma de decisiones del tratamiento D en el segundo experimento

Periode 1 von 20 Verbleibende Zeit [sec]: 19

Zufällig ausgewählte Rangfolge:

Erstes Prinzip: L
Zweites Prinzip: R
Drittes Prinzip:
Viertes Prinzip:

Strategie von A	Strategie von B	Gewinn von A	Gewinn von B
X	X	1	0.5
X	Y	-0.5	0
Y	X	0	0.2
Y	Y	0.4	1

Bitte entscheiden Sie sich nun für X oder Y:

En cada sesión se siguió el mismo protocolo. Los tratamientos sólo diferían en los términos de la variable que estaba siendo manipulada: la posibilidad de utilizar *principios normativos* para coordinar las expectativas que prescriben cómo coordinar las acciones (ver Tabla 8).

Tabla 8: Características de los tratamientos

	Tratamiento B	Tratamiento P	Tratamiento D
Número de sesiones	1	1	1
Períodos por sesión	20	20	20
Grupos (<i>matching groups</i>)	8	8	8
Participantes	32	32	32
Variable de tratamiento	Ninguna	Principios	Principios + Dictador
Coordinación	75%	65,31%	80,94%
Ganancia media	8,8€	8,1€	9,3€

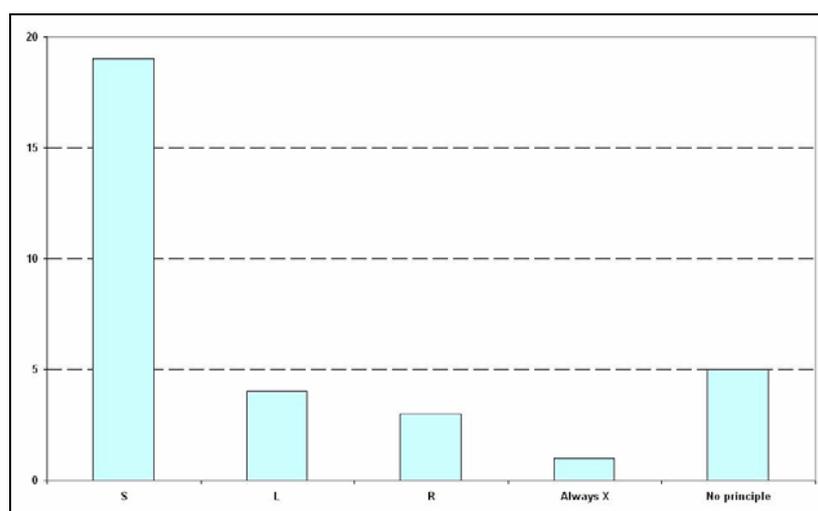
7.3.2. Resultados del segundo experimento

Con este segundo experimento intentaremos responder fundamentalmente a tres cuestiones distintas. En primer lugar, ¿realmente siguen los sujetos principios normativos al tomar parte en juegos de coordinación? En segundo lugar, ¿es el comportamiento de los sujetos consistente con los primeros principios? En tercer lugar, ¿mejora el uso de principios la coordinación? Para responder a la primera y a la segunda preguntas analizaremos los tratamientos independientemente. Para responder a la tercera cuestión compararemos los resultados de los tres tratamientos.

A. Tratamiento base

En el tratamiento B (tratamiento base) los participantes no fueron inducidos a utilizar ningún *principio normativo*. Sin embargo, como se puede comprobar en la Figura 18, el comportamiento de más de la mitad de los participantes fue consistente con el principio *S*, en concreto 19 de los 32 participantes seleccionaron de forma repetida la estrategia más eficiente. Además, 4 sujetos siguieron el principio *L*, 3 sujetos siguieron el principio *R* y 1 sujeto siempre escogió la estrategia *X*. Ningún sujeto alternó entre ambas estrategias (principio *A*) y el comportamiento de 5 sujetos no es susceptible de ser clasificado de acuerdo con ninguno de los ocho principios descritos en la segunda parte de la tesis.

Figura 18: Principios seguidos realmente en el tratamiento B



La preponderancia del comportamiento guiado por un criterio de eficiencia es consistente con los resultados obtenidos en el experimento anterior y, como ya poníamos de relieve, contraria a varios conocidos estudios experimentales, los cuales no encontraban dicha motivación por la eficiencia en sus investigaciones (ver Van Huyck, Battalio y Beil, 1990; 1991; Van Huyck, Cook y Battalio, 1997). A la vista de los resultados obtenidos en nuestros dos experimentos, nos preguntamos a qué puede deberse esta aparente contradicción. Volviendo a algo que ya poníamos de manifiesto anteriormente, la inmensa mayoría de los estudios experimentales sobre coordinación se centran en juegos simétricos, mientras que nuestra clase de juegos básica presenta varias asimetrías: no sólo ambos roles prefieren diferentes equilibrios (*convenciones de comportamiento*), sino que también sólo uno de los dos jugadores puede “perder dinero” (tener desutilidad) en cada juego particular. Desde luego que, en este segundo experimento, el mecanismo del *velo de ignorancia* impide que los participantes decidan en los juegos de acuerdo con su identidad de rol, pero este hecho no explica por sí solo la preferencia por el comportamiento eficiente sobre el comportamiento influido por cuestiones de riesgo, por ejemplo. Además, los resultados obtenidos en el primer

experimento sugieren que el conocimiento de los roles (ausencia del *velo de ignorancia*) no supone una diferencia estadísticamente significativa en los principios que guían las decisiones individuales en la clase de juegos aquí analizada.

Por tanto, el primer resultado de este segundo experimento tiene que ver con el predominio del principio de eficiencia incluso cuando ningún principio es enunciado explícitamente. La incidencia del principio *R* (*Risk dominance*) (como ha sido resaltado en los trabajos de Van Huyck, Battalio y Beil, 1990; 1991; Van Huyck, Cook y Battalio, 1997) o el principio *L* (*Loss avoidance*) (sugerido, entre otros, por Cachon y Camerer, 1996 y Rydval y Ortmann, 2005) es mínima. Por último, no tenemos evidencia del uso de ningún tipo de comportamiento en alternancia¹²⁶.

B. Tratamiento de principios

En el tratamiento P, los sujetos tenían que ordenar los principios pero no tenían que comunicar el ranking resultante a la persona con la que estaban interactuando. El tratamiento estaba diseñado para comprobar si, después de ordenar los principios, los sujetos se sentían obligados a seguir sus primeros principios. En el experimento anterior veíamos que sólo los primeros principios importaban y que el comportamiento real de los participantes nunca estaba determinado por otro principio que no fuera el primero. Por tanto, al analizar este tratamiento consideraremos los dos aspectos siguientes: ¿qué principios fueron ordenados como primeros?, y, ¿fue el comportamiento consistente con dichos principios?

¹²⁶ Utilizando datos experimentales procedentes de varias sesiones piloto, Schotter y Sopher (2003: n15) argumentan que la emergencia de comportamientos en alternancia en un juego de la “batalla de los sexos” requiere que el juego sea jugado repetidamente un número suficiente de rondas.

Las barras de color más oscuro de la figura 19 representan la frecuencia de los principios ordenados en primer lugar. De los cuatro principios, la opción más frecuente fue *Sum maximization* (*S*), seguido de *Loss avoidance* (*L*), *Alternating* (*A*) y *Risk dominance* (*R*). Este ranking se asemeja al ranking general obtenido en el primer experimento. Por el contrario, las barras más claras dibujan la frecuencia de los principios que fueron realmente seguidos por los participantes en sus acciones. Como podemos observar, no sólo *S* fue el principio más veces seleccionado como primer principio, sino que también fue el más ampliamente seguido. De hecho, los resultados del tratamiento P no son muy diferentes a los obtenidos en el primer tratamiento; en este caso 17, en vez de 19 sujetos, sobre un total de 32 siguieron el principio de eficiencia. Además, aquellos que seleccionaron el principio *S* mostraron un alto nivel de consistencia con dicho principio. La Figura 20 representa los niveles de consistencia de aquellos que seleccionaron cada una de las cuatro convenciones normativas como primer principio.

Figura 19: Principios ordenados como primeros y principios realmente seguidos en el tratamiento P

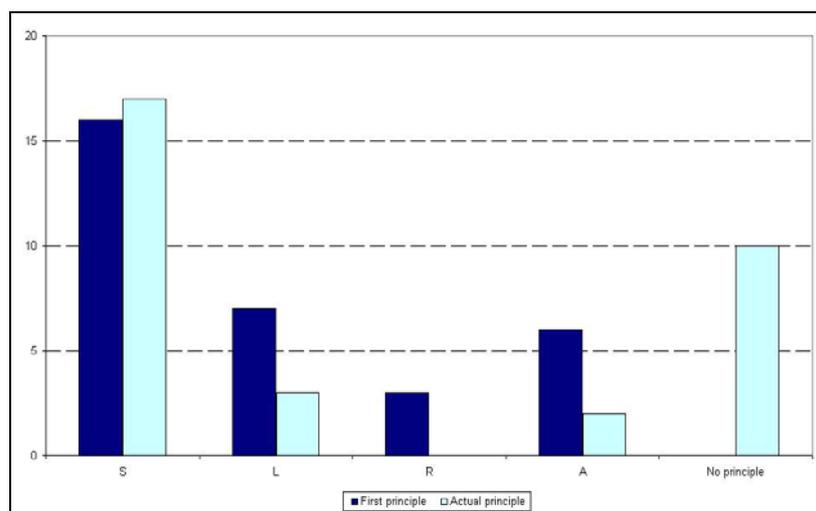
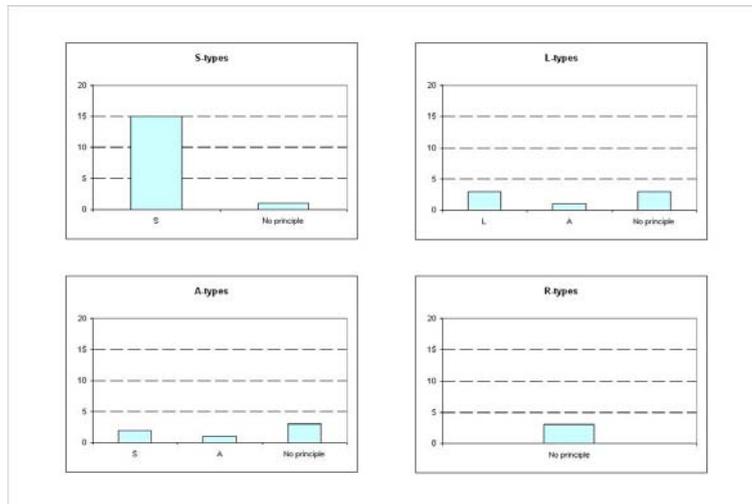


Figura 20: Consistencia con el comportamiento prescrito en el tratamiento P



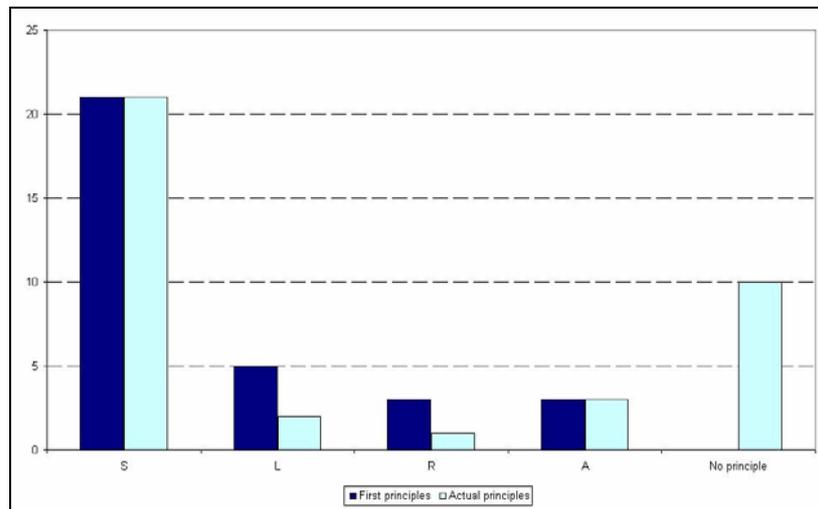
Todos los sujetos que seleccionaron *S* como su primer principio, menos uno, realmente lo siguieron. Por el contrario, sólo 4 de los 16 participantes que seleccionaron cualquier otro primer principio se mostraron consistentes con éste. La consistencia con el principio *S* es estadísticamente diferente a la consistencia con los demás principios (U de Mann-Whitney test; $z = -3,897$, $p = 0,001$). Este resultado coincide con la diferencia en las pautas de consistencia entre el principio *S* y los demás principios obtenida en el primero de los experimentos. Por tanto, el segundo resultado de este experimento pone de manifiesto que la consistencia con los principios depende críticamente del principio que ha sido seleccionado. En este sentido, la consistencia con el criterio de eficiencia es especialmente alta.

C. Tratamiento con dictador

La Figura 21 muestra las frecuencias de los principios que fueron ordenados en primer lugar, así como los principios que fueron realmente seguidos en el tratamiento D. De nuevo, *Sum maximization* (21 sujetos) fue el principio que más veces fue ordenado en

primer lugar, seguido de *Loss avoidance* (5 sujetos), *Risk dominance* (3 sujetos) y *Alternating* (3 sujetos).

Figura 21: Principios ordenados en primer lugar y principios realmente seguidos en el tratamiento D



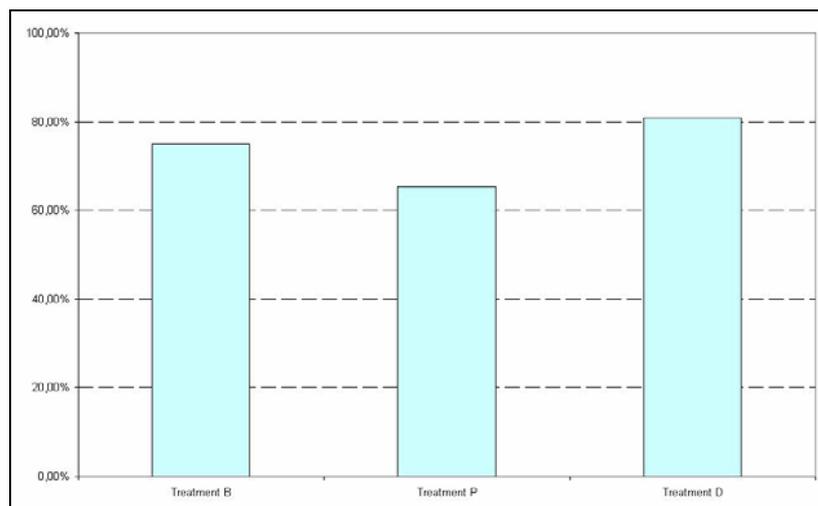
Una vez que hemos verificado que *Sum maximization* sigue desempeñando un papel predominante incluso en un entorno estratégico de elección de principios, nos preguntamos si el comportamiento de los sujetos fue consistente con el primer principio sugerido por el *dictador aleatorio*. Aquí tenemos que considerar 64 comportamientos diferentes, dado que cada sujeto tomó una decisión a la luz de la propuesta de cada uno de los *dictadores aleatorios* a las que se enfrentó (una durante las primeras diez rondas y otra durante las diez segundas). 44 de los 64 comportamientos fueron consistentes con el primer principio propuesto por el dictador aleatorio; 8 de los 64 se desviaron de dicha propuesta para seguir el principio *S* y 12 comportamientos no pueden ser clasificados de acuerdo con alguno de los principios considerados. En general, los participantes fueron consistentes con la prescripción del *dictador* y la única razón para desviarse del mismo era la búsqueda de la eficiencia. El resultado es que cuando los participantes podían ordenar los principios estratégicamente, principalmente ordenaron en primer lugar el

principio *S* y fueron mayoritariamente consistentes con él. Por tanto, como tercer resultado de este experimento diremos que la posibilidad de ordenar los principios estratégicamente no altera la preponderancia de y la consistencia con el principio de eficiencia.

D. Coordinación

¿Cómo afectaron los comportamientos descritos anteriormente a la coordinación? La Figura 22 presenta los niveles de coordinación alcanzados en cada uno de los tres tratamientos. La coordinación media en este experimento fue del 68,54%, casi idéntica al 68,78% obtenido en el primer experimento. Aunque existen diferencias estadísticamente significativas entre los tratamientos (Test de Kruskal-Wallis; $\chi^2 = 12,667$, $d.f. = 2$, $p = 0,047$), los niveles de coordinación no se incrementaron linealmente como se podía esperar.

Figura 22: Coordinación según tratamiento



Los participantes lo hicieron relativamente bien en el tratamiento base, en el que ningún tipo de principio normativo fue inducido, alcanzando una tasa de coordinación media bastante alta (75%). Sin embargo, en contra de un argumento mantenido en el primer experimento, la tasa de coordinación disminuyó en más de 10 puntos, hasta el 65,31% en el tratamiento P, en el que fueron sugeridos los cuatro principios mencionados. El nivel de coordinación se incrementó hasta el 80,94% sólo cuando el mismo ranking era compartido por los dos sujetos que estaban interactuando (tratamiento D).

Para interpretar estos resultados debemos tener en cuenta una de las características definitoria de la noción de *convención normativa* planteadas en la primera parte de la tesis: su carácter colectivo. En este sentido, para que un principio funcione ambos jugadores tienen que ser mutuamente conscientes de que dicho principio debería ser aplicado. Así, mientras que la introducción de los principios en el tratamiento D mejora el nivel general de coordinación, la introducción de los principios en el tratamiento P no hace lo propio. Sin duda, este hallazgo se encuentra en la base de nuestra propuesta acerca de cómo los *principios* o *convenciones normativas* facilitan la coordinación en *convenciones de comportamiento*. Sin embargo, esta importante conclusión no nos dice nada a propósito de por qué la coordinación en el tratamiento P fue significativamente menor que en el tratamiento base. En nuestra opinión, la razón para este último hecho es que, al seguir mayoritariamente un criterio de eficiencia, los sujetos que participaron en el tratamiento B lograron coordinarse con bastante frecuencia en la convención eficiente. Por tanto, la focalidad del equilibrio eficiente en dicho tratamiento contribuyó notablemente a que se alcanzara un elevado nivel de coordinación. Sin embargo, cuando otros principios fueron sugeridos, el comportamiento general no estuvo tan concentrado

en un criterio de eficiencia y la mitad de los sujetos eligieron principios distintos a *S*. La implementación de la estrategia prescrita por dichos principios fue desastrosa (sólo 4 de los 16 sujetos implementaron los principios correctamente). Como resultado, en general, los sujetos no podían beneficiarse en este caso de una alta coordinación espontánea en el resultado eficiente del juego.

Teniendo en cuenta todo lo anterior, el cuarto resultado de este segundo experimento nos dice que para que los principios funcionen es indispensable que sean mutuamente conocidos y aceptados por aquellos que participan de los mismos.

7.3.3. Conclusiones del segundo experimento

Este segundo experimento fue diseñado para clarificar algunos de los resultados obtenidos en el primero y seguir, de este modo, avanzando en la comprobación de las hipótesis teóricas derivadas de nuestro trabajo en las dos primeras partes de la tesis¹²⁷. En el mismo hemos eludido el problema informativo presente en el primer experimento, manteniendo un conocimiento constante de los parámetros numéricos de los juegos y un desconocimiento permanente del rol propio. Además hemos introducido dos tratamientos nuevos para estudiar dos aspectos concretos de nuestra teoría. En primer lugar, el tratamiento B supone un grupo de control real, en el que los sujetos se ven sometidos al mismo tipo de diseño, aunque no disponen de ninguna información acerca de los principios normativos. En segundo lugar, el tratamiento D nos ha servido para

¹²⁷ De nuevo, recordamos la advertencia que se ya se realizaba en una nota anterior acerca del proceso de realización de los experimentos. En este sentido, el diseño original de este segundo experimento era más ambicioso y se acercaba más al abordaje empírico de los aspectos normativos de la noción de convención social. Sin embargo, debido a la dificultad de dicha empresa, se pospuso la contemplación de aspectos más genuinamente sociológicos – como el diseño de mecanismos de reprobación social – para experimentos posteriores, y se decidió reforzar algunas de los resultados derivados del primer experimento.

comprobar cómo el conocimiento común de los principios incrementa la probabilidad de que éstos sean implementados correctamente y, de este modo, aumenta los niveles medios de coordinación. ¿Qué conclusiones nos ha permitido sacar en claro este nuevo diseño?

Las conclusiones principales obtenidas en este segundo experimento han sido las siguientes:

- Incluso cuando no se sugiere ningún principio, los participantes siguen principalmente un criterio de eficiencia en la clase de juegos que aquí estamos considerando.
- Este resultado es robusto tanto a la sugerencia de otros principios, como a la posibilidad de emplear éstos estratégicamente.
- La consistencia con los principios depende críticamente del principio seleccionado. Los participantes son especialmente consistentes con el principio de eficiencia.
- Para que los principios funcionen adecuadamente, es necesario que sean conocidos y aceptados conjuntamente.

En definitiva, estos resultados nos permiten seguir avanzando en el conocimiento acerca del uso de principios normativos en la toma de decisiones en entornos de coordinación asimétrica. Dado que los resultados presentados son bastante consistentes

con los obtenidos en el primer experimento, consideramos que el análisis del uso de principios, el estudio de las pautas de consistencia de los comportamientos con respecto a cada principio en particular y la evaluación de los resultados de aplicar estos principios son aspectos cruciales para entender cómo las personas solucionan problemas recurrentes de coordinación. Para ello, y contrariamente a la práctica generalizada en la teoría económica, no sólo deberíamos prestar atención a las pautas de comportamiento que realmente observamos, sino también a los criterios normativos que dan soporte a dichas pautas.

7. 4. Conclusiones

En la introducción de este capítulo planteábamos dos objetivos. En primer lugar, nos preguntábamos si el modelo teórico desarrollado en la segunda parte de la tesis podía ser implementado en el laboratorio y, así, permitirnos su estudio con sujetos reales. En segundo lugar, nos cuestionábamos hasta qué punto el diseño abstracto de la segunda parte tenía que ver con el comportamiento real de los sujetos en una situación experimental.

Con respecto a la primera de las cuestiones tenemos que decir que no hemos encontrado demasiados problemas en trasladar el modelo teórico evolutivo desarrollado en el capítulo 5 a una situación de laboratorio; eso sí, una situación estática que apenas nos permite decir algo acerca de las posibles extensiones dinámicas del modelo. Hay varios datos que podemos usar para mostrar hasta qué punto los sujetos captaron la lógica del diseño que pretendíamos implementar. Por ejemplo, tanto la selección de “buenos principios” (principios evolutivamente estables), como las tasas de consistencia

con los principios demuestran la persecución del objetivo de maximización del pago esperado por parte de los sujetos¹²⁸. Como resultado, el nivel medio de coordinación se situó en torno al 70% en ambos experimentos, bastante por encima del 50% esperado si los participantes hubieran seleccionado en todo momento sus estrategias de forma aleatoria. La coincidencia en los niveles de coordinación entre ambos experimentos también es una muestra de la consistencia entre el diseño de ambos, pudiendo entender el segundo como una simple extensión y un refinamiento del primero.

Dicho lo anterior, nos preguntamos: ¿qué aporta este estudio experimental a los resultados obtenidos en la tesis hasta el momento?

En primer lugar, refuerza nuestra distinción analítica entre *convenciones normativas* y *convenciones de comportamiento*. Como hemos tenido oportunidad de comprobar en el segundo experimento, las *convenciones normativas* no sólo se muestran eficaces para prescribir *convenciones de comportamiento* concretas cuando son presentadas de modo explícito, sino también cuando no se hace ninguna referencia a las mismas, como ocurría en el tratamiento base del segundo experimento. Este resultado apunta hacia el hecho de que las decisiones estratégicas están fuertemente influenciadas por criterios normativos, algo que, sin embargo, necesita ser estudiado en entornos distintos para que pueda ser generalizado. Si retomamos la metáfora de “el juego de la vida” (Güth y Papel, 2002) a la que hacíamos referencia en el capítulo 5 y nuestra pretensión es postular nuestro diseño como un mecanismo teórico general, éste deberá ser puesto a prueba en situaciones completamente distintas. Por ejemplo, deberíamos probar la validez de nuestros principios normativos en situaciones mucho más conflictivas que las

¹²⁸ Como poníamos de manifiesto en el capítulo anterior, el supuesto de que las decisiones individuales deben depender de los incentivos materiales propuestos es central para el tipo de diseño experimental que estamos empleando.

que se han dado en nuestros experimentos. Del mismo modo, deberíamos seguir explorando la emergencia de otros principios que se unan a los ocho considerados. Otras innovaciones posibles tendrían que ver con la posibilidad de permitir el aprendizaje de los sujetos a través de la implementación de juegos repetidos que incorporen información pública sobre el resultado de cada juego particular, o, en un sentido muy weberiano, cómo afectaría al comportamiento de los sujetos la posibilidad de imponer sanciones a los que se desvían de “su primer principio”¹²⁹.

En segundo lugar, más allá de nuestra distinción analítica, otro de los resultados fuertes que hemos encontrado es la consistencia mantenida por los participantes con los principios seleccionados. Este hecho nos anima a seguir explorando la “función cognitiva” de las convenciones sociales, en cuanto mecanismos de coordinación social. Además, esto nos debe llamar la atención sobre las posibles inercias en el seguimiento de las convenciones sociales; unas inercias que, en bastantes ocasiones, podrían ir contra cualquier criterio de racionalidad. Éste es un hecho que puede ser explorado en investigaciones que empleen situaciones más conflictivas y en las que sea más fácil distinguir cuándo los sujetos están siendo consistentes con el criterio de racionalidad individual y cuándo no lo están siendo.

En tercer lugar, existe todo un terreno para explorar en relación con los problemas informativos relacionados con el seguimiento de las convenciones sociales. En nuestros experimentos hemos manipulado la información referente a tres variables: los parámetros numéricos del juego, el rol de cada participante y los propios principios normativos. Una extensión natural de esta investigación sería explorar todas las

¹²⁹ Ambas extensiones se encuentran en la agenda del programa de investigación abierto en el *Strategic Interaction Group* al que hacíamos referencia con anterioridad.

combinaciones posibles de información o desinformación respecto de estas tres variables. Además, también podría manipularse la información relativa a la fuente de legitimidad de los principios, en el sentido de diseñar una fase previa en la que éstos sean diseñados colectivamente, sean sometidos a votación, etc¹³⁰.

Por último, no queremos dejar pasar un resultado que, si bien no tiene tanto que ver con nuestras hipótesis teóricas de partida, se muestra especialmente relevante en ambos experimentos: la preeminencia del criterio de eficiencia. Al contrario que los resultados anteriores, este último resultado no se deriva directamente de nuestro diseño y, hasta cierto punto, nos conduce a plantearnos una posible regularidad empírica en el seno de la población objeto de estudio. Si esto fuera así, sólo habría una forma de saberlo: a través del uso de comparaciones entre poblaciones. Por tanto, se podrían abordar cuestiones del tipo: ¿tiene que ver el comportamiento eficiente observado con que el experimento haya sido llevado a cabo en Alemania? ¿Tiene que ver con la experiencia de los participantes en sesiones experimentales anteriores?, etc. Como señalábamos en el capítulo 6, sólo mediante la comparación entre muestras extraídas de poblaciones que, *ex ante*, son consideradas distintas podemos dilucidar *ex post* el sentido de dicha diferencia.

En definitiva, el uso del método experimental nos ha permitido establecer una serie de resultados estadísticamente significativos que refuerzan nuestro modelo teórico, así como vislumbrar nuevas posibles vías de estudio de la noción de convención social.

¹³⁰ Sobre este último punto – la fuente de legitimidad de los principios, en especial su vinculación con procedimientos de voto – está ya siendo explorado, aunque aún con resultados poco concluyentes. Ver Miller y Vanberg (2006).

Apéndice 1: Instrucciones del primer experimento^{131,132}

Bienvenido y gracias por participar en este experimento. Por favor, lee las instrucciones cuidadosamente. De ahora en adelante, cualquier comunicación con los demás participantes está prohibida. De otra forma, tendríamos que excluirte del experimento y de todos los pagos. Si tienes cualquier pregunta, por favor levanta la mano. Iremos hacia tu sitio y responderemos tus preguntas individualmente.

Durante el experimento utilizamos ECU (unidad de moneda experimental) en vez de Euro. Tu ganancia experimental será calculada en ECU. Al final del experimento los ECU que hayas ganado serán convertidos en Euros (1 ECU = 0,5 Euros) y la cantidad obtenida se te pagará en metálico.

En este experimento, vas a interactuar con otro participante y cada uno de vosotros desempeñará un rol que, para simplificar, es bien A ó B. Ambos participantes pueden elegir entre dos estrategias X_A ó Y_A (para el participante A) y X_B ó Y_B (para el participante B).

Estas estrategias conducen a los siguientes pagos:

¹³¹ La transcripción que aquí se realiza tanto de estas instrucciones, como de las correspondientes al segundo experimento, es una traducción al castellano de las originales en alemán.

¹³² En este apéndice proporcionamos las instrucciones del tratamiento N. Las instrucciones del resto de tratamientos son sólo parcialmente diferentes. Las diferencias ya han sido puestas de manifiesto en el subapartado 7.2.1.

Pago para el participante A	Pago para el participante B	Si el participante	
		A selecciona	B selecciona
1	a	X_A	X_B
d	0	X_A	Y_B
0	b	Y_A	X_B
c	1	Y_A	Y_B

Los valores a, b, c y d cumplirán las siguientes condiciones:

$$1 > a > b > 0 \quad \text{y} \quad 1 > c > 0 > d > -1$$

Si X_i ó Y_i es mejor para ti depende de la decisión de tu compañero en esta ronda, es decir qué estrategia selecciona. Por tanto, tú deberías estar interesado en acordar colectivamente una constelación de estrategias (X_A, X_B) , (X_A, Y_B) , (Y_A, X_B) , (Y_A, Y_B) .

Para llegar a dicho acuerdo, te proponemos principios mediante los cuales puedes seleccionar una de las cuatro constelaciones de estrategias. (Estos principios serán explicados en la página siguiente).

En primer lugar, tienes que ordenar los principios. No tienes por qué (aunque puedes) ordenar todos los principios. Es suficiente con que indiques tu prioridad. Entonces, conociendo ambos rankings (el tuyo y el del participante con el que estás interactuando), puedes elegir bien la estrategia X o la estrategia Y. En cada ronda, aprenderás, qué rol desempeñas (A ó B) así como los valores de los parámetros a, b, c y d. Después, juegas otras nueve rondas del juego (es decir, en total juegas 10 rondas). En cada ronda, los valores de los parámetros a, b, c, d y tu rol (A ó B) cambiarán. El ranking que has seleccionado permanece igual.

Después de las 10 rondas, serás emparejado aleatoriamente con un nuevo compañero con el que interaccionas durante otras 10 rondas. Mantendrás tu ranking y conocerás el ranking de tu compañero. En cada ronda, conocerás de nuevo qué rol desempeñas (A ó B), así como los valores de los parámetros a, b, c y d. Al final de estas 20 rondas, te informaremos de los valores de los parámetros a, b, c y d en las rondas individuales, si has sido participante A ó B en las rondas individuales y tu pago en cada ronda.

Para compensar posibles pérdidas, cada participante recibe una cantidad básica de 4 ECU (2 Euros). Si acabaras el experimento con un balance negativo, te pediremos que lo compenses haciendo algunos trabajos esporádicos (a 32 ECU por hora) en nuestro instituto. Si no estás de acuerdo, por favor levanta la mano. Entonces tienes que abandonar el experimento.

Ahora te presentaremos los principios para seleccionar una de las cuatro constelaciones de estrategias:

A

En las rondas pares jugamos X y en las rondas impares Y, de modo que ambos participantes confían en la elección de X ó Y.

G

Usamos la estrategia que es mejor para el otro jugador si éste juega ambas estrategias con la misma probabilidad, es decir el participante A debería jugar Y_A y el participante B debería jugar X_B .

I

Usamos ambas estrategias con tal probabilidad que al otro no le importa lo que elige.

L

Para evitar pérdidas, jugamos Y, es decir Y_A, Y_B .

M

Usamos el principio que tiene el mejor de los resultados posibles, es decir el participante A debería jugar siempre Y_A y el participante B siempre X_B .

R

Usamos la constelación de estrategias donde cada participante reacciona óptimamente a la elección del otro y el producto del correspondiente incremento del pago resultante de un cambio de estrategia es máximo.

Para el participante A, el pago se incrementa en $1 - 0 = 1$ para X_B y $c - d$ para Y_B . Para el participante B el incremento es $a - 0 = a$ para X_A y en $1 - b$ para Y_A . Como sólo en (X_A, X_B) y (Y_A, Y_B) ambos reaccionan óptimamente a la elección del otro, los participantes deberían usar (X_A, X_B) si $a > (c - d)(1 - b)$ y (Y_A, Y_B) si $a < (c - d)(1 - b)$.

S

Usamos la combinación de estrategias para la que la suma de pagos es máxima, es decir los participantes juegan (X_A, X_B) si $a > c$ y (Y_A, Y_B) si $a < c$.

U

Usamos la estrategia que es mejor si el otro usa ambas estrategias con la misma probabilidad, es decir el participante A debería jugar X_A si $1 + d > c$ y Y_A si $1 + d < c$.
El participante B debería jugar X_B si $a + b > 1$ y Y_B si $a + b < 1$.

Round	Gains		
	Strategy	Gains A	Gains B
1	X_A, X_B	1	0,5
	X_A, Y_B	-0,5	0
	Y_A, X_B	0	0,2
	Y_A, Y_B	0,4	1

Round	Gains		
	Strategy	Gains A	Gains B
2	X_A, X_B	1	0,8
	X_A, Y_B	-0,1	0
	Y_A, X_B	0	0,1
	Y_A, Y_B	0,7	1

Round	Gains		
	Strategy	Gains A	Gains B
3	X_A, X_B	1	0,9
	X_A, Y_B	-0,3	0
	Y_A, X_B	0	0,8
	Y_A, Y_B	0,6	1

Round	Gains		
	Strategy	Gains A	Gains B
4	X_A, X_B	1	0,7
	X_A, Y_B	-0,2	0
	Y_A, X_B	0	0,6
	Y_A, Y_B	0,5	1

Round	Gains		
	Strategy	Gains A	Gains B
5	X_A, X_B	1	0,6
	X_A, Y_B	-0,4	0
	Y_A, X_B	0	0,5
	Y_A, Y_B	0,7	1

Round	Gains		
	Strategy	Gains A	Gains B
6	X_A, X_B	1	0,3
	X_A, Y_B	-0,8	0
	Y_A, X_B	0	0,2
	Y_A, Y_B	0,6	1

Round	Gains		
	Strategy	Gains A	Gains B
7	X_A, X_B	1	0,7
	X_A, Y_B	-0,6	0
	Y_A, X_B	0	0,2
	Y_A, Y_B	0,9	1

Round	Gains		
	Strategy	Gains A	Gains B
8	X_A, X_B	1	0,5
	X_A, Y_B	-0,1	0
	Y_A, X_B	0	0,2
	Y_A, Y_B	0,5	1

Round	Gains		
	Strategy	Gains A	Gains B
9	X_A, X_B	1	0,2
	X_A, Y_B	-0,2	0
	Y_A, X_B	0	0,1
	Y_A, Y_B	0,5	1

Round	Gains		
	Strategy	Gains A	Gains B
10	X_A, X_B	1	0,3
	X_A, Y_B	-0,2	0
	Y_A, X_B	0	0,2
	Y_A, Y_B	0,6	1

Apéndice 2: Instrucciones del segundo experimento

Instrucciones generales (comunes a todos los tratamientos)

Bienvenido y gracias por participar en este experimento. Por favor, lee las instrucciones cuidadosamente. De ahora en adelante, cualquier comunicación con los demás participantes está prohibida. De otra forma, tendríamos que excluirte del experimento y de todos los pagos. Si tienes cualquier pregunta, por favor levanta la mano. Iremos hacia tu sitio y responderemos tus preguntas individualmente.

Durante el experimento utilizamos ECU (unidad de moneda experimental) en vez de Euro. Tu ganancia experimental será calculada en ECU. Al final del experimento los ECU que hayas ganado serán convertidos en Euros (1 ECU = 0,5 Euros) y la cantidad obtenida se te pagará en metálico.

En este experimento, vas a interactuar con otro participante y cada uno de vosotros desempeñará un rol que, para simplificar, es bien A ó B. Ambos participantes pueden elegir entre dos estrategias X_A ó Y_A (para el participante A) y X_B ó Y_B (para el participante B). Sin embargo, sólo conocerás el rol que te ha sido asignado al final del experimento, después de haber tomado todas tus decisiones.

Estas estrategias conducen a los siguientes pagos:

Pago para el participante A	Pago para el participante B	Si el participante	
		A selecciona	B selecciona
1	a	X_A	X_B
d	0	X_A	Y_B
0	b	Y_A	X_B
c	1	Y_A	Y_B

Los valores a, b, c y d cumplirán las siguientes condiciones:

$$1 > a > b > 0 \quad \text{y} \quad 1 > c > 0 > d > -1$$

Si X_i ó Y_i es mejor para ti depende de la decisión de tu compañero en esta ronda, es decir qué estrategia selecciona. Por tanto, tú deberías estar interesado en acordar colectivamente una constelación de estrategias (X_A, X_B) , (X_A, Y_B) , (Y_A, X_B) , (Y_A, Y_B) .

Instrucciones específicas para el tratamiento B

Puedes elegir bien la estrategia X o bien la estrategia Y. En cada ronda conocerás los valores de los parámetros a, b, c y d. Entonces, jugarás otras nueve rondas del juego (es decir, en total juegas 10 rondas). En cada ronda, los valores de los parámetros a, b, c y d cambiarán.

Después de las 10 rondas, serás emparejado aleatoriamente con un nuevo compañero con quien interactuarás durante otras 10 rondas. En cada ronda, conocerás de nuevo los valores de los parámetros a, b, c y d. Al final de las 20 rondas, te informaremos sobre los valores de los parámetros a, b, c y d en las rondas individuales, si has sido participante A o participante B en las rondas individuales y sobre tu pago en cada ronda.

Para compensar posibles pérdidas, cada participante recibe una cantidad básica de 5 ECU (2,5 Euros). Si acabaras el experimento con un balance negativo, te pediremos que lo compenses haciendo algunos trabajos esporádicos (a 32 ECU por hora) en nuestro instituto. Si no estás de acuerdo, por favor levanta la mano. Entonces tienes que abandonar el experimento.

Antes de que el experimento comience, te pediremos que respondas correctamente a un conjunto de preguntas dirigidas a mejorar tu comprensión del experimento. Por favor rellena el cuestionario cuidadosamente.

Instrucciones específicas para el tratamiento P

Para llegar a dicho acuerdo, te proponemos principios mediante los cuales puedes seleccionar una de las cuatro constelaciones estratégicas. (Estos principios serán explicados en la página siguiente).

En primer lugar, tienes que ordenar los principios. No tienes por qué (aunque puedes) ordenar todos los principios. Es suficiente con que indiques tu prioridad. Entonces, conociendo tu ranking, puedes elegir bien la estrategia X o la estrategia Y. En cada ronda, aprenderás qué rol desempeñas (A ó B) así como los valores de los parámetros a, b, c y d. Después, juegas otras nueve rondas del juego (es decir, en total juegas 10 rondas). En cada ronda, los valores de los parámetros a, b, c y d cambiarán. El ranking que has seleccionado permanece igual.

Después de las 10 rondas, serás emparejado aleatoriamente con un nuevo compañero con el que interaccionas durante otras 10 rondas. Mantendrás tu ranking. En cada ronda, conocerás de nuevo los valores de los parámetros a, b, c y d. Al final de estas 20 rondas, te informaremos de los valores de los parámetros a, b, c y d en las rondas individuales, si has sido participante A ó B en las rondas individuales y tu pago en cada ronda.

Para compensar posibles pérdidas, cada participante recibe una cantidad básica de 5 ECU (2,5 Euros). Si acabaras el experimento con un balance negativo, te pediremos que lo compenses haciendo algunos trabajos esporádicos (a 32 ECU por hora) en nuestro instituto. Si no estás de acuerdo, por favor levanta la mano. Entonces tienes que abandonar el experimento.

Ahora te presentaremos los principios para seleccionar una de las cuatro constelaciones de estrategias:

A

En las rondas pares jugamos X y en las rondas impares Y, de modo que ambos participantes confían en la elección de X ó Y.

L

Para evitar pérdidas, jugamos Y, es decir Y_A , Y_B .

R

Usamos la constelación de estrategias donde cada participante reacciona óptimamente a la elección del otro y el producto del correspondiente incremento del pago resultante de un cambio de estrategia es máximo.

Para el participante A, el pago se incrementa en $1 - 0 = 1$ para X_B y $c - d$ para Y_B . Para el participante B el incremento es $a - 0 = a$ para X_A y en $1 - b$ para Y_A . Como sólo en (X_A, X_B) y (Y_A, Y_B) ambos reaccionan óptimamente a la elección del otro, los participantes deberían usar (X_A, X_B) si $a > (c - d)(1 - b)$ y (Y_A, Y_B) si $a < (c - d)(1 - b)$.

S

Usamos la combinación de estrategias para la que la suma de pagos es máxima, es decir los participantes juegan (X_A, X_B) si $a > c$ y (Y_A, Y_B) si $a < c$.

Antes de que el experimento comience, te pediremos que respondas correctamente a un conjunto de preguntas dirigidas a mejorar tu comprensión del experimento. Por favor rellena el cuestionario cuidadosamente.

Instrucciones específicas para el tratamiento D

Para llegar a dicho acuerdo, te proponemos principios mediante los cuales puedes seleccionar una de las cuatro constelaciones estratégicas. (Estos principios serán explicados en la página siguiente).

En primer lugar, tienes que ordenar los principios. No tienes porqué (aunque puedes) ordenar todos los principios. Es suficiente con que indiques tu prioridad. Entonces, un ranking será seleccionado aleatoriamente (bien tu ranking o bien el ranking de tu compañero) y este ranking será mostrado a tu compañero y a ti. Después de esto, puedes elegir bien la estrategia X o bien la estrategia Y. En cada ronda, aprenderás los valores de los parámetros a, b, c y d. Después, juegas otras nueve rondas del juego (es decir, en total juegas 10 rondas). En cada ronda, los valores de los parámetros a, b, c, d y d cambiarán. El ranking que has seleccionado permanece igual.

Después de las 10 rondas, serás emparejado aleatoriamente con un nuevo compañero con el que interaccionas durante otras 10 rondas. Uno de los dos ranking será seleccionado aleatoriamente de nuevo (bien tu ranking o el ranking de tu nuevo compañero). En cada ronda, conocerás de nuevo los valores de los parámetros a, b, c y d. Al final de estas 20 rondas, te informaremos de los valores de los parámetros a, b, c y d en las rondas individuales, si has sido participante A ó B en las rondas individuales y tu pago en cada ronda.

Para compensar posibles pérdidas, cada participante recibe una cantidad básica de 5 ECU (2,5 Euros). Si acabaras el experimento con un balance negativo, te pediremos que lo compenses haciendo algunos trabajos esporádicos (a 32 ECU por hora) en nuestro instituto. Si no estás de acuerdo, por favor levanta la mano. Entonces tienes que abandonar el experimento.

Ahora te presentaremos los principios para seleccionar una de las cuatro constelaciones de estrategias:

(La descripción de los principios, la cual es similar a la presentada en el tratamiento P, es introducida aquí de nuevo).

Antes de que el experimento comience, te pediremos que respondas correctamente a un conjunto de preguntas dirigidas a mejorar tu comprensión del experimento. Por favor rellena el cuestionario cuidadosamente.

Apéndice 3: Juegos realmente jugados en el segundo experimento

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 1	X _A , X _B	1	0.5
	X _A , Y _B	-0.5	0
	Y _A , X _B	0	0.2
	Y _A , Y _B	0.4	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 5	X _A , X _B	1	a
	X _A , Y _B	D	0
	Y _A , X _B	0	b
	Y _A , Y _B	C	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 2	X _A , X _B	1	0.8
	X _A , Y _B	-0.1	0
	Y _A , X _B	0	0.1
	Y _A , Y _B	0.7	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 6	X _A , X _B	1	a
	X _A , Y _B	D	0
	Y _A , X _B	0	b
	Y _A , Y _B	C	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 3	X _A , X _B	1	0.9
	X _A , Y _B	-0.3	0
	Y _A , X _B	0	0.8
	Y _A , Y _B	0.6	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 7	X _A , X _B	1	0.7
	X _A , Y _B	-0.6	0
	Y _A , X _B	0	0.2
	Y _A , Y _B	0.9	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 4	X _A , X _B	1	0.7
	X _A , Y _B	-0.2	0
	Y _A , X _B	0	0.6
	Y _A , Y _B	0.5	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 8	X _A , X _B	1	0.5
	X _A , Y _B	-0.1	0
	Y _A , X _B	0	0.2
	Y _A , Y _B	0.5	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 5	X _A , X _B	1	0.6
	X _A , Y _B	-0.4	0
	Y _A , X _B	0	0.5
	Y _A , Y _B	0.7	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 9	X _A , X _B	1	0.2
	X _A , Y _B	-0.2	0
	Y _A , X _B	0	0.1
	Y _A , Y _B	0.5	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 6	X _A , X _B	1	0.3
	X _A , Y _B	-0.8	0
	Y _A , X _B	0	0.2
	Y _A , Y _B	0.6	1

Rounds	Payoffs		
	Behavior	Payoffs A	Payoffs B
Period 10	X _A , X _B	1	0.3
	X _A , Y _B	-0.2	0
	Y _A , X _B	0	0.2
	Y _A , Y _B	0.6	1

CAPÍTULO 8

CONCLUSIONES GENERALES

Esta tesis comenzaba poniendo de relieve una llamativa laguna en la investigación social contemporánea: aunque es imposible concebir una sociedad sin convenciones, las distintas ciencias sociales han sido incapaces de ponerse de acuerdo sobre las características que definen a éstas, así como los mecanismos que posibilitan su existencia. Por ello, el objetivo transversal del presente trabajo ha sido la elaboración de una propuesta teórica que de cuenta de la complejidad de la noción de convención social. Para ello, hemos dividido nuestro esfuerzo en tres grandes bloques que representan otras tantas aproximaciones metodológicas a este objeto de estudio. Las conclusiones que se presentan a continuación seguirán el mismo esquema desarrollado en el resto de la tesis; primero serán expuestas aquéllas derivadas de una aproximación analítica a la noción de convención, después aquéllas que tienen que ver con una aproximación teórica y, por último, aquellas obtenidas a partir de una aproximación empírico-experimental.

La principal conclusión alcanzada a través de la aproximación analítica, desarrollada en la primera parte de la tesis, ha sido constatar que las convenciones sociales tienen una naturaleza dual, un carácter bifronte. Tal como han señalado otros autores para el caso de las instituciones (Kliemt, 1986; Crawford y Ostrom, 1995; Ostrom, 2005), las convenciones son, al mismo tiempo, tanto regularidades estables en el comportamiento social como principios normativos que guían dicho comportamiento. Para llegar a dicha conclusión, en el segundo y el tercer capítulo hemos revisado,

respectivamente, las teorías económica y sociológica de las convenciones sociales.

La reconstrucción de la noción de convención social como *convención de comportamiento* predominante en la teoría económica contemporánea nos ha permitido refinar nuestra descripción del componente de regularidad de comportamiento presente en toda convención social. Nuestro tipo ideal de *convención de comportamiento* queda definido por las siguientes tres características: primero, se trata de una regularidad empíricamente observable en el comportamiento humano, es decir, una regularidad que ha de ser constatada de forma recurrente en situaciones de interacción social; segundo, dicha regularidad es indeterminada y tiene un rango limitado de arbitrariedad, en tanto que los participantes en la misma se mostrarían indiferentes entre dicha convención y otra pauta de comportamiento social alternativa si ésta existiera; tercero, se trata de una regularidad estable, ya que una vez establecida los sujetos individuales no tienen incentivos para desviarse unilateralmente de ella. Desde el punto de vista de la teoría de juegos, estas tres características se ajustan a la definición de una *convención de comportamiento* como cualquier equilibrio estable en un juego que tiene dos o más equilibrios estables. Aquí la noción de equilibrio estable está tomada de la moderna teoría de juegos evolutiva desarrollada a partir de los trabajos de Maynard Smith (1982). Intuitivamente, una convención sería una pauta estable de comportamiento alcanzada en una sociedad después de un proceso de evolución social o cultural.

Como hemos visto, el componente de regularidad de comportamiento supone sólo una de las dos caras que componen nuestra noción de convención social. La otra cara de la convención es la definición de ésta como un principio normativo que guía la acción social. Para dar cuenta de dicho componente normativo hemos recurrido a la

reconstrucción del tipo ideal de *convención normativa* propio de la teoría sociológica.

Mientras que la mayoría de teóricos económicos pertenecientes a la corriente principal de la economía (*mainstream economics*) no tendrían mayores reparos en aceptar el tipo ideal de *convención de comportamiento* presentado, no ocurre así con la noción sociológica de *convención normativa*. En este sentido, puede considerarse mayor la originalidad de nuestro segundo tipo ideal de convención. Para elaborarlo, hemos reconstruido la línea que va desde de la definición de convención social como un tipo específico de orden social realizada por Weber (1964) hasta la propuesta de convención como principio normativo de la filósofa analítica Margaret Gilbert (1989). Nuestra noción sociológica de convención como principio normativo queda así definida por las cuatro características siguientes: primero, se trata de un principio de acción aceptado de forma tácita y conjunta por una colectividad; segundo, la fuente de obligatoriedad del mismo radica en la existencia de mecanismos de aprobación de los comportamientos contrarios a éste; tercero, la motivación que tienen los individuos para aprobar o desaprobar una acción de acuerdo con la *convención normativa* existente está relacionada con la equiparación de la convención con un acuerdo tácito y, por tanto, la desviación de la convención desencadena una actitud similar a la resultante de la trasgresión de un acuerdo; cuarto, mientras que las normas morales definen las conductas como "buenas" o "malas", las convenciones normativas definen las acciones como "apropiadas" e "inapropiadas".

Tras exponer ambos tipos ideales de convención, la primera parte de la tesis concluye aportando una definición provisional de convención social que no es sino una integración de los componentes conductual y normativo a los que hemos hecho

referencia con anterioridad en un único marco analítico. Dicho esto, ¿hasta qué punto podemos afirmar que nuestro intento de reconstrucción analítica de la noción de convención social ha tenido éxito? En primer lugar, nuestra definición nos permite distinguir la noción de convención social de otro tipo de regularidades sociales, como la "costumbre", la "moda" o la "situación de intereses", así como de otro tipo de instituciones, como las "normas morales" o las "leyes". En segundo lugar, nuestra definición recoge, siguiendo a autores como Sugden (1986; 1989), un modo sencillo de formalizar matemáticamente la noción de convención social como un equilibrio estable en un juego que contiene una multiplicidad de equilibrios, tal como la definíamos anteriormente. En tercer lugar, la definición aporta un mecanismo explicativo acerca del porqué de la conformidad con la convención, equiparando ésta con un acuerdo tácito. Por último, nuestra definición dual nos ha permitido elaborar una tipología de convenciones sociales – *convenciones establecidas*, *convenciones emergentes*, *convenciones difusas* y *convenciones decadentes* – que, aunque aquí sólo haya sido esbozada esquemáticamente, puede ser de utilidad para la realización de investigaciones empíricas posteriores en el ámbito de las convenciones sociales. Por tanto, consideramos que el trabajo realizado en la primera parte de la tesis contribuye a asentar los fundamentos teóricos y metodológicos necesarios para dar cuenta de un modo riguroso de la noción de convención social.

A pesar de lo expuesto anteriormente, no cabe duda de que aún quedan algunos aspectos relevantes, en relación con la noción de convención social, que necesitan de una mayor atención teórica en el futuro. Entre éstos no es menor la distinción, bastante ambigua en la teoría social contemporánea, entre las nociones de convención social y norma social. Dicho problema ha sido resuelto en esta tesis adoptando la postura

mantenida por Weber (1964), en la que el concepto de convención social es bastante amplio y abarca la mayoría de instituciones sociales que han recibido la denominación de norma social en la literatura sociológica posterior. En este sentido no hemos entrado en las discusiones de la diferencia entre ambos conceptos planteadas por autores como Ullmann-Margalit (1977) – la cual distingue entre “normas de dilema del prisionero”, “normas de coordinación” (a las que pertenecen las convenciones) y “normas de parcialidad” –, Elster (1989) – el cual establece una tipología de normas dentro de las cuales aparece el concepto de “equilibrio convencional” – o Coleman (1990) – quien distingue entre “normas esenciales” y “normas convencionales”. Nuestra elección terminológica weberiana hace que hayamos pospuesto para otro momento una mayor dedicación a dicho debate. Otro de los temas que la primera parte de la tesis ha dejado abierto es el de las motivaciones que se hallan detrás de la conformidad con la convención. Sin embargo, este último sí que ha sido abordado en la segunda y la tercera parte de la tesis. De ello nos ocuparemos a continuación.

El objetivo de la segunda parte de la tesis era formalizar los argumentos presentados en la primera. En concreto, pretendíamos dotar de una estructura formal, empleando la teoría de juegos, a la definición de convención social con la que concluíamos la primera parte. Partiendo de nuestra definición dual, dicha estructura formal se compone de dos elementos: primero, una clase de juegos básica, que permite derivar posibles *convenciones de comportamiento* a partir de los múltiples problemas de coordinación contenidos en dicha clase de juegos y, segundo, una serie de principios o *convenciones normativas* que prescriben comportamientos concretos en los juegos contenidos en dicha estructura básica. ¿Qué aporta a nuestro trabajo la formalización de ambos componentes?

Una clase de juegos es una estructura genérica que establece los límites de múltiples estructuras de interacción posibles. Así, cada representación matricial de un juego concreto que se encuentre dentro de dichos límites se ajustaría a las condiciones de nuestra clase de juegos, en cuanto a la ordenación de las preferencias de los jugadores y, por tanto, al número y las características de los equilibrios. Esta elección metodológica nos permite captar la intuición de que una convención concreta puede emerger no sólo en una estructura de interacción concreta y perfectamente definida, sino también en estructuras de interacción similares y cuya definición numérica concreta no esté del todo clara. Por tanto, nuestra clase de juegos nos permite formalizar el componente de regularidad estable de comportamiento de la convención pero, ¿cómo formalizamos el componente normativo?

Para formalizar el componente normativo de nuestra definición de convención social hemos definido una serie de principios normativos que prescriben cómo comportarse en nuestra clase de juegos. Cada principio queda definido por una sentencia normativa que nos indica cómo actuar en cada situación concreta de interacción en el seno de la clase de juegos definida. Así, un mismo comportamiento observado puede relacionarse tanto con la estructura de interacción definida (componente conductual), como con el posible principio que ha conducido a dicho comportamiento (componente normativo). La elección de principios normativos que, en apariencia, puede haber resultado arbitraria responde a la descripción de los comportamientos más comunes para la clase de juegos estudiada por parte de numerosos trabajos experimentales.

La segunda parte de la tesis no se ha limitado a una formalización de nuestra definición de convención social, sino que, a partir de la misma, hemos elaborado un marco de análisis apoyado en las definiciones proporcionadas por la teoría de juegos evolutiva. Si bien las herramientas de análisis evolutivo empleadas han sido bastante estándares – recurriendo al concepto de estrategia evolutivamente estable –, en esta tesis hemos introducido una variante en los análisis evolutivos al uso, la cual posiblemente constituye uno de los aspectos más originales del presente trabajo. Siguiendo a autores como Macy (1997), nuestro objetivo no ha sido el estudio de la estabilidad evolutiva de los comportamientos concretos – como tradicionalmente ha venido realizando la teoría de juegos evolutiva – sino el análisis de la estabilidad evolutiva de los principios que guían dichos comportamientos. En otras palabras, no nos hemos ocupado de analizar si el comportamiento X es evolutivamente estable en una población P_0 , sino si el principio Pr que prescribe el comportamiento X , el comportamiento Y o una combinación de ambos comportamientos según se den las circunstancias, es evolutivamente estable en dicha población.

¿Qué resultados hemos obtenido a partir de nuestro análisis evolutivo? En primer lugar, hemos constatado que, dada la clase de juegos básica empleada, existe una multiplicidad de principios o convenciones normativas evolutivamente estables. En otras palabras, dependiendo de las concreciones numéricas que puede adoptar nuestra clase de juegos, más de un principio de entre los considerados podría ser estable en una población determinada. Esto nos ha permitido añadir, a la definición propuesta en la primera parte, el rango limitado de arbitrariedad del que también goza la noción de *convención o principio normativo*. En segundo lugar, hemos demostrado que, de nuevo para nuestra clase de juegos, existe una *multiplicidad genética* de convenciones

normativas, es decir, que, dadas las circunstancias, podrían coexistir varios principios normativos en una población. Este segundo resultado es interesante en tanto que abunda en la distinción entre los dos componentes de nuestra definición de convención, en el sentido de que una regularidad de comportamiento – un determinado equilibrio en los términos de la teoría de juegos – podría estar apoyada en más de un principio normativo.

Si bien, los resultados obtenidos son alentadores, en tanto que coinciden con algunas de nuestras intuiciones básicas acerca de cómo operan las convenciones sociales, tenemos que ser muy conscientes de los límites que plantea nuestro análisis. Primero, como hemos señalado en varias ocasiones, los resultados obtenidos sólo son válidos para la clase de juegos y los principios definidos; cualquier generalización directa a otros escenarios sería inapropiada. Segundo, aunque el concepto de estrategia (en nuestro caso principio) evolutivamente estable tiene un regusto a análisis dinámico, en esencia se trata de un concepto estático que tiene que ver con las condiciones que debería cumplir una estrategia o principio para mostrarse estable ante las presiones evolutivas. Sería interesante el análisis de las dinámicas evolutivas – como la conocida *dinámica del replicador* (ver Weibull, 1995) – que podrían darse a partir del escenario descrito. Ambos ejercicios – el estudio de otras clases de juegos y otros principios, así como el estudio dinámico de nuestro marco de análisis – supondrían extensiones naturales que pueden ser realizadas a la segunda parte de esta tesis.

Si la segunda parte nos ha servido fundamentalmente para demostrar que los argumentos desarrollados en la primera pueden ser formalizados en el marco de la teoría de juegos, el objetivo de la tercera era estudiar hasta qué punto dicha estructura formal

seguía siendo válida en una situación experimental de laboratorio en la que tomaban parte sujetos reales. Para ello, nuestro marco teórico fue puesto a prueba en dos experimentos llevados a cabo en el laboratorio del *Max Planck Institute of Economics*, en Jena (Alemania). Sin embargo, antes de exponer la metodología y los resultados de cada uno de estos experimentos, y ante la ausencia de reflexiones metodológicas en la sociología contemporánea respecto a la validez de los experimentos económicos, el penúltimo capítulo de esta tesis supone un paréntesis metodológico dedicado a la evaluación del método experimental.

La reflexión metodológica, acerca de la validez de los experimentos económicos en sociología, planteada en el capítulo 6, nos ha permitido alcanzar dos conclusiones que han servido como justificación metodológica de los resultados obtenidos en nuestro análisis experimental. La primera de ellas es la distinción entre lo que hemos denominado *experimentos de orientación empírica* y *experimentos de orientación teórica*. Mientras que los primeros tratan de analizar una situación empírica concreta, empleando muestras representativas y apenas manipulando la realidad, los segundos están explícitamente diseñados para manipular y, de este modo, poner a prueba los postulados de una teoría, hipótesis teórica o mecanismo explicativo preexistente. Dado que el tipo de diseño empleado en esta tesis puede ser encuadrado en esta segunda categoría, el resto del capítulo se ocupó de analizar los problemas de validez de este segundo tipo de experimentos. En este sentido, la segunda de las conclusiones alcanzadas en este capítulo metodológico tiene que ver con una discusión metodológica planteada por Kanazawa (1999a) y hace referencia al papel secundario desempeñado por el problema de la validez externa en los *experimentos de orientación teórica*. Así, este tipo de experimentos, al estudiar mediante la comparación de grupos de contraste la

validez de una determinada teoría, no requiere de un diseño que tenga en cuenta las características concretas de una determinada población. A pesar de ello, el capítulo plantea algunas de las limitaciones de estos experimentos, especialmente relacionadas con la *artificialidad* propia de los experimentos de laboratorio.

Una vez discutida la validez de los experimentos económicos en sociología, la tesis concluye con la presentación de los resultados de nuestra aproximación experimental a la noción de convención social. Los resultados obtenidos en los dos experimentos realizados, con el objetivo de poner a prueba nuestro marco teórico, han servido para reforzar los argumentos presentados en las dos primeras partes de la tesis. En primer lugar, la definición de estrategia evolutivamente estable (EEE) ha resultado ser especialmente apropiada para dar cuenta de la noción de *convención normativa* – recordemos que más de nueve de cada diez participantes en el primer experimento seleccionaron principios que se ajustaban a dicha definición. Éste es un resultado que nos anima a continuar estudiando las condiciones bajo las cuales los distintos *principios* o *convenciones normativas* son evolutivamente estables, abundando sobre la idea de que es necesario abandonar la noción de estabilidad presente en la definición de equilibrio de la teoría de juegos estándar por una idea de equilibrio en la línea de la noción de equilibrio (combinación de estrategias) evolutivamente estable empleada en esta tesis.

El segundo resultado relevante de nuestros experimentos tenía que ver con la consistencia entre el comportamiento concreto de los individuos y la prescripción realizada por las *convenciones normativas*. Este resultado refuerza otra de las hipótesis básicas de nuestro trabajo: el componente normativo de las convenciones. Al analizar

los comportamientos individuales llegábamos a la conclusión de que, en su mayoría (alrededor de ocho de cada diez), dichos comportamientos estaban completamente determinados por las recomendaciones derivadas de los principios. Así, el carácter normativo de las convenciones quedaba reflejado en la existencia de un motivo para la acción – la prescripción realizada por un principio concreto – distinto al cálculo racional individual; aunque en multitud de ocasiones ambos pudieran coincidir. Nótese que la existencia de dicha motivación diferente al cálculo racional individual – similar a la esgrimida por Elster (1989b: 113) para el caso de las normas sociales – no nos aclara si los sujetos acomodaban su comportamiento a la prescripción de los principios por una cuestión de “ahorrar costes cognitivos” o porque valoraban intrínsecamente la bondad de un principio concreto. Pero, que no seamos capaces de apuntar la razón última de cada de una de las acciones observadas, no quiere decir que no seamos capaces de afirmar, como hemos hecho más arriba, el componente normativo de las convenciones.

Un tercer resultado obtenido en ambos experimentos nos da alguna pista sobre esta “razón última” para la acción que no podíamos determinar analizando sólo la consistencia de los comportamientos con respecto a los principios. No todos los principios o convenciones normativas fueron valorados del mismo modo. Así, más de la mitad de los sujetos participantes en ambos experimentos preferían interactuar en un “mundo” en el que el principio normativo fuera la eficiencia. Es decir, seleccionaban aquel principio que maximizaba el bienestar conjunto de ambos participantes como el más adecuado para alcanzar una coordinación duradera. Este resultado se veía reforzado por el hecho de que la consistencia mostrada por los sujetos respecto al criterio de eficiencia era significativamente mayor a la mostrada hacia otros principios. Por tanto, parece que el éxito en el establecimiento de una convención que resuelva un problema

recurrente de coordinación social también tiene que ver con las características internas de dicha convención, es decir, hasta qué punto consigue una coordinación de los comportamientos satisfactoria para todas las partes. En el contexto limitado de nuestra clase de juegos básica dichas características están más cerca de la propuesta de contrato social utilitarista de Harsanyi que de la propuesta rawlsiana. Dicho esto, podríamos preguntarnos si tal conclusión puede tener que ver con el contexto cultural en el que los experimentos fueron realizados – recordemos que ambos estudios fueron llevados a cabo en Alemania. Ésta es una cuestión que sólo podría ser resuelta si el diseño concreto empleado fuera replicado en otro contexto en el sospecháramos que podrían existir diferencias – por ejemplo, un país del sur de Europa como España. Dicha comparación se escapa de los objetivos de esta tesis y, de nuevo, se presenta como una interesante extensión de la misma.

El cuarto resultado relevante derivado de nuestros experimentos hace referencia a la pregunta: ¿cómo afecta a la coordinación de los sujetos la introducción de criterios normativos? Como hemos tenido la oportunidad de comprobar en la presentación de resultados de los experimentos en el capítulo 7, el empleo de criterios normativos incrementa sustancialmente los niveles de coordinación entre los sujetos experimentales. Este resultado refuerza la consideración del papel desempeñado por las *convenciones normativas* en la estabilización de las *convenciones de comportamiento*. De este modo, nos anima a continuar estudiando experimentalmente cómo interaccionan dichos principios normativos con otros elementos analizados por los economistas experimentales (ver Camerer, 2003: capítulo 6) con el objetivo de estudiar cómo coordinan satisfactoriamente los individuos sus acciones.

Las conclusiones obtenidas a partir de los experimentos realizados no hacen sino reforzar, como ya lo hacía antes nuestro análisis teórico-evolutivo, la idea central de esta tesis: sólo mediante el análisis combinado de los aspectos conductuales y normativos podremos ser capaces de dar cuenta de la complejidad de las convenciones sociales. Dicho objetivo no es sencillo y, como hemos puesto de manifiesto, aún quedan multitud de cuestiones teóricas y problemas empíricos sin resolver. Sin embargo, esperamos que este trabajo haya servido, al menos, para establecer las bases para una investigación futura fructífera en uno de los terrenos más apasionantes de la teoría social: el estudio de las convenciones sociales.

REFERENCIAS BIBLIOGRÁFICAS

- AGASSI, J. (1960), "Methodological Individualism", *British Journal of Sociology*, Vol. 11 (3), pp. 244-270.
- AGUIAR, F., H. CRIADO y F. HERREROS (2003), "Sociología y elección racional", en S. Giner (ed.), *Teoría sociológica moderna*, Barcelona, Ariel.
- AGUIAR, F. (2004), "Teoría de la decisión e incertidumbre: modelos normativos y descriptivos", *EMPIRIA*, nº 8, pp. 139-160.
- ALLAIS, M. (1953), "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine", *Econometrica*, Vol. 21, pp. 503-546.
- AUMANN, R. J. (1976), "Agreeing to Disagree", *Annals of Statistics*, Vol. 4, pp. 1236-1239.
- AXELROD, R. (1981), "The Emergence of Cooperation Among Egoists", *American Political Science Review*, Vol. 75, pp. 306-318.
- (1986), *La evolución de la cooperación*, Madrid, Alianza Editorial.
- BASKHAR, V. y F. VEGA-REDONDO (2004), "Migration and the Evolution of Convention", *Journal of Economic Behavior & Organization*, Vol. 55 (3), pp. 397-418.
- BATIFOULIER, P. (2001), *Théorie des conventions*, París, Economica.
- BATIFOULIER, P. y G. de LARQUIER (2001), « La convention en théorie des jeux », en P. Batifoulier (2001), *Théorie des conventions*, París, Economica.

- BERNINGHAUS, S., K. M. EHRHART y C. KESER (2002), “Conventions and Local Interaction Structures: Experimental Evidence”, *Games and Economic Behavior*, Vol. 39, pp. 177-205.
- BERNINGHAUS, S., W. GÜTH y H. KLIEMT (2005), “Conventions. Some Conventional and Some Not So Conventional Wisdom”, *Homo Oeconomicus*, Vol. 22 (2), pp. 147-168.
- BIBOW, J., P. LEWIS y J. RUNDE (2005), “Uncertainty, Conventional Behavior, and Economic Sociology”, *American Journal of Economics and Sociology*, Vol. 64 (2), pp. 507-532.
- BIGGART, N. W. y T. D. BEAMISH (2003), “The Economic Sociology of Conventions: Habit, Custom, Practice, and Routine in Market Order”, *Annual Review of Sociology*, Vol. 29, pp. 443-464.
- BINMORE, K. (1994), *Teoría de juegos*, Madrid, McGraw Hill.
- (1998), *Just Playing. Game Theory and the Social Contract*, The MIT Press.
- (2005), *Natural Justice*, Oxford, Oxford University Press.
- BOYER, R. y A. OLÈAN (1992), “How do Conventions Evolve?”, *Journal of Evolutionary Economics*, Vol. 2, pp. 165-177-.
- BROWN, S. R. y L. E. MELAMED (1990), *Experimental Design and Analysis*, Thousand Oaks, Sage.
- CACHON, G. P. y C. CAMERER (1996), “Loss-avoidance and Forward Induction in Experimental Coordination Games”, *Quarterly Journal of Economics*, Vol. 111, pp. 165-194.

CAMERER, C. F. (1998), "Behavioral Economics: Reunifying Psychology and Economics", Proceedings of the National Academy of Sciences of the United States of America, Vol. 96 (19), pp. 10575-10577.

(2003a), "Coordination", C. F. Camerer, *Behavioral Game Theory*, Princeton, Princeton University Press.

(2003b), *Behavioral Game Theory*, Princeton, Princeton University Press.

CAMPBELL, D. T. y J. C. STANLEY (1963), *Experimental and Quasi-experimental Designs for Research*, Boston, Houghton Mifflin Company.

CARNAP, R. (1962), *Logical Foundations of Probability*, Routledge & Keagan Paul: The University of Chicago Press.

CHAPIN, F. S. (1931), *Experimental Designs in Sociological Research*, Nueva York, Harper y Row.

CHWE, M. (2001), *Rational Rituals: Culture, Coordination, and Common Knowledge*, Princeton, Princeton University Press.

COLEMAN, J. S. (1987), "Norms as Social Capital", en G. Radnitzky y P. Bernholz (eds.), *Economic Imperialism*, Nueva York, Paragon.

(1990a), *Foundations of Social Theory*, Cambridge (MA), Harvard University Press.

(1990b), "On the Emergence of Norms", en M. Hechter, K. D. Opp y R. Wippler (eds.), *Social Institutions: Their Emergence, Maintenance and Effects*, Berlin, Walter de Gruyter.

COOPER, R., D. V. DEJONG, R. FORSYTHE y T. W. ROSS (1989), "Communication in the Battle of the Sexes Game: Some Experimental Results", *RAND Journal of Economics*, Vol. 20 (4), PP. 568-587.

- (1993), "Forward induction in the Battle of the Sexes Game", *American Economic Review*, Vol. 83 (5), pp. 1303-1316.
- CRAWFORD y OSTROM (1995), "A Grammar of Institutions", *American Political Science Review*, Vol. 89 (3), pp. 582-600.
- CROSON, R. (2002), "Why and How to Experiment: Methodologies from Experimental Economics", *University of Illinois Law Review*, Vol. 2002, pp. 921-945.
- CUBITT, R. P. y. R. SUGDEN (2003), "Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory", *Economics and Philosophy*, Vol. 19, pp. 175-210.
- CUMMING, S. (2003), "Language and Power", mimeo.
<http://www.linguistics.ucsb.edu/faculty/cumming/ling50/emcdata.htm>
- DEQUECH, D. (2005), "Cognition and Valuation: Some Similarities and Contrasts between Institutional Economics and the Economics of Convention", *Journal of Economic Issues*, Vol. 39 (2), pp. 465-473.
- DUPUY, J. P., F. EYMARD-DUVERNAY, O. FAVEREAU, A. ORLÉAN, R. SALAIS y L. THÉVENOT (1989), « Introduction », *Revue Economique*, Vol. 2, pp. 141-145.
- ELSTER, J. (1989a), *The Cement of Society*, Cambridge, Cambridge University Press.
- (1989b), *Nuts and Bolts for the Social Sciences* Cambridge, Cambridge University Press.
- ENGEL, C. (2005), *Generating Predictability. Institutional Analysis and Design*, Cambridge, Cambridge University Press.

- FAVEREAU, E. (1986), "La formalisation du rôle des conventions dans l'allocation des ressources", R. Salais y. L. Thévenot (eds.), *Le Travail Marchés, règles, conventions*, Paris, Economica.
- FAVEREAU, O. y E. LAZEGA (2002), *Conventions and Structures in Economic Organization*, Cheltenham, Edward Elgar.
- FISCHBACHER, U. (1999), "z-Tree-Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual", *Workin Paper no 21*, Institute for Empirical Research in Economics, University of Zurich.
- FRANCISCO, A. de (1997), *Sociología y cambio social*, Barcelona, Ariel.
- GIDDENS, A. (1979), *Central Problems in Social Theory*, Berkeley, University of California Press.
- (1984), *The Constitution of Society: Outline of the Theory of Structuration*, Berkeley, University of California Press.
- GILBERT, M. (1983), "Notes on the Concept of Social Convention", *New Literary History*, Vol. 14 (2), pp. 225-251.
- (1989), *On Social Facts*, Londres, Routledge.
- (2001), "Collective Preferences, Obligations, and Rational Choice", *Economics and Philosophy*, 17, pp. 109-119.
- GINER, S. (1997), "Intenciones humanas, estructuras sociales: para una lógica de la situación", M. Cruz (ed.), *Acción humana*, Barcelona, Ariel.
- GINTIS, H. (2006), "The Unification of the Behavioral Sciences", *Behavioral and Brain Sciences*, en prensa.

- GREENWOOD, E. (1976 [1945]), *Experimental Sociology. A Study in Method*, Nueva York, Octagon Books.
- GREINER, B. (2004), “The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics”, *Working Paper Series in Economics*, nº 10, University of Cologne.
- GUALA, F. (2005), *The Methodology of Experimental Economics*, Cambridge, Cambridge University Press.
- GÜTH, W., R. SCHMITTBERGER y B. SCHWARZE (1982), “An Experimental Analysis of Ultimatum bargaining”, *Journal of Economic Behavior & Organization*, Vol. 3 (4), pp. 367-388.
- GÜTH, W. y H., KLIEMT (1998), “The Indirect Evolutionary Approach. Bridging the Gap Between Rationality and Adaptation”, *Rationality and Society*, Vol. 10, pp. 377-399.
- GÜTH, W. y H. BESTER (1998), “Is Altruism Evolutionarily Stable?”, *Journal of Economic Behavior & Organization*, Vol. 34 (2), pp. 193-209.
- GÜTH, W. y R. TIETZ (1990), “Ultimatum Bargaining Behavior - A survey and Comparison of Experimental Results”, *Journal of Economic Psychology*, Vol. 11 (3), pp. 417-449.
- GÜTH, W. y S. NAPEL (2002), “Inequality Aversion in a Variety of Games - An Indirect Evolutionary Analysis”, *Discussion Papers on Strategic Interaction*, nº 23-02, Max Planck Institute of Economics, Jena, Germany.
- HAMMERSTEIN, P. y R. SELTEN (1994), “Game Theory and Evolutionary Biology”, R. Aumann (ed.), *Handbook of Game Theory with Economic Applications*, Amsterdam, Elsevier.

HARDIN, R. (1971), "Collective Action as an Agreeable n-Prisoners' Dilemma", *Behavioral Science*, Vol. 16-5, pp. 472-481.

(1982), *Collective Action*, Londres, John Hopkins University Press.

HARSANYI, J. y R. SELTEN (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, The MIT Press.

HARUVY, E. y D. O. STAHL (2004), "Deductive versus Inductive Equilibrium Selection: Experimental Results", *Journal of Economic Behavior & Organization*, Vol. 53, pp. 319-331.

HAYASHI, N. y A. YOSANO (2005), "Trust and belief about others: Focusing on judgment accuracy of others' trustworthiness", *Sociological Theory and Methods*, Vol. 20 (1), pp. 59-80.

HAYEK, F. A. von (1945), "The Use of Knowledge in Society", *American Economic Review*, Vol. 35, pp. 519-530.

(1967), "Notes on the Evolution of Rules of Conduct", en F. A. von Hayek, *Studies in Philosophy, Politics, and Economics*, Chicago, University of Chicago Press.

HECHTER, M. K. D. OPP y R. WIPPLER (1990), *Social Institutions. Their Emergence, Maintenance and Effects*, Nueva York, Aldine de Gruyter.

HECHTER, M. y K. D. OPP (2001), *Social Norms*, Nueva York, Russell Sage Foundation.

HEDSTRÖM, P. y R. SWEDBERG (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge, Cambridge University Press.

HENRICH, J., R. BOYD, S. BOWLES, C. CAMERER, H. GINTIS, Richard MCELREATH y E. FEHR (2001), "In search of Homo Economicus: Experiments in 15 Small-Scale Societies", *American Economic Review*, Vol. 91(2), pp. 73-79.

HOBBS, T. (1993 [1651]), *Leviatán: la materia, forma y poder de un Estado eclesiástico y civil*, Madrid, Alianza.

HUME, D. (1975 [1777]), *An Enquiry concerning the Principles of Morals*, Oxford, Oxford University Press.

(1978 [1740]), *A Treatise on Human Nature*, Oxford, Oxford University Press.

JHERING (1916), *Der Zweck im Recht* (vol. II), Leipzig, Breitkopf & Härtel.

JUNTA DE ANDALUCÍA (2006), *Emergencias 112*, documento electrónico: <http://www.juntadeandalucia.es/gobernacion/opencms/portal/SubdireccionEmergencias/Emergencias112/ContenidosEspecificos/index?entrada=tematica&tematica=68&subtematica=70>.

KAGEL, J. H. y. A. E. ROTH, *The Handbook of Experimental Economics*, Princeton, Princeton University Press.

KAHNEMAN, D., J. NETSCH y R. THALER (1986), “Fairness and the Assumptions of Economics”, *Journal of Business*, Vol. 59, pp. 285-300.

KAHNEMAN, D. y. A. TVERSKY (1979), “Prospect Theory: An Analysis of Decision under Risk”, *Econometrica*, Vol. 47, pp. 263-291.

KANAZAWA, S. (1999a), “Using Laboratory Experiments to Test Theories of Corporate Behavior”, *Rationality and Society*, vol. 11 (4), pp. 443-61.

(1999b), “Testing Macro Organizational Theories in Laboratory Experiments”, *Social Sciences Research*, Vol. 28, pp. 66-87.

KEYNES, J. M. (1973), “The General Theory of Employment, Interest and Money”, en J. Keynes, *The Collected Writings of John Maynard Keynes* (Vol. VI), Londres, Macmillan.

- KLIEMT, H. (1986), *Las instituciones morales*, Barcelona, Editorial Alfa.
- KNIGHT, J. (1995), *Explaining Social Institutions*, Ann Arbor, University of Michigan Press.
- KOLLOCK, P. (1994), "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment and Trust", *American Journal of Sociology*, Vol. 100 (2), pp. 313-345.
- (1998), "Social Dilemmas: The Anatomy of Cooperation", *Annual Review of Sociology*, vol. 24, pp. 183-214.
- KOMORITA, S. S. (1994), "Introduction", S. S. Komorita (ed.), *Social Dilemmas*, Madison, Brown and Benchmark.
- KRISTOF, N. D. y. S. WUDUNN (1994), *China Wakes. The Struggle for the Soul of a Rising Power*, Nueva York, Times Books.
- LATSIS, J. S. (2004), *Theories and the Arbitrary Economy*, Universidad de Cambridge.
- (2005), "Is there redemption for conventions?", *Cambridge Journal of Economics*, Vol. 29, pp. 709-727.
- LAZEGA, E. y. O. FAVEREAU (2002), "Introduction", O. Favereau y. E. Lazega (ed.), *Conventions and Structures in Economic Organization*, Cheltenham, Edward Elgar.
- LEDYARD, J. (1995), "Public goods: A survey of experimental research", J. Kagel y A. Roth (eds.), *Handbook of Experimental Economics*, Princeton, Princeton University Press.
- LEVIN, I. P. (1999), *Relating Statistics and Experimental Design. An Introduction*, Thousand Oaks, Sage.

- LEWIS, D. (1969), *Conventions. A Philosophical Study*, Oxford, Basil Blackwell.
- LINDE, E. y. J. P. (2003), *La coordinación de las policías*, Madrid, Colex.
- LUCE, R. D. y. H. RAIFFA (1957), *Games and Decisions*, Nueva York, Wiley.
- MACY, M. (1997), "Identity, Interest and Emergent Rationality", *Rationality and Society*, Vol. 9 (4), pp. 427-448.
- MARKOVSKY, B y S. R. THYE (2001), "Social influence on paranormal beliefs », *Sociological Perspectives*, Vol. 44 (1), pp. 21-44.
- MARTÍN, P. (2003), "La cooperación entre las fuerzas y cuerpos de seguridad: perspectivas de futuro", E. Linde y J. Pérez (eds.), *La coordinación de las policías*, Madrid, Colex.
- MAYNARD SMITH, J. y. G. R. PRICE (1973), "The Logic of Animal Conflict", *Nature*, Vol. 246, pp. 15-18.
- MAYNARD SMITH, J. (1974), "The theory of Games and the Evolution of Animal Conflicts", *Journal of Theoretical Biology*, Vol. 47, pp. 209-221.
- (1982), *Evolution and the Theory of Games*, Cambridge, Cambridge University Press.
- McADAMS, R. H. y. J. NADLER (2005), "Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/Dove Game", *Journal of Empirical Legal Studies*, Vol. 2 (1), pp. 87-124.
- MENGER, K. (1963), *Problems in Economic and Sociology*, Urbana, University of Illinois Press.
- MILGROM, P. (1981), "An Axiomatic Characterization of Common Knowledge", *Econometrica*, nº 49 (1), pp. 219-222.

- MILL, J. S. (1891), *System of Logic Ratiocinative and Inductive*, Londres, Longmans, Green and Co.
- MILLER, L. M. (2004), “Escenarios para la acción colectiva”, *Revista Internacional de Sociología*, nº 39, pp. 167-197.
- (2007), “Coordinación y acción colectiva”, *Revista Internacional de Sociología*, nº 46.
- MILLER, L. M. y C. VANBERG (2006), “The Influence of a Social Norm on Behavior in a Dictator Game”, mimeo.
- MILLER, S. (1992), “On conventions”, *Australian Journal of Philosophy*, Vol. 70 (4), pp. 435-444.
- MORROW, J. D. (1994), *Game Theory for Political Scientists*, Princeton, Princeton University Press.
- MOSER, C. A. y. G. KALTON (1992 [1971]), “Experiments and Investigations”, en C. A. Moser y. G. Kalton, *Survey Methods in Social Investigation*, Hants, Gower.
- MOSTERÍN, J. (2006), *La Hélade. Historia del pensamiento*, Madrid, Alianza.
- NABOKOV, V. (2002 [1955]), *Lolita*, Barcelona, Anagrama.
- NOGUERA, J. A. (2006), “Why We Need an Analytical Sociological Theory?”, *Papers*, nº 80.
- NORTH, D. (1986), “The New Institutional Economics”, *Journal of Institutional and Theoretical Economics*, Vol. 142, pp. 230-237.
- (1990), *Institutions, Institutional Change, and Economic Performance*, Nueva York, Cambridge University Press.

OAKLEY, A. (1998), "Experimentation and social interventions: a forgotten but important history", *BMJ*, Vol. 317, pp. 1239-1242.

O'NEILL, B. (2000), *Honor, Symbols and War*, Ann Arbor, University of Michigan Press.

OLSON, M. (1971), *The Logic of Collective Action. Public Goods and the Theory of the Groups*, Cambridge (MA), Harvard University Press.

ORLÉAN, A. (1999), *Le pouvoir de la finance*, Paris, Odile Jacob.

(2004), *Analyse économique des conventions*, Paris, Quadrige/Presses Universitaires de France.

OSTROM, E. (1986), "An Agenda for the Study of Institutions", *Public Choice*, Vol. 48, pp. 3-25.

(1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Nueva York, Cambridge University Press.

(2000), "Collective Action and the Evolution of Social Norms", *Journal of Economic Perspectives*, Vol. 14 (3), pp. 137-158.

(2005), *Understanding Institutional Diversity*, Princeton, Princeton University Press.

OSTROM, V. (1980), "Artisanship and Artifact", *Public Administration Review*, Vol. 40, pp. 309-317.

(1986), "A fallabilist's Approach to Norms and Criteria of Choice", en F. K. Kaufmann, G. Majone y V. Ostrom (eds.), *Guidance, Control, and Evaluation in the Public Sector*, Berlin, Gruyter.

- (1991), *The Meaning of American Federalism: Constituting a Self-Governing Society*, San Francisco, Institute for Contemporary Studies Press.
- RAWLS, J. (1971), *A Theory of Justice*, Cambridge (MA), Harvard University Press.
- REY, P. (2006), "Economía experimental y teoría de juegos", mimeo.
- RIKER, W. (1980), "Implications from the Disequilibrium of Majority Rule for the Study of Institutions", *American Political Science Review*, Vol. 74, pp. 432-446.
- ROBLES, J. M. (2004), "Racionalidad acotada: heurísticos y acción individual", *Theoria*, Vol. 14 (1), pp. 37-46.
- RYDVAL, O. y. A. ORTMANN (2005), "Loss Avoidance as Selection Principle: Evidence from Simple Stag-hunt Games", *Economics Letters*, Vol. 88, pp. 101-107.
- SACCONI, L. y. S. MORETTI (2002), "Fuzzy norms, default reasoning and equilibrium selection in games under unforeseen contingencies and incomplete knowledge", *LIUC Papers*, pubblicazione periodica dell'Università "Carlo Cattaneo".
- SCHELLING, T. (1960), *The Strategy of Conflict*, Cambridge (MA), Harvard University Press.
- SCHOTTER, A. (1981), *The Economic Theory of Social Institutions*, Cambridge, Cambridge University Press.
- SCHOTTER, A. y. B. SOPHER (2003), "Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study", *Journal of Political Economy*, Vol. 111 (3), pp. 498-529.
- SHELLY, R. K. y P. T. MUNROE (1999), "Do women engage in less task behaviour than men?", *Sociological Perspectives*, Vol. 42 (1), pp. 49-67.

- SMATKA, J., J. SKVORETS, T. SOZANSKI Y J. MAZUR (1998), "Conflict in Networks", *Sociological Perspectives*, Vol. 41 (1), pp. 49-66.
- SKYRMS, B. (1996), *Evolution and the Social Contract*, Cambridge, Cambridge University Press.
- STAMMLER, R. (1914), *Wirtschaft und Recht: nach der materialistischen Geschichtsauffassung ; eine sozialphilosophische Untersuchung*, Leipzig, Veit.
- STRAUB, P. G. (1995), "Risk Dominance and Coordination Failures in Static Games", *The Quarterly Review of Economics and Finance*, Vol. 35 (4), pp. 339-363.
- SUGDEN, R. (1986), *The Economics of Welfare, Rights and Co-operation*, Oxford, Basil Blackwell.
- (1989), "Spontaneous Order", *Journal of Economic Perspectives*, Vol. 3 (4), pp. 85-97.
- (1995), "The Coexistence of Conventions", *Journal of Economic Behavior and Organization*, Vol. 28 (2), pp. 241-256.
- (1998), "Conventions", P. Newman, *The New Palgrave Dictionary of Economics and the Law*, Albany, Macmillan.
- (2005), "Experiment, Theory, World: A Symposium on the Role of Experiments in Economics", *Journal of Economic Methodology*, Vol. 12 (2), pp. 177-184.
- THERBÖRN, G. (2002), "Back to Norms! On the Scope and Dynamics of Norms and Normative Action", *Current Sociology*, Vol. 50 (6), pp. 863-880.
- TÖNNIES (1909), *Die Sitte*, Frankfurt am Main, Rütten & Loening.
- ULLMANN-MARGALIT, E. (1977), *The Emergence of Norms*, Oxford, Oxford University Press.

- VAN DAMME, E. (1987), *Stability and Perfection of Nash Equilibria*, Londres, Springer.
- VAN HUYCK, J., R. C. BATTALIO y R. O. BEIL (1990), “Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure”, *American Economic Review*, Vol. 80, pp. 234-248.
- (1991), “Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games”, *Quarterly Journal of Economics*, Vol. 106, pp. 885-911.
- VAN HUYCK, J., R. C. BATTALIO y F. W. RANKIN (1997), “On the Origin of Conventions: Evidence from Coordination Games”, *The Economic Journal*, Vol. 107, pp. 576-596.
- VAN HUYCK, J., J. P. COOK y R. C. BATTALIO (1997), “Adaptative Behavior and Coordination Failure”, *Journal of Economic Behavior & Organization*, Vol. 32, pp. 483-503.
- VAN LANGE, P., W. LIEBRAND, D. MESSICK y H. WILKE (1992), “Introduction and Literature Review”, en D. Messick, H. Wilke y W. Liebrand, *Social Dilemmas. Theoretical Issues and Research Findings*, Oxford, Pergamon Press.
- VANDERSCHRAAF, P. (1998), “Knowledge, Equilibrium and Convention”, *Erkenntnis*, Vol. 49, pp. 337-369.
- VON NEUMANN, J. y O. MORGERNSTEIN (1947), *Theory of Games and Economic Behavior*, Princeton, Princeton University Press.
- WEBER, M. (1964 [1922]), *Economía y sociedad*, Madrid, Fondo de Cultura Económica.
- WEIBULL, J. W. (1995), *Evolutionary Game Theory*, Cambridge (MA), The MIT Press.

- (1998), "What Have We Learned from Evolutionary Game Theory So Far?", *IUI working Papers Series*, N° 487.
- WEIGELIN, E. (1919), *Sitte, Recht und Moral*, Leipzig, Walther Rothschild.
- WILLER, R., L. TROYER y M. J. LOVAGLIA (2005), "Influence over observers of structural power: An experimental investigation", *Sociological Quarterly*, Vol. 46 (2), pp. 263-277.
- WILLIANSO, O. E. (1985), *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, Nueva York, Free Press.
- WILSON, R. K. y C. M. RHODES (1997), "Leadership and Credibility in N-Person Coordination Games", *The Journal of Conflict Resolution*, Vol. 41 (6), pp. 767-791.
- YAMAGISHI, T., S. KANAZAWA, R. MASHIMA y S. TERAJ (2005), "Separating Trust from Cooperation in a Dynamic Relationship. Prisoners' Dilemma with Variable Dependence", *Rationality and Society*, Vol. 17 (3), pp. 275-308.
- YOUNG, H. P. (1993), "The Evolution of Conventions", *Econometrica*, Vol. 61, pp. 57-84.
- (1996), "The Economics of Convention", *Journal of Economic Perspectives*, Vol. 10 (2), pp. 105-122.
- (1998a), *Individual Strategy and the Social Structure: An Evolutionary Theory of Institutions*, Princeton, Princeton University Press.
- (1998b), "Social Norms and Economic Welfare", *European Economic Review*, Vol. 42, pp. 821-830.