CARACTERIZACIÓN DE MECANISMOS DE APRENDIZAJE E INTERACCIÓN SOCIAL MEDIANTE MÉTODOS MATEMÁTICOS DE LA FÍSICA ESTADÍSTICA

PEDRO JOSÉ MÉNDEZ TOVAR

TRABAJO DE GRADO PARA OPTAR AL TÍTULO DE MATEMÁTICO

ASESOR:
JAVIER MONTOYA MARTINEZ

PROGRAMA DE MATEMÁTICAS
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD DE CARTAGENA

CARTAGENA DE INDIAS
MARZO DE 2017

Agradecimientos

Quiero empezar dando mis agradecimientos primeramente a Dios, quien creo fervientemente me ha permitido alcanzar cada logro en mi vida, y en quien confío me dará la sabiduría necesaria para seguir adelante.

Sabiendo que no existe forma de compensar una vida de arduo sacrificio y dedicación, agradezco a mis padres PEDRO MÉNDEZ FLOREZ y AIFFA TOVAR GONZALES, quienes han sido grandes ejemplos de vida para mí, contribuyendo en gran manera a la conquista de esta meta y de mi formación profesional, ofreciéndome siempre su apoyo, estímulo y consejos incondicionalmente, poniendo toda su fe en mí y convenciéndome que siempre podré alcanzar logros mayores.

De igual manera agradezco a mi esposa LUISA INÉS ZÚÑIGA RAMOSs, porque siempre tuvo una palabra de ánimo cuando pensé abandonar mis objetivos, mostrándome así su sincero amor y afecto. A mis profesores por la comprensión y tolerancia que tuvieron al brindarme sus saberes a lo largo de mis estudios, siendo ejemplos para seguir siempre adelante. En especial a JOHN EDUARDO REALPE y JAVIER MONTOYA MARTINEZ, por ser mis maestros y tutores en los últimos años como estudiante de pregrado, posibilitando así la conclusión de mi formación profesional, de quienes he aprendido no sólo en lo que concierne al aspecto académico, sino a la formación ética y moral.

Al término de esta etapa de mi vida académica, quiero expresar un profundo agradecimiento a mi familia en general, a mis amigos, compañeros y a quienes con ayuda, apoyo y comprensión me alentaron a lograr esta realidad. Sólo deseo que entiendan que el logro mío es logro de ustedes, que mi esfuerzo es inspirado en ustedes, y que este peldaño alcanzado significa el comienzo de un arduo trabajo por continuar.

Índice general

Índice de figuras

Introducción

El aprendizaje es el proceso mediante el cual se adquiere una determinada habilidad, se asimila una información o se adopta una nueva estrategia de conocimiento y acción. Uno de los procesos cognitivos responsables de la propagación social es el aprendizaje social, entendido de manera general como aquel proceso por el cual la adquisición de información nueva por parte de los individuos se produce o está favorecida por su exposición recíproca en un entorno común. Al tener en cuenta la cantidad de agentes que interfieren y las interacciones entre ellos, es importante considerar el análisis de tal aprendizaje a la luz de la física estadística, los sistemas complejos y la teoría de juegos.

A lo largo del siglo XX, la física estadística, cuyo mayor éxito fue interpretar la termodinámica como el resultado estadístico macroscópico (emergente) de la interacción entre un enorme número de átomos o moléculas (agentes microscópicos), ha ampliado sus objetos de estudio a un variado tipo de "sistemas complejos", lo cual le ha valido la actual denominación de "Física Estadística de los Sistemas Complejos".

Los sistemas complejos, caracterizados más por su comportamiento rico y complicado que por su definición intrínseca, aparecen en diversas áreas: Física, Matemática, Biología, Química, Ingeniería, Economía, etc. Lo que caracteriza estos sistemas es la presencia de un número muy elevado de agentes que interactúan entre sí y como consecuencia surgen comportamientos emergentes, los cuales están caracterizados por ser independientes del comportamiento aislado de cada agente involucrado.

La teoría de juegos es la principal herramienta usada para modelar interacciones extratégicas en biología evolutiva y las ciencias sociales. Tradicionalmente, la teoría de juegos estudia los equilibrios de juegos simples. Definimos un juego complicado como aquel en el que hay muchos movimientos posibles, y además, muchos pagos condicionales posibles para esos moviemientos.

La comunidad científica ha comprendido que la investigación tradicional necesita adaptarse para abordar los nuevos problemas, y desarrollar métodos eficaces para comprender su naturaleza. Por otro lado, uno de los resultados de esta "ampliación de miras" es que el estudio de los sistemas complejos no afecta sólo a la ciencia e investigación básicas, sino a ámbitos mucho más aplicados de la innovación: desde el estudio y decodificación del genoma humano, al análisis y predicción de evolución de indicadores y magnitudes económicas (Bolsa, datos macroeconómicos) o industriales (consumos eléctricos o de agua), pasando por el diseño y fabricación de nuevos materiales (para la industria semiconductora, plásticos y polímeros, etc.) o el estudio de la meteorología y la dinámica oceánica global, por citar algunos ejemplos.

Entre estos intereses de investigación y aplicabilidad, surge la curiosidad sobre el estudio de las interacciones entre seres humanos, incluyendo la forma en la que se llevan a cabo algunos procesos como las dinámicas de aprendizaje en una determinada red de individuos, vista a la luz de la teoría de sistemas complejos y la teoría de juegos sobre redes.

Capítulo 1

Métodos y conceptos

1.1. Teoría de juegos

La teoría de juegos es el estudio matemático de la interacción entre agentes independientes y con intereses propios. Para tener un juego debemos tener al menos dos agentes o jugadores cuyas decisiones interaccionan de forma que pueden afectar los intereses de los otros jugadores, un conjunto de estrategias disponibles para cada jugador y una matriz de pagos que registre la recompensa para cada jugador en función de las decisiones tomadas.

Un ejemplo sencillo es el arquero atajando un penal. Los jugadores son el arquero y el que patea el penal. Las estrategias del arquero (para simplificar) diremos que son tirarse a la izquierda o a la derecha y las estrategias del que patea son tirar la pelota a la izquierda o a la derecha. La matriz de pagos para el arquero puede ser 1 si ataja y -1 si no ataja.

Un juego es llamado *trivial* cuando los pagos se pueden definir solo con las estrategias de un jugador y por tanto existe una estrategia ganadora. Todo juego no trivial debe tener algún aspecto de *conflicto de intereses*, aunque se puede pensar en juegos en los cuales el único problema es la coordinación, ya que los intereses de los jugadores coinciden. En estos juegos, el *conflicto* se presenta en la regla de coordinación entre los jugadores [28]. Por ejemplo, imagine un juego en el que tenemos a los jugadores Alicia y Bernardo, quienes quieren ir al cine, pero Alicia prefiere las películas de romance mientras Bernardo prefiere las de acción. El juego tiene una regla muy sencilla: si no se ponen de acuerdo no van al cine y la matriz de pagos sería la representada en la figura 1.1, donde el primer número de cada pareja ordenada es el pago de Bernardo y el segundo es el de Alicia. Si van a ver una pelíula de acción Bernardo gana más, pero Alicia no pierde todo porque al menos va al cine y viceversa. Es claro que si no se ponen de acuerdo pierden ambos. Este es un juego de coordinación donde las partes deben ceder para no perderlo todo [4].

Tanto la redacción como la lectura sobre la teoría de juegos ha crecido considerablemente en los últimos años, extendiéndose a disciplinas tan diversas como ciencias políticas, biología, psicología, economía, lingüistca, sociología y ciencias computacionales (entre otras). En este capítulo trataremos de mostrar

	B acción	**B romance**
A acción	(2,1)	(0,0)
A romance	(0,0)	(1,2)

Figura 1.1: Matriz de pagos para un juego de coordinación.

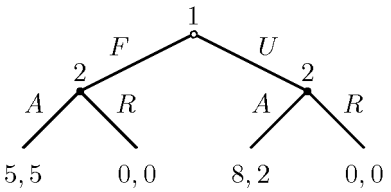

Figura 1.2: Un juego en forma extensiva

el tipo de introducción básica y breve que alguien con un interés urgente por la teoría de juegos estaría necesitando. Se abarcan los principales tipos de juegos, sus representaciones y los conceptos primordiales necesarios para analizarlos.

La teoría de juegos es estudiada principalmente por matemáticos y economistas, siendo la microeconomía su principal área de aplicación inicial. Pero ¿qué hace un científico computacional publicando textos sobre teoría de juegos? Como muchas otras disciplinas, la ciencia computacional ha sido influenciada en gran manera por la teoría de juegos. Así que no es extraño encontrar que una parte razonable del material de estudio computacional sobre *sistemas multiagentes* trate acerca de la teoría de juegos. Ese material se puede clasificar en dos tipos: Uno básico, y un material más avanzado y relevante para la inteligencia artificial (AI) y para la ciencia computacional. Se tratará aquí de abarcar el primer tipo de material.

Existen muchos textos sobre la teoría de juegos, algunos de ellos muy bien elaborados. Un buen estudio de la teoría de juegos debería considerarlos seriamente, por lo cual se mencionarán algunas referencias importantes.

Los juegos se han clasificado tradicionalmente en *juegos cooperativos* y *no cooperativos*. La diferencia radica en las posibilidades de comunicación, negociación y coordinación que se permite a los jugadores. Los juegos *no cooperativos* son aquellos en los que cada agente actúa siguiendo exclusivamente su propio interés y los jugadores no pueden hacer acuerdos vinculantes. A medida que ampliamos las posibilidades de cooperación, de comunicación y acuerdos, los juegos pasan a ser cooperativos. Un elemento central en los juegos cooperativos, o en el enfoque cooperativo en las situaciones de conflicto, es la existencia de un poder superior capaz de hacer cumplir los acuerdos entre las partes en conflicto. Actualmente existen muchos trabajos que intentan desarrollar una justificación no cooperativa de las soluciones cooperativas, integrando las dos ramas [28].

1.1.1. Juegos en Forma Extensiva

La presentación de juegos en forma extensiva modela juegos con algun orden que se debe considerar. Los juegos se presentan como árboles (como se muestra en la figura 1.2). Cada vértice o nodo representa un punto donde el jugador toma decisiones. El jugador se especifica por un número situado junto al vertice. Las líneas que parten del vértice representan acciones posibles para el jugador. Las recompensas se especifican en las hojas del árbol.

En el juego que se muestra en el ejemplo hay dos jugadores. El jugador 1 mueve primero y elige F o U. El jugador 2 ve el movimiento del jugador 1 y elige A o R. Si el jugador 1 elige U y entonces el jugador 2 elige A, el jugador 1 obtiene 8 y el jugador 2 obtiene 2.

1.1.2. Juegos en Forma Normal

La teoría de juegos estudia lo que pasa cuando agentes con intereses propios interactúan. Pero, ¿qué significa decir que los agentes tienen "intereses propios"? No necesariamente significa que quieran causar algún daño a los otros agentes con los que interactúan, ni que tengan cuidado solo de sí mismos. Más bien significa que cada agente tiene su propia descripción de cuáles "estados del mundo" (situaciones de juego) le agradan, y que él actúa intentando conseguir que dichos estados se den [23].

La aproximación dominante para modelar los intereses de un agente es la *teoría de la utilidad*. Esta aproximación teórica cuantifica la preferencia o el rechazo por parte de un agente hacia un conjunto de alternativas disponibles, y describe cómo esas preferencias cambain cuando un agente enfrenta inseguridad sobre la alternativa que recibirá. Específicamente, una *función utilidad* es un mapeo de estados del mundo a números reales. Esos números son interpretados como medidas del nivel de felicidad de un agente en un estado dado. Cuando el agente no tiene certeza del estado del mundo que enfrenta, su utilidad se define como el valor esperado de su función utilidad con respecto a la distribución de probabilidad apropiada sobre los estados [23].

Definición de juegos en forma normal

La forma normal, también conocida como la forma estratégica o matricial, es la representación más familiar de interacciones estratégicas en teoría de juegos. Un juego escrito de esta forma llega a representar la utilidad de cada jugador para cada estado del mundo, en especial cuando los estados del mundo dependen solo de las acciones combinadas de los jugadores. Considerar este caso especial puede parecer poco interesante. Sin embargo, produce que arreglos en los que el estado del mundo también depende de la aleatoriedad en el desarrollo del juego (juegos Bayesianos) puedan ser reducidos a juegos en forma normal. En efecto, hay reducciones a forma normal para otras representaciones de juegos, tales como aquellos juegos que involucran un elemento de tiempo (juegos en forma extensiva). Como muchas otras representaciones de interés pueden ser reducidas a esta representación, podría decirse que la forma normal es la más fundamental en la teoría de juegos [23].

Definición 1.1 *Un juego en forma normal (finito, de n personas) es una tupla* (N, A, u)*, donde:*

- N *es un conjunto finito de jugadores, indizados por* i*;*
- $A = A_1 \times \cdots \times A_n$*, donde* A_i *es un conjunto finito de acciones disponibles para el jugador* i*. Cada vector* $a = (a_1, \ldots, a_n) \in A$ *es llamado un perfil de acción;*
- $u = (u_1, \ldots, u_n)$*, donde* $u_i : A \mapsto \mathbb{R}$ *es una función de utilidad (o de pagos) con valores reales para el jugador* i *[23].*

En un juego en forma normal los jugadores eligen sus estrategias de forma simultánea, es decir, que cada jugador elige su jugada son conocer las decisiones de los demás. Cada jugador recibe un pago $u_i(a_1, \ldots, a_n)$, dependiendo de las estrategias elegidas por los demás.

Como ejemplo prototipo de juegos en forma normal o estratégica destacan los juegos con dos jugadores, donde cada uno de ellos tiene un número finito de estrategias, por lo que las funciones de pago pueden representarse en una doble matriz. Tales juegos suelen denominarse **bimatriciales**.

		Jugador 2	
		C	D
Jugador 1	C	$(-1,-1)$	$(-3,0)$
	D	$(0,-3)$	$(-2,-2)$

Figura 1.3: Matriz de pagos en el Dilema del Prisionero.

	C	D
C	R	S
D	T	P

Figura 1.4: Generalización de los pagos en el dilema del prisionero

1.1.3. Ejemplos de juegos en forma normal

Dilema del prisionero

Este es uno de los juegos más conocidos en la teoría de juegos. Puede describirse inicialmente como sigue:

Dos personas (jugadores) han cometido un crimen y ahora se encuentran en salas separadas de una estación de policía. El fiscal ha hablado con cada uno sin que el otro lo supiera, diciéndole a ambos: "Si confiesas y aceptas testificar en contra del otro jugador, y el otro jugador no confiesa, entonces te dejaré ir. Si ambos confiesan, entonces los enviaré a ambos a prisión durante 2 años. Si no confiesas y el otro jugador sí lo hace, entonces serás condenado a 3 años en prisión. Si nadie confiesa, entonces ambos serán condenados por un cargo de menor gravedad, puesto que tenemos suficiente evidencia para condenarlos y cada uno irá a prisión por 1 año".

En este juego hay dos jugadores, disponiendo cada uno de dos estrategias puras C y D. C representa "cooperar" y D representa "desertar" (del inglés *defect*). La figura 1.3 muestra la matriz de pagos para este juego. Para cada una de las cuatro casillas de esa matriz, la primera componente de la pareja ordenada indica el pago para el jugador 1 y la segunda componente indica el pago para el jugador 2, en función del par de acciones ejercidas respectivamente. Nótese que los pagos en la matriz representan el tiempo perdido en términos de años en prisión. El término *cooperar* se refiere a la cooperación con el otro jugador al no confesar. El término *desertar* se refiere a confesar y aceptar testificar, rompiendo así el acuerdo (implícito) con el otro jugador [22].

El dilema del prisionero ha sido ampliamente usado para modelar situaciones en las cuales la cooperación mutua prima para el mejor resultado en términos sociales, pero los desertores pueden beneficiarse individualmente. En general, los pagos en este juego pueden describirse como lo muestra la matriz de la figura 1.4, donde la cooperación mutua produce la recompensa R, la deserción mutua produce el castigo P, y la mezcla de las opciones produce el pago más bajo S para el jugador incauto que cooperó y para el desertor produce el pago más alto T, considerado como la *tentación* (de engañar al otro jugador para obtener el máximo beneficio). La esencia del dilema está centrada en la desigualdad $T > R > P > S$. Ambos jugadores preieren un resultado en el que el oponente coopere, pero la mejor opción para ambos es desertar. En particular, la tentación de engañar ($T > R$) y el miedo a ser engañado ($S < P$) pueden poner la cooperación en riesgo [8].

		Jugador 2	
		E	F
Jugador 1	E	(1, 2)	(0, 0)
	F	(0, 0)	(1, 2)

Figura 1.5: Matriz de pagos de un juego asimétrico.

Juego de la gallina

El juego de la gallina es una competición de automovilismo o motociclismo en la que dos participantes conducen un vehículo en dirección al del contrario; el primero que se desvia de la trayectoria del choque pierde y es humillado por comportarse como una gallina. El juego se basa en la idea de crear presión psicológica hasta que uno de los participantes se echa atrás. Este juego es considerado como un juego estático con información completa.

Juegos simétricos y asimétricos

Un juego *simétrico* es un juego en el que las recompensas por jugar una estrategia en particular dependen solo de las estrategias que empleen los otros jugadores y no de quien las juegue. Si las identidades de los jugadores pueden cambiarse sin que vambien las recompensas de las estrategias, entonces el juego es simétrico. Muchos de los juegos 2×2 mas estudiados son simétricos. Las representaciones estándar del dilema del prisionero y el juego de la gallina son juegos simétricos. A los juegos que no cumplen las condiciones de simetría se les llama *asimétricos*. Un ejemplo es el representado en la matriz de pagos de la figura 1.5.

Juegos de pago común

Hay ciertos tipos de juegos en forma normal restringidos que merecen una mención especial. El primer tipo son los llamados juegos de pago común. Estos son juegos en los cuales para cada perfil de acción elegido, todos los jugadores tienen el mismo pago.

Definición 1.2 (Juegos de pago común) *Un juego de pago común es aquel en el que para todos los perfiles de acción $a \in A_1 \times \cdots \times A_n$ y cada par de agentes i, j, se cumple que $u_i(a) = u_j(a)$.*

También son llamados juegos de *coordinación pura* o *juegos en equipo*. En estos juegos los agentes no tienen *conflicto de intereses*; Su único desafío es coordinarse con una acción que sea máximamente beneficiosa para todos.
Por ejemplo, imagine dos personas conduciendo uno hacia el otro en un país sin normas de tránsito, donde cada uno debe decidir independientemente si conducir sobre el lado derecho o el lado izquierdo de la carretera. Si los conductores eligen el mismo lado, tendrán una utilidad alta, en caso contrario, tendrán una baja utilidad. La matriz de pagos es la mostrada en la figura 1.6.

Juegos de suma cero

En el otro extremo del espectro de juegos de coordinación pura, se encuentran los juegos de suma cero, los cuales son más propiamente llamados *juegos de suma constante*. A diferencia de los juegos de pago común, los juegos de suma constante son significativos principalmente en el contexto de los juegos de dos jugadores.

Definición 1.3 (Juegos de suma constante) *Un juego en forma normal de dos jugadores es de suma constante si existe una constante c tal que para cada perfil de acción $a \in A_1 \times A_2$ se cumple que $u_1(a) + u_2(a) = c$.*

	Izquierda	Derecha
Izquierdad	(1,1)	(0,0)
Derecha	(0,0)	(1,1)

Figura 1.6: Pagos en juego de conductores

	Cara	Cruz
Cara	(1,-1)	(-1,1)
Cruz	(-1,1)	(1,-1)

Figura 1.7: Pagos en Matching pennies

Por conveniencia, cuando hablemos de juegos de suma constante más adelante, estaremos asumiendo que $c = 0$, es decir, tendremos un juego de suma cero. Si los juegos de pago común representan situaciones de coordinación pura, los juegos de suma cero representan situaciones de competencia pura, donde la ganancia de un jugador va en detrimento del otro jugador. La razón por la que los juegos de suma cero son más significativos para dos agentes es que si se permiten más agentes, cualquier juego puede ser convertido en un juego de suma cero al agregar un jugador artificial cuyas acciones no impacten los pagos para los otros agentes, y cuyos pagos son escogidos para hacer cada suma de pagos venga a ser cero [23].

Un ejemplo clásico de juegos de suma cero es *Matching pennies*. Cada uno de los dos jugadores tiene que escoger simultáneamente cara o cruz. Si ambos jugadores escogen la misma opción, el jugador 1 recibe un dólar de parte del jugador 2. En caso contrario, el jugador 1 paga un dólar al jugador 2. La matriz correspondiente se presenta en la figura 1.7, donde la primera componente de cada pareja ordenada es el pago para el jugador 1 y la segunda componente es el pago para el jugador 2 [28].

Batalla de los sexos

En general, los juegos tienden a incluir elementos de coordinación y de competencia. El Dilema del Prisionero lo hace, aunque de una forma bastante paradógica. He aquí otro juego que incluye ambos elementos. En este juego, llamado *Batalla de los sexos*, un esposo y su esposa desean ir al cine, y ellos pueden escoger entre dos películas: "Arma Letal (AL)" y "Maravilloso Amor (MA)". La preferencia es ver la película juntos, pero la esposa (jugador 1) prefiere MA y el esposo (jugador 2) prefiere AL. La matriz de pagos es mostrada en la figura 1.8 [23].

1.1.4. Estrategias en juegos en forma normal

Hasta ahora hemos definido las acciones disponibles para cada jugador en un juego, pero no hemos hablado aún de su conjunto de estrategias o sus opciones disponibles. Ciertamente, un tipo de estrategia es seleccionar una sola acción y ejecutarla. Podemos llamar a ésta una *estrategia pura*, y usaremos la notación que hemos desarrollado para representar acciones. A la opción de cada agente de escoger

		Esposo	
		AL	MA
Esposa	AL	(2,1)	(0,0)
	MA	(0,0)	(1,2)

Figura 1.8: Batalla de los sexos

la estrategia pura la llamaremos *perfil de estrategia pura.*

Los jugadores podrían también seguir otro tipo de estrategia menos obvia: hacer una aleatorización sobre el conjunto de acciones disponibles, siguiendo alguna distribuición de probabilidad. Tal estrategia es llamada una estrategia mixta. A pesar de que no sea inmediatamente obvio el porqué un jugador debería introducir la aletoriedad dentro de sus opciones de acción, de hecho, en un juego multiagente el rol de las estrategias mixtas es fundamental [23].

Definición 1.4 (Estrategia mixta) *Sea (N, A, u) un juego en forma normal, y para cualquier conjunto X sea* $\prod(X)$ *el conjunto de todas las distribuciones de probabilidad sobre X. Entonces el conjunto de estrategias mixtas para el jugador i es* $S_i = \prod(A_i)$

Definición 1.5 (Perfil de estrategia mixta) *El conjunto de perfiles de estrategia mixta es simplemente el producto cartesiano de conjuntos de estrategia mixta* $S_1 \times \cdots \times S_n$.

Denotamos por $s_i(a_i)$ la probabilidad de que una acción a_i sea aplicada bajo la estrategia mixta s_i. El subconjunto de acciones que asignadas con probabilidad positiva por la estrategia mixta s:i es llamado el *soporte* de s_i [23].

Definición 1.6 (Soporte) *El soporte de una estrategia mixta* s_i *para un jugador i es el conjunto de estrategias puras* $\{a_i | s_i(a_i) > 0$

Note que una estrategia pura es un caso especial de una estrategia mixta para la cual el soporte consta de una sola acción. En el otro extremo tenemos las *estrategias completamente mixtas.* Una estrategia es completamente mixta si tiene un soporte lleno (es decir, si asigna una probabilidad diferente de 0 a cada acción) [23].

No hemos definido aún los pagos de jugadores dado un perfil de estrategia particular, puesto que la matriz de pagos se define solo para el caso especial de perfil de estrategia pura. Pero la generalización de estrategias mixtas es simple, y se basa en la noción básica de teoría de desiciones. Intuitivamente, primero calculamos la probabilidad de alcanzar cada resultado dado un perfil de estrategia, y luego calculamos el promedio de los pagos para los resultados, medido por las probabilidades de cada resultado. Formalmente definimos la utilidad esperada como sigue (abusando de la notación, usamos u_i para la utilidad y para la utilidad esperada) [23].

Definición 1.7 (Utilidad esperada de una estrategia mixta) *Dado un juego en forma normal* (N, A, u)*, la utilidad esperda* u_i *para el jugador i del perfil de estrategia mixta* $s = (s_1, \ldots, s_n)$ *es definida como*

$$u_i(s) = \sum_{a \in A} u_i(a) \prod_{j=1}^{n} s_j(a_j).$$

1.2. El aprendizaje

La palabra *aprendizaje* hace referencia a la adquisición de nuevos conocimientos o la modificación y refuerzo de conocimientos, comportamientos, habilidades, valores o preferencias existentes, implicando síntesis de información.

1.2.1. Modelos de aprendizaje

Existen varios modelos de aprendizaje en teoría de juegos y una buena forma de evaluarlos es a través de los datos obtenidos experimentalmente. Los modelos de aprendizaje pueden dividirse en: Enfoques evolutivos, aprendizaje reforzado, aprendizaje basado en la atracción del peso de la experiencia (EWA), imitación, aprendizaje con dirección y reglas de aprendizaje [30].

- Los enfoques **evolutivos** asumen que los organismos ya traen un programa previo; es decir, nacen con las estrategias que utilizarán para jugar. Este tipo de modelos son más aplicables a los animales con estrategias heredables o a la evolución de la cultura humana [7].
- Los enfoques de **aprendizaje reforzado** dan un paso adelante a los modelos anteriores puesto que toman en cuenta la sofisticación cognitiva que pueden tener los jugadores. Éstos asumen que cada estrategia es reforzada por los pagos que se recibirán. Además, asumen que los individuos tienen una habilidad de razonamiento imperfecta y puede ser aplicado a jugadores que no tienen ningún conocimiento previo acerca de las características del juego [30].
- Hay algunos enfoques que se basan en las **creencias** que los jugadores forman acerca de los otros, además se reconoce que estas creencias son modificadas por la historia de cada jugador. En este aprendizaje por creencia, los jugadores calculan los pagos esperados y eligen las estrategias con pagos esperados muy altos más frecuentemente que el resto de las estrategias. Además, se asume que la estrategia jugada recientemente y que dio buenas ganancias, tiene una probabilidad muy alta de ser elegida nuevamente [30].
- En el **aprendizaje anticipatorio** se asume que los jugadores no solo toman en cuanta el pasado de los otros, sino que además ponen su atención en las ganancias de sus contrincantes y asumen que estos resultados se relacionarán con las posibles estrategias que los contrincantes realizarán en el futuro. Así que no solo se trata de "adivinar" la conducta de los otros por medio de la observación de su pasado [24, 6].
- La teoría de **aprendizaje con dirección**, también combina la idea del aprendizaje por medio de creencias y del aprendizaje por refuerzo. En este caso, cada jugador utiliza una estrategia que ya ha elegido anteriormente y que le ha resultado funcional y la ajusta, dirigiéndola a las siguientes situaciones que se le presentan, de esta manera el aprendizaje dirige o limita su actuación [30].
- Los modelos basados en **reglas de aprendizaje** asumen que los jugadores tienen reglas de decisión y este tipo de procesos son los responsables de la elección de estrategias. Por lo tanto, las personas aprenden qué regla de decisión utilizar más que cuál estrategia usar [30].
- **El aprendizaje por atracción causada y ponderada por la experiencia**, llamado generalmente EWA (por su nombre original *Experience Weighted Attraction*) puede considerarse como una familia de reglas de aprendizaje que asumen que el aprendizaje por refuerzo y el aprendizaje basado en creencias son extremos de un continuo de experiencias y hace una conjugación de ambos enfoques. Sin embargo, agrega un elemento más, se asume que los jugadores con cada movimiento actualizan los valores o pesos de su información [30]. Nos interesa ver, en particular, cómo puede este mecanismo de aprendizaje explicar interacciones y dilemas sociales actuales.

1.2.2. Aprendizaje EWA

Se refiere al aprendizaje dado por la atracción causada y ponderada por la experiencia. Esto es lo que resume su nombre original en inglés *Experience Weighted Attraction*, del cual se desprende la sigla EWA. La palabra "ponderada" (weighted) en este sentido hace referencia al peso de importancia y de atracción que un jugador da a una determinada acción o estrategia de juego dependiendo de la experiencia alcanzada. Matemáticamente, este tipo de aprendizaje es descrito de la siguiente manera: Considérese un juego en el que participan p jugadores. Cada uno de ellos puede escojer de un conjunto

de N acciones (estrategias puras) en cada momento que tengan los jugadores para ejecutar una acción. En el modelo EWA la probabilidad para un jugador $\mu \in \{1, \ldots, p\}$ de escoger una acción $i \in \{1, \ldots, N\}$ en un tiempo t es

$$x_i^\mu(t) = \frac{e^{\beta Q_i^\mu(t)}}{\sum_k e^{\beta Q_k^\mu(t)}}, \tag{1.1}$$

donde las $\{Q_i^\mu\}$ son llamadas *atracciones* o *inclinaciones*. La idea básica es que $Q_i^\mu(t)$ indica la "atracción" del jugador μ hacia la acción i en un tiempo t, basado en qué tan exitosa ha sido la estrategia i en el pasado. El parámetro $\beta \geq 0$ es llamado la *intensidad de la elección*. Para $\beta = 0$ los jugadores seleccionan acciones con igual propabilidad (es decir, el juego es completamente aleatorio), y para $\beta \to \infty$ la elección de cada jugador es decisiva, es decir, que un jugador siempre escogerá una misma acción para valores dados de Q_i^μ, a saber, la acción con la atracción más alta [12].

La regla de actualización para las atracciones $\{Q_i^\mu\}$ en el modelo EWA indica:

$$Q_i^\mu(t+1) = \frac{\phi \mathcal{N}(t) Q_i^\mu(t) + [\delta + (1-\delta) I(i, s_\mu(t)]\, \prod^\mu(i, \mathbf{s}_{-\mu}(t))}{\mathcal{N}(t)}, \tag{1.2}$$

El valor de $\mathcal{N}(t)$ es actualizado de acuerdo a la ecuación

$$\mathcal{N}(t+1) = \phi(1-k)\mathcal{N}(t) + 1, \tag{1.3}$$

La notación es explicada a continuación:

- Escribimos $s_\mu(t) \in \{1, \ldots, N\}$ para la acción que el jugador μ toma en un instante t, en una determinada ejecución de las dinámicas. La notación $-\mu$ categoriza a todos los jugadores diferentes a μ, es decir, $-\mu$ es el conjunto $\{1, \ldots, p\} \backslash \{\mu\}$. La notación $\mathbf{s}_{-\mu}(t)$ representa el conjunto de acciones que los oponentes del jugador μ tomaron en una partida determinada. Así, $\mathbf{s}_{-\mu}(t) \in \{1, \ldots, N\}^{p-1}$ es un vector de $p-1$ componentes, siendo cada una de estas una de las acciones posibles.

- Para un dado $\mathbf{s}_{-\mu}(t) \in \{1, \ldots, N\}^{p-1}$, la cantidad $\prod^\mu(i, \mathbf{s}_{-\mu}(t))$ es un elemento de la matriz de pagos e indica el pago que el jugador μ recibe cuando utiliza la estrategia pura i y enfrentando las acciones $\mathbf{s}_{-\mu}(t)$ de los otros agentes en un tiempo t.

- La variable $\mathcal{N}$ indica un factor de de ponderación. Es necesario especificar una condición inicial, la cual es entonces actualizada de acuerdo a la relación (1.3). Cuando $\phi(1-k) = 1$ esto no es más que el número de veces que se ha jugado. En este caso, como $\mathcal{N}$ se cancela en el primer término del numerador en el miembro derecho de la ecuación 1.2, y como divide al segundo término, a medida que el tiempo avanza la influencia de las actualizaciones se vuelven más y más pequeñas, es decir, los movimientos pasados tienen más peso que los movimientos recientes y el comportamiento se vuelve "estable".

- La notación $I(\cdot, \cdot)$ se establece para la función indicadora (también llamada la función delta de Kronecker), es decir, $I(a, b) = 1$ si $a = b$ y $I(a, b) = 0$ en caso contrario.

- El parámetro δ especifica el peso relativo dado a las estrategias que son jugadas contra aquellas que no son jugadas. En el caso de $\delta = 1$ los jugadores actualizan todas las atracciones Q_i^μ en cada ronda,independientemente de qué acciones tomaron en realidad. La opción $\delta = 0$ corresponde al caso en el que solo los puntajes de las estrategias que son realmente usadas en una ronda dada son actualizados después de esa ronda.

- El parámetro k se interpola entre el aprendizaje por refuerzo promedio ($k = 0$) y el aprendizaje por refuerzo acumulativo ($k = 1$); tenemos $\mathcal{N}(t) = 1$ para todo t si $k = 1$, las atracciones Q_i^μ entonces representan el resultado acumulativo para todas las jugadas pasadas (dependiendo de la elección de ϕ potencialmente descontada con el tiempo), para $k = 0$ el factor normalizador $\mathcal{N}(t)$ crece con el tiempo.
- El parámetro ϕ especifica el peso de los resultados de jugar en el pasado lejano relativo a iteraciones más recientes. Si $k = 1$ y $\phi = 1$ todas las experiencias pasadas llevan el mismo peso, sin importar cuánto tiempo ha transcurrido, para $\phi = 0$ solo la ronda más reciente afecta las decisiones futuras de los jugadores. Los valores intermedios de ϕ corresponden a un descuento exponencial.

1.3. Sistemas complejos y teoría evolutiva de juegos

La evolución de la cooperación entre individuos que a su vez deben competir es aún uno de los grandes temas científicos abiertos (En orden de relevancia científica, aparece en la lista de los 25 primeros del número de aniversario de la revista *Science*[1] de 2005). Por un lado la competencia Darwiniana es necesaria para mejorar el rendimiento de los individuos. Sin embargo, la cooperación es omnipresente en la naturaleza e indispensable para lograr los grandes saltos evolutivos de los seres vivos (células eucariotas, reproducción sexuada, organismos multicelulares, comportamiento social y lenguaje, entre otros).

Dada la dinámica y la complejidad de las interacciones de los agentes considerados en estos casos, la física de los sistemas complejos aborda estos problemas utilizando modelos de la teoría de juegos. En este punto coinciden claramente dos tópicos importantes para seguir abordando la temática de la cooperación y el aprendizaje: Los *sistemas adaptativos complejos* desde la física y la *teoría evolutiva de juegos* desde las matemáticas.

1.3.1. Sistemas adaptativos complejos

Los sistemas complejos, caracterizados más por su comportamiento rico y complicado que por su definición intrínseca, aparecen en diversas áreas: Física, Matemática, Biología, Química, Ingeniería, Economía, etc. Lo que caracteriza estos sistemas es la presencia de un número muy elevado de agentes que interactúan entre sí y como consecuencia surgen comportamientos emergentes, los cuales están caracterizados por ser independientes del comportamiento aislado de cada agente involucrado. Dentro del estudio de estos sistemas, se consideran los llamados *sistemas adaptativos complejos*.

Los sistemas adaptativos complejos conservan las siguientes caracterísiticas especiales [20]:

- Están compuestos por un gran número de elementos fundamentales interactuando, enviando y recibiendo señales y ejecutando una secuencia de reglas de interacción.
- No hay un control centralizado.
- Cambian su estructura reorganizando sus elementos fundamentales para adaptarse a su entorno cambiante.
- El proceso de adaptación los hace difíciles de entender y controlar.

[1] En español, "Ciencia", es una revista científica y órgano de expresión de la Asociación Estadounidense para el Avance de la Ciencia, cuyo nombre original es *American Association for the Advancement of Science (AAAS)*.

Algunos ejemplos en el ámbito natural son: el cerebro, el sistema inmune, la ecología, los embriones, las colonias de hormigas, la dinámica celular del cáncer y la conciencia; y en el ámbito social/humano podemos considerar como ejemplos a las ciudades, los partidos políticos, las comunidades científicas, las economías locales y globales, los movimientos sociales, la cultura, las organizaciones, el lenguaje, la sustentabilidad ambiental, entre otras.

De acuerdo a John Holland [18], las características fundamentales dentro de estos sistemas son:

1. **Evolución:** Los elementos fundamentales que constituyen a los sistemas adaptativos complejos evolucionan y aprenden. Durante su proceso de evolución, los elementos fundmamentales mejoran su

2. **Comportamiento agregado:** El comportamiento agregado emerge de las iteracciones entre los elementos a diferentes escalas. En este comportamiento de interés para entenderlo y posiblemente modificarlo. Por ejemplo, en ecología se tienen los eslabones de la cadena alimenticia; en economía se tienen los eslabones de la cadena de demanda y suministro hasta llegar al producto interno bruto.

3. **Capacidad de anticipación:** A fin de adaptarse al entorno cambiante, los elementos de los sistemas adaptativos complejos desarrollan reglas condicionadas (condición/acción, IF/THEN) de interacción para anticipar las consecuencias de ciertas respuestas al entorno. Dado que los elementos son gobernados por sus propias reglas de interacción, constantemente las revisan e incluso las ajustan. Por lo tanto, la estructura de los sistemas complejos adaptativos está basada en estas reglas condicionadas de interacción.

   Esta estructura, habilita a los sistemas adaptativos complejos para adaptarse al entorno, presentando nuevas formas de comportamiento emergente. La adaptación de los sistemas adaptativos complejos requiere la solución a dos problemas asociados al cambio en las reglas de interacción condicionadas:

   i. Asignación de crédito.
   ii. Descubrimiento de nuevas reglas.

1.3.2. Teoría de juegos evolutivos

La teoría de juegos evolutivos difiere de la teoría de juegos clásica en que se concentra en las dinámicas de la estrategia en lugar de sus equilibrios. A pesar de su nombre, la teoría de juegos evolutivos se aplica más en economía que en biología.

Los objetivos originales de la toría de juegos consistían en encontrar principios generales del comportamiento racional. Se esperaba que éste resultara óptimo contra un comportamiento irracional. Se analizaban entonces experimentos imaginarios entre jugadores perfectamente racionales, y que sabían que sus oponentes también lo eran, y que usarían una estrategia similar. Resultó que se estaba pidiendo demasiado, que la racionalidad de los jugadores era una condición demasiado restrictiva y, en última instancia, perjudicial.

La "especie" de los jugadores racionales entró en el camino de la extinción al introducirse la doctrina de la "mano temblorosa": ¿Qué pasaría si un jugador creyese que, ocasionalmente su oponente haría la jugada "incorrecta" en lugar de la "correcta"? Es una cuestión sumamente "racional" de considerar. ¿Qué tan "ocasional" podría ser ese comportamiento? ¿Tiene relevancia que los jugadores se equivoquen "a propósito", habiendo concebido la jugada correcta, pero no llevándola a cabo? [1]

	Teoría de juegos clásica	Teoría de juegos evolutivos
Jugadores	Racionales	Irracionales
Estrategias	Elegibles	Heredadas
Interacción	Todos a la vez	Muestreo aleatorio de una población
Resultados	Utilidad	Capacidad reproductiva (fitness)

Figura 1.9: Diferencias de enfoques de estudio.

Una vez abierta la puerta de la irracionalidad, no hubo vuelta atrás: los jugadores ya no son "lógicos perfectos". Esta situación resultó en extremo favorable para la teoría, ya que abrió las puertas a innumerables aplicaciones en las ciencias sociales, desde la ética a la economía, desde la política hasta el comportamiento animal. Al no estar restringidos por el comportamiento racional, los jugadores pueden aprender, adaptarse, evolucionar. En lugar de encontrar la solución "perfecta" basada en consideraciones a priori, se trata ahora de estudiar el comportamiento dinámico de modelos de juegos definidos por estrategias, pagos y mecanismos de adaptación. No está para nada claro que la situación vaya a alcanzar un estado estacionario.

Ahora bien, ¿por qué es esto relevante para la biología? Todos los aspectos del comportamiento animal tienen alguna influencia en uno de los conceptos centrales de la evolución: la *fitness* (estrictamente, el número de descendientes vivos que el animal produce). La selección natural tiene por efecto que el comportamiento de un animal tienda a maximizar su fitness. El comportamiento óptimo que maximiza la fitness resulta de un balance entre beneficios y costos.

Por ejemplo, consideremos un animal macho que produce un sonido para atraer a las hembras y aparearse. La intensidad del sonido que produce tiene asociado un beneficio y un costo: cuanto más fuerte chille, por ejemplo, de más lejos lo oirán sus potenciales parejas; pero chillar más fuerte cuesta más energía. Ambas tendencias tendrán por efecto que la intensidad óptima se encuentre en algún valor intermedio, en principio seleccionado naturalmente. Ahora bien, en general este animal no se encontrará solo, sino que habrá otros machos tratando de atraer a las mismas potenciales parejas. De manera que el éxito de su comportamiento dependerá no solamente de cuán fuerte chille, sino de cuán fuerte chillen los demás. Una estrategia posible, por ejemplo, sería quedarse callado en medio de sus congéneres chillones. De esta manera no incurre en el gasto asociado a la llamada de apareamiento. Este comportamiento recible el nombre de *satélite*. Si las hembras no son muy selectivas al acercarse a los machos y se aparean indiscriminadamente con ellos, el macho satélite probablemente tenga una ventaja sobre los otros, ya que percibe los beneficios sin incurrir en los costos. En todo caso, lo que está claro es que la fitness ahora no depende simplemente de la relación entre costos y beneficios, sino que depende también de las estrategias, o modos de comportamiento, que ejerzan los participantes [1].

Podemos resumir las principales diferencias de los enfoques de la teoría de juegos evolutivos y la teoría de juegos clásica en la tabla de la figura 1.9.

Capítulo 2

El aprendizaje en el Dilema del Prisionero

La cooperación y la deserción están en el núcleo de todo dilema social [9]. Mientras que los individuos cooperadores contribuyen al bienestar colectivo a costo personal, los desertores son indiferentes ante los posibles beneficios colectivos buscando un bien individual. ¿Cómo aporta esto al estudio de los dilemas sociales? Estudios experimentales apuntan a que los humanos presentan características especiales en la evolución y las dinámicas con las que actualizan sus estrategias a lo largo de diferentes iteraciones, sugieriendo que, por ejemplo, la cooperación de cada individuo no depende tanto de los pagos como de la cooperación que observa en el resto de juegadores, siendo más propensos a contrubuir tanto como sus colegas lo hagan [11, 17, 13].

Estas evoluciones en el comportamiento y actualizaciones de estrategias evidencian diferentes tipos de mecanismos de aprendizaje en los humanos. Es entonces importante analizar cómo el estudio del comportamiento humano en el Dilema del Prisionero contribuye al estudio de dinámicas de aprendizaje o viceversa. En particular, teniendo modelos matemáticos como el del aprendizaje EWA, se puede discutir acerca del nivel de relación que se establecería entre esos modelos y la características matemáticas propias del Dilema del Prisionero.

Básicamente, muchos dilemas sociales pueden ser formalizados como juegos de dos personas donde cada jugador puede cooperar (C) o desertar (D). El Dilema del prisionero ha sido ampliamente usado para modelar situaciones en las que la cooperación mutua conduce al mejor resultado en términos sociales, pero los desertores pueden obtener los mejores beneficios individuales.

2.1. La cooperación en humanos

Para entender la trascendencia del aprendizaje en el Dilema del Prisionero es necesario analizar las particularidades de los comportamientos cooperativos de los humanos que han jugado éste y otros juegos afines. El dilema del prisionero ha sido objeto de diferentes tipos de investigación, desde diversas áreas del conocimiento, tanto a nivel teórico como experimental.

Resultados experimentales [26, 14, 17] muestran que los humanos frecuentemente actúan de forma más cooperativa de lo que dictaría el simple interés personal. Una posible razón de ello, es una situación del Dilema del prisionero. El dilema del prisionero repetido en el tiempo hace que la “no cooperación” se castigue de una manera más severa y que el caso contrario, es decir la intención o acción de cooperar,

se premie más de lo que podría sugerir el problema original.

Se mencionarán algunos estudios y experimentos realizados (cuyos procedimientos y resultados serán discutidos en la siguiente sección) en los que se ha descubierto que la cooperación puede tomar diversas condicionalidades y las estrategias de los individuos se actualizan según algunos patrones y modelos.

2.1.1. La evolución de la cooperación

Investigadores como Martin Nowak han estudiado intensamente el surgimiento de la cooperación en la biología evolutiva en numerosas situaciones. Sus simulaciones matemáticas les han permitido demostrar que en ciertas condiciones, la cooperación surge como una estrategia evolutiva ventajosa. Esas condiciones están ligadas a los costes y beneficios de la cooperación. Por ejemplo H. Ohtsuki y M. Nowak probaron que en una red social, donde cada individuo trata de imitar al azar a los individuos más exitosos con los que tiene relación, si k es el número de individuos con los que otro está relacionado en promedio, c el coste promedio de cooperar y b el beneficio promedio obtenido de cada individuo cooperador, la estrategia de cooperación se torna dominante cuando $\frac{b}{c} > k$ [25].

2.1.2. Cooperación Condicional (CC)

La reciprocidad o cooperación condicional es aquella cooperación basada en elegir estrategias dependiendo de cuánta cooperación se reciba de los otros jugadores. Ha sido estudiada en juegos de dos jugadores a través del concepto de estrategias reactivas, que hace referencia a las acciones, reacciones e intervenciones planificadas en respuesta a los comportamientos identificados en los demás. La más famosa de éstas es *Tit-For-Tat*, la cual consiste en jugar lo que el oponente jugó en la ronda anterior. Más adelante observaremos datos experimentales acerca de esta estrategia [2].

Las estrategias reactivas generalizan esta idea considerando que los jugadores escogen sus acciones con probabilidades que dependen de la acción previa del oponente. Un desarrollo más avanzado fue considerar las estategias reactivas de memoria uno, en las cuales las probabilidades dependen de la acción previa de ambos jugadores. En juegos multijugadores, la cooperación condicional ha sido reportada en experimentos que se discutirán más adelante [11, 29].

En particular, el análisis de dos experimentos a gran escala con humanos jugando un Dilema del Prisionero multijugador iterado en red [17, 13] extendió esta idea al mencionar que la cooperación de un jugador también depende de su propia acción previa, dando paso a la regla conocida como cooperación condicional de ánimo cambiante, llamada originalmente en inglés *moody conditional cooperation* (MCC).

2.1.3. Cooperación condicional de ánimo cambiante o *Moody Conditional Cooperation* (MCC)

La estrategia MCC puede describirse matemáticamente de la siguiente manera [15]: Si en la ronda previa el jugador desertó, cooperará con probabilidad

$$p_D = q \tag{2.1}$$

(independientemente de la cooperación observada), mientras que si en la ronda previa el jugador cooperó, cooperará otra vez con probabilidad

$$p_C(x) = px + r \tag{2.2}$$

(sujeto a la restricción $p_C(x) \leq 1$), donde x es la fracción de vecinos cooperativos.

En otras palabras, la probabilidad de que un jugador que coopera en una ronda dada t del juego decida cooperar en la ronda inmediatamente posterior $t + 1$ aumenta linealmente con la fracción de agentes que cooperan en su interacción con él en t, y se mantiene constante si el jugador en cuestión por el contrario no coopera en t.

Hay muchas evidencias que soportan este comportamiento, como los experimentos a gran escala mencionados anteriormente [17, 13] y otro sobre el Dilema del Prisionero en modo multijugador [16].

2.2. Experimentos de cooperación en humanos

Se cita a continuación una serie de experimentos realizados con humanos, relacionados con las dinámicas de cooperación observadas en interacciones sociales reales. Mencionaremos los métodos y los resultados de reelevancia obtenidos para el estudio de las estrategias y condicionalidades relacionadas con la cooperación.

2.2.1. El comportamiento social según la teoría de juegos

Entender las interacciones entre las personas y sus contactos sociales es un problema clave para dilucidar la forma en la que funciona la sociedad y cómo ésta contribuye a la mejora del bienestar individual. El origen evolutivo de la cooperación entre individuos no emparentados es una cuestión sin resolver que afecta a varias disciplinas. Entre los distintos mecanismos propuestos para explicar cómo puede aparecer la cooperación destaca la existencia de una estructura en la población que determine las interacciones entre individuos.

Muchos modelos han explorado analítica y computacionalmente los efectos de dicha estructura, sobre todo en el marco del Dilema del Prisionero, pero los resultados obtenidos dependen enormemente de muchos detalles, tales como el tipo de estructura considerada o la dinámica evolutiva. Por tanto, era preciso llevar a cabo un trabajo experimental diseñado apropiadamente para identificar qué características de las que integran los modelos son las relevantes. En la tesis doctoral de Jelena Grujić [14] se investiga cómo la estructura espacial influye en la promoción de la cooperación.

Para ello, se diseñó un experimento con el fin de estudiar la aparición de cooperación cuando las personas juegan al Dilema del Prisionero iterado. Los voluntarios que participaron en este experimento jugaron al Dilema del Prisionero en una red de tamaño considerable. Los parámetros del experimento se escogieron para promover la cooperación en la mayor medida posible, partiendo de las predicciones de los modelos teóricos.

En el experimento, los voluntarios jugaron el Dilema del Prisionero 2×2 (en pareja) con cada uno de sus ocho vecinos (vecindad de Moore[1]) tomando solo una acción: cooperar (C) o desertar (D), siendo esta la misma contra todos los oponentes. Los pagos resultantes para cada jugador fueron calculados sumando los pagos de sus ocho interacciones.

[1] Conjunto de las ocho celdas que rodean una celda central en un enrejado cuadrado bidimensional.

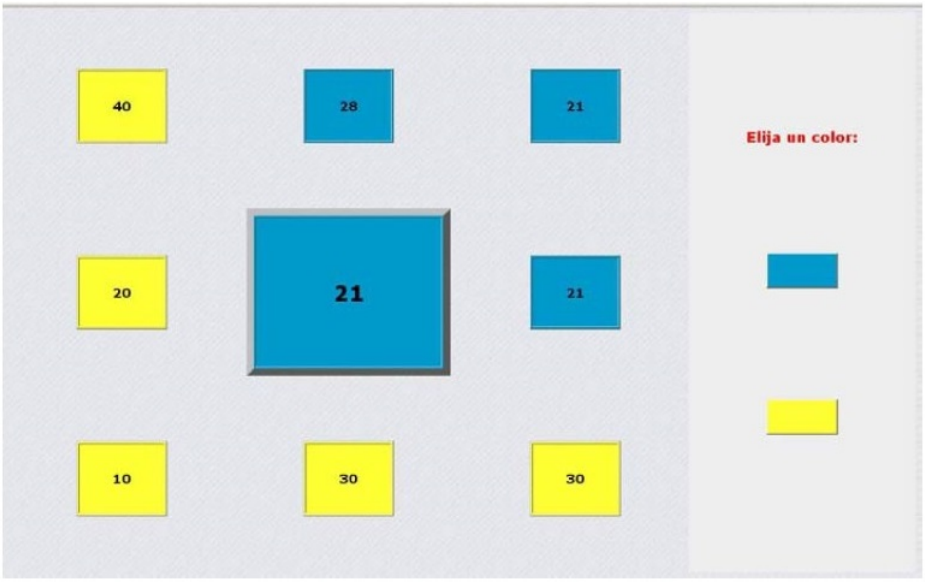

Figura 2.1: Información dada en la elaboración del experimento [14].

Los pagos en el Dilema del Prisionero fueron ajustados a 7 centavos de euro para la cooperación mutua, 10 centavos para el desertor enfrentando a un cooperador, y 0 centavos para cualquier jugador que enfrenta a un desertor. Con estas opciones (un cooperador y un desertor reciben el mismo pago contra un desertor) la deserción no es una estrategia de riesgo dominante, lo cual aumenta la posibilidad de que surja la cooperación. Además, para evitar el *efecto framing*[2], las dos acciones fueron siempre referidas en términos de colores (azul para C y amarillo para D), y el juego nunca fue referido como "Dilema del Prisionero" en el material facilitado a los voluntarios. Sin embargo, los jugadores fueron propiamente informados de las consecuencias de elegir cada acción, y les fueron dados algunos ejemplos en la introducción. Después de cada ronda, los jugadores fueron informados de las acciones tomadas por sus vecinos y sus correspodientes pagos (ver figura 2.1).
El experimento completo consistió de tres partes: *experimento 1*, *control* y *experimento 2*.

- En el **experimento 1** los jugadores mantuvieron la misma posición en la red con los mismos vecinos a lo largo de todo el experimento.
- En el **control** se removió el efecto de redes revolviendo los jugadores en cada ronda.
- Finalmente en el **experimento 2**, los jugadores fueron otra vez fijados en una red, aunque en posiciones diferentes del experimento 1.

La figura 2.2 representa los porcentajes totales de acciones cooperativas en cada una de las rondas de las tres partes del experimento:

- El *experimento 1* comienza con un porcentaje alto de cooperación, sobre el 50 %, que decae rápidamente hasta alcanzar un nivel más o menos constante después de 25 rondas.
- El *experimento 2* exhibe el mismo comportamiento, pero el nivel de cooperación inicial es mucho más bajo (32 %) y la transición es más corta.

[2]Las personas reaccionan a una opción particular de diferentes formas dependiendo cómo se les presente: como una ganancia o una pérdida.

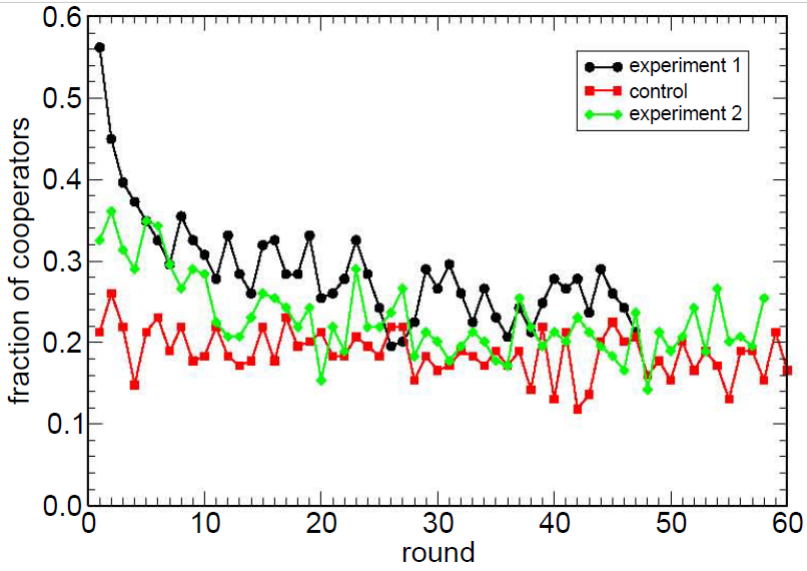

Figura 2.2: Fracción de cooperadores en cada ronda de las tres partes del experimento. El nivel de cooperación decae a un nivel bajo pero distinto de cero [14].

- Contrario a esto, la parte de *control* muestra una fracción constante de acciones cooperativas, fluctuando alrededor del 20 %. Esto es una señal clara de que los jugadores se dieron cuenta que el hecho de que los vecinos cambiaran después de cada ronda hacía que fuera imposible tratar de lograr un entorno mutuamente rentable, lo cual trataron de establecer al principio de los experimentos 1 y 2.

Los resultados indican que el nivel de cooperación no mejora por la existencia de una red, manteniéndose la fracción de cooperadores en un 20 % aproximadamente. Estos resultados se pueden explicar a través de la existencia de heterogeneidad y de una estrategia de cooperación condicional generalizada, en la que la probabilidad de cooperar depende de la cooperación de los otros participantes en el juego y también de la acción previa del jugador.

Estas conclusiones han tenido un gran impacto en la manera en la que la Teoría de Juegos en grafos[3] se usa para modelar las interacciones humanas en grupos estructurados. De hecho, se propone (al igual que en [15]) un modelo basado en agentes en el que coexisten tres diferentes estrategias compatibles con las observaciones experimentales: cooperación, deserción y cooperación condicionada generalizada.

Se consideraron grupos de $n = 2, 3, 4, 5$ jugadores y se calcularon los pagos para cada tipo de jugador en el equilibrio utilizando cadenas de Markov[4]. De esta manera, se demostró que para los grupos de tamaño menor que $n = 4$ existe un punto interior en el cual las tres estrategias coexisten. La correspondiente cuenca de atracción disminuye al aumentar el número de jugadores, mientras que para $n = 5$ no se pudo encontrar ningún punto de atracción interior. Finalmente, se observó que para el límite cuando n tiende a infinito, dicho atractor no existe.

Así pues, estos experimentos en red sugieren que la cooperación puede depender de la acción previa

[3]Conjunto de objetos llamados vértices o nodos unidos por enlaces llamados aristas o arcos. En teoría de juegos y física estadística, los grafos permiten estudiar las interrelaciones entre individuos que interactúan unos con otros.

[4]Proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende solamente del evento inmediatamente anterior. Esta característica de falta de memoria recibe el nombre de propiedad de Markov.

del jugador, pero al mismo tiempo se probó teóricamente que ese tipo de comportamiento no puede coexistir con jugadores que nunca cooperan y con cooperadores en grupos formados por más de 5 personas. Por ello, se decidió diseñar un experimento que reprodujese el esquema teórico. Así se confirmó la existencia de cooperadores condicionales y un nivel de cooperación bajo en grupos formados por más de dos miembros.

Soprendentemente, se vió que el comportamiento de los jugadores en grupos de dos individuos es cualitativamente diferente a las situaciones donde este número es mayor. El experimento se prolongó durante 100 rondas, lo cual permitió estudiar el régimen a largo plazo. Cuando se juega al Dilema del Prisionero por parejas en esta situación, el nivel de cooperación, tras una caída inicial, se incrementa significativimente y llega a un nivel de más del 80 %.

Además, se reanalizaron los datos del experimento de Traulsen [29], en el que los voluntarios jugaban al Dilema del Prisionero con sus cuatro vecinos más cercanos en una red de tamaño 4×4. El experimento tenía dos tratamientos:

- Uno **espacial**, donde los jugadores tenían una posición fija en la red durante todo el experimento
- Uno **no espacial**, en el cual los jugadores cambiaban sus posiciones en la red después de cada ronda.

Se analizaron estadísticamente las decisiones individuales y se dedujo con qué modelo o modelos de Teoría de Juegos evolutivos se pueden conectar. No se encontró ninguna diferencia entre ambos tratamientos. Sin embargo, las estrategias que usan los jugadores no corresponden con las que se suelen estudiar en Teoría de Juegos evolutivos.

Finalmente, utilizando simulaciones numéricas, se vio cómo los mecanismos de actualización obtenidos en los experimentos no favorecen la cooperación en la estructura espacial. Como apoyo a estas conclusiones, se compararon los resultados de experimentos diferentes. Aunque hay diferencias, ciertas características parecen ser universales. Así, el nivel de cooperacion se muestra bajo en todos los experimentos, a pesar de que muchos modelos teóricos predicen una promoción de la cooperación, y la estructura de la población (la red) parece no tener ningún efecto sobre el nivel de cooperación.

En todos los experimentos se observa cooperación condicional generalizada, aunque también es posible describir el comportamiento observado con otras reglas, si bien de manera menos universal que con la anterior.

2.2.2. Experimentos con el Juego del Bien Público

Además del Dilema del Prisionero, otro juego que se ha utilizado para estudiar la cooperación de individuos dentro de una colectividad es el llamado *juego del Bien Público*.

Imaginemos que 4 ciudadanos reciben 20 unidades monetarias cada uno y deben decidir qué parte de esa suma destinan a un proyecto común o bien público. Por cada unidad monetaria que se destine al proyecto, cada uno de los 4 ciudadanos recibirá un beneficio de 0,4 unidades (con independencia de que haya o no contribuido), de forma que el beneficio conjunto será de 1,6 unidades.

A pesar de esa gran rentabilidad del proyecto, cada ciudadano tendrá la tentación de hacerse el de "la vista gorda" a la hora de financiarlo y sacarle provecho bajo ningún costo personal (free-riding[5]), pues por cada unidad que aporte sólo recibirá un beneficio individual de 0,4 unidades. Ahora bien,

[5]El término inglés *free-riding*, muy usado en juegos no cooperativos, hace alusión a usar un transporte público sin pagar pasajes.

si todos ceden a esa tentación, el proyecto público no saldrá adelante y el grupo no aprovechará esa gran oportunidad. La experiencia muestra que, cuando el juego se desarrolla entre personas que no se conocen ni pueden comunicarse entre sí, muchos jugadores terminan sucumbiendo a la tentación y la cooperación acaba decayendo.

La situación descrita es conocida en Teoría de Juegos como el "Juego del Bien Público" (Del inglés *Public Good Game*) porque ilustra el dilema social que surge cuando un bien o una inversión beneficia a todos los miembros de un grupo social, con independencia de que contribuyan a sufragarlo. En tal caso, el ideal egoísta para cada miembro es que el bien se produzca, pero que lo financien los demás y, en consecuencia, existe la tentación de actuar como un *free-rider* ("aprovechado","oportunista").

Dado que este problema se plantea con frecuencia en muchas situaciones sociales, el juego del bien público ha sido objeto de muchos experimentos. A continuación se citarán algunos con resultados relevantes para las temáticas aquí tratadas.

Dinámica de contribuciones y cooperación en un juego de bien público

Uno de los experimentos que se han basado en el juego del bien público se encuentra en un artículo de los investigadores españoles Pablo Brañas y María Paz Espinosa [3]. Éstos llevaron a cabo el experimento en mayo de 2007 con 48 estudiantes de la Universidad de Granada, con el fin analizar cómo los jugadores crean espectativas del juego y cómo esto afecta sus actualización de estrategias y sus contribuciones al bien colectivo.

A los participantes se les propuso realizar dos tipos de tareas. La primera, consistió en jugar un juego de bien público durante cinco periodos o rondas de juego, organizados en grupos de 4 individuos, explicándoles que estarían jugando con los mismos compañeros todos los periodos. En cada periodo los sujetos podrían contribuir entre 0 y 100 unidades de moneda para una cuenta colectiva, pero también podrían guardarlos para una cuenta individual privada. Después de cada periodo, el dinero total aportado a la cuenta colectiva por los 4 jugadores se multiplicaría por 1.5 y lo que resultara sería dividido equitativamente entre los 4 jugadores. La ganancia final de cada jugador es la suma de los pagos obtenidos en cada uno de los cinco periodos.

Después de transcurrir los cinco periodos y recibir los pagos correspondientes, se continuó con la segunda tarea. Para ésta, se le preguntó a cada jugador sobre lo que creeía que había aportado cada uno de los 48 participantes en cada uno de los periodos, considerando las contribuciones observadas durante los cinco periodos. Las contribuciones promedio en la tarea 1 (contributions) y los promedios de las creencias de los jugadores recolectadas en la tarea 2 (beliefs) se representan en la figura 2.3.

La diferencia media entre las acciones y las creencias fue relativamente pequeña para los primeros tres periodos. Sin embargo, el promedio de la diferencia incrementó en los periodos 4 y 5. Hubo un *efecto de fin de juego* (desviación en la secuencia de comportamiento) en el periodo 4, pero la predicción media no lo incorporó. Los individuos no predijeron el efecto de fin de juego y las contribuciones y creencias divergieron.

En general se obtuvieron los siguientes datos:

- Contribuciones: El 25 % (12 de 48) de los individuos disminuyeron su contribución en el periodo 4; 12,5 % no contribuyeron en el periodo 5; pero un alto porcentaje de individuos (23 %) no disminuyeron su contribución al acercarse el final del juego.
- Creencias: 25 de 48 individuos (52 %) no predijeron ningún efecto de fin de juego; 10 individuos

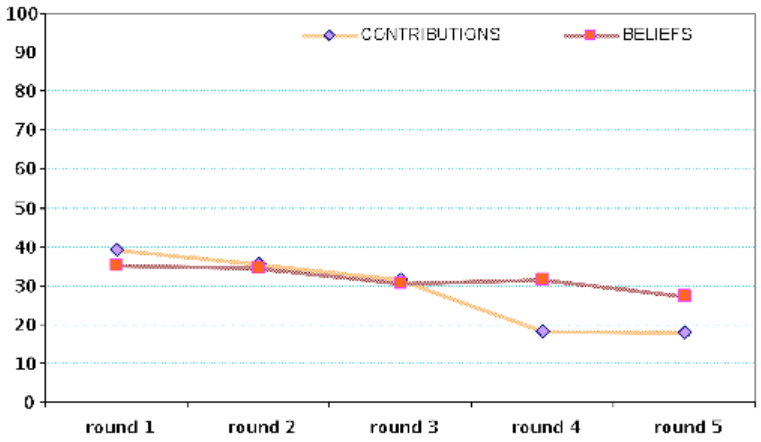

Figura 2.3: Comparación del promedio contribuciones reales (contributions)en la tarea 1 con el promedio contribuciones individuales según las creencias de los mismos jugadores (beliefs) en la tarea 2 [3]

(21 %) creyeron que el efecto de fin de juego ocurriría en el último periodo y solo uno hizo la predicción correcta (disminuyo en el periodo 4).

Un análisis regresivo permitió medir la importancia relativa de los antecedentes y los datos en la formación de las creencias de los individuos. Uno de los resultados más importantes es que los antecedentes de comportamiento tienen una influencia muy significativa en comparación con los datos y valores observados por los jugadores a medida que transcurre el juego.

Este análisis sugiere que, antes de jugar, los sujetos no esperan hacer razonamientos acerca de periodos anteriores, ni siquiera en los últimos periodos, y que las actualizaciones de estrategias de cooperación usando los valores y datos observados mientras transcurre el juego son lentas. Con esto se ve además, que la tasa de disminución de la contribución y la sostenibilidad de la cooperación puede estar relacionada con la forma en la que los jugadores aprenden y crean expectativas.

Sobre los porcentajes medios de aportación se encontró que:

- Son positivos (es decir, los jugadores no son "aprovechados"), pero siempre están por debajo del 100 %. Un porcentaje frecuente de aportación inicial es del 40-60/
- Son tanto mayores cuanto mayor sea la rentabilidad per capita del proyecto.
- Son mayores cuando los jugadores han tenido posibilidad de comunicarse entre sí antes del juego (supuesto que habíamos excluido en nuestro ejemplo).
- Suele ir bajando cuando el juego se juega en rondas sucesivas entre desconocidos.

Dos economistas suizos, Ernst Fehr y Simon Gächter [10], introdujeron en el juego una pequeña variante: añadieron una fase final en la que los jugadores, tras conocer cuánto ha contribuido cada uno al bien público, tienen oportunidad de castigarse entre sí. En particular, por cada unidad monetaria a la que renuncie el castigador, el destinatario del castigo perderá 3 unidades. Se trata, pues, de un castigo muy eficaz, pero "altruistao", porque quien lo ordena soporta un coste, por lo cual es llamado también "castigo altruista" (altruistic punishment). La figura 2.4 presenta la comparación entre la cooperación media presentada en uno de los experimentos sin incluir castigos y la presentada incluyendo castigos..

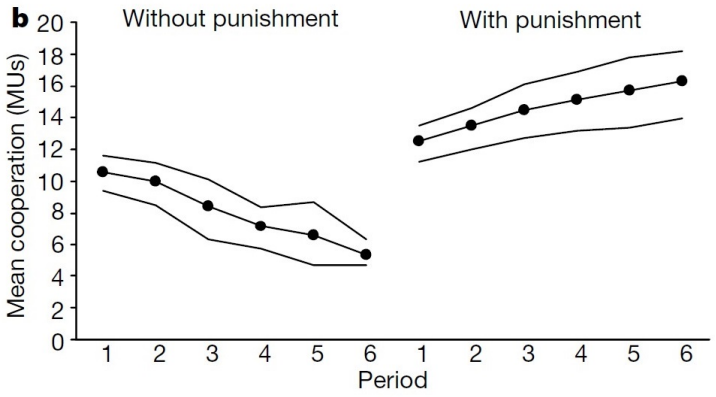

Figura 2.4: Relación de la cooperación media (eje vertical) en cada periodo de juego (eje horizontal) observada en uno de los experimentos de Bien Público [10]. Los primeros seis periodos se jugaron sin castigos y los otros seis con castigos.

Los experimentos de Fehr y Gächter son concluyentes: un elevado porcentaje de jugadores (el 84 %, en uno de los experimentos) suele aplicar castigos altruistas; los más proclives a castigar son quienes más contribuyen; y, en fin, la amenaza del castigo eleva mucho la aportación media de los jugadores. Así pues, "el acto de castigar, aunque costoso para el castigador, proporciona un beneficio a los demás miembros de la población al inducir a que los potenciales no-cooperadores aumenten sus inversiones" [10].

En [11] se realizó un experimento basado en el juego del Bien Público, con las siguientes especificaciones

2.2.3. El dilema del prisionero iterado

En 1984 fueron publicados los estudios de Robert Axelrod[6] acerca de una extensión al escenario clásico del dilema del prisionero que denominó D*ilema del Prisionero Iterado* [2]. Aquí, los participantes deben escoger una y otra vez su estrategia mutua, y tienen memoria de sus encuentros previos. Axelrod invitó a compañeros académicos a lo largo del mundo a idear estrategias automatizadas para competir en un torneo del Dilema del Prisionero Iterado. Los programas que participaron variaban ampliamente en la complejidad del algoritmo: hostilidad inicial, capacidad de perdón y estrategias similares.

Axelrod descubrió que cuando se repiten estos encuentros durante un largo periodo de tiempo con muchos jugadores, cada uno con distintas estrategias, las estrategias egoístas tendían a ser peores a largo plazo, mientras que las estrategias altruistas eran mejores, juzgándolas únicamente con respecto al interés propio. Usó esto para mostrar un posible mecanismo que explicase lo que antes había sido un difícil punto en la teoría de la evolución: ¿Cómo puede evolucionar un comportamiento altruista desde mecanismos puramente egoístas en la selección natural?.

Se descubrió que la mejor estrategia era la que se traduce como *Tal para cual*, del inglés *Tit-for-Tat*, que fue desarrollada y presentada en el torneo por Anatol Rapoport[7]. Consiste simplemente en coope-

[6](1943) Matemático, profesor de ciencias políticas y políticas públicas en la Universidad de Míchigan, Estados Unidos.
[7](1911 - 2007) Científico ruso. Uno de los primeros en aplicar modelos matemáticos avanzados de redes a problemas sociales.

rar en la primera ronda del juego, y después de eso elegir lo que tu oponente eligió la ronda anterior. Era el más simple de todos los programas presentados, y fue la estrategia que ganó el concurso.

La estrategia *Tal para cual* permite la recuperación ocasional de quedarse encerrado en un círculo de deserciones. La probabilidad exacta depende de la alineación de los oponentes. Es la mejor estrategia cuando se introducen problemas de comunicación en el juego. Esto significa que a veces tu jugada se transmite incorrectamente a tu oponente: tú cooperas pero tu oponente cree que has desertado.

Axelrod afirmaba que el éxito de la estrategia *Tal para cual* se debía a dos motivos. El primero es que es "amable", esto es, comienza cooperando y sólo deserta como respuesta a la deserción de otro jugador, así que nunca es el responsable de iniciar un ciclo de deserciones mutuas. El segundo es que se le puede provocar, al responder siempre a lo que hace el otro jugador. Castiga inmediatamente a otro jugador si éste deserta, pero igualmente responde adecuadamente si cooperan de nuevo. Este comportamiento claro y directo significa que el otro jugador entiende fácilmente la lógica detrás de las acciones de la estrategia, y puede por ello encontrar una forma de trabajar con él productivamente.

No es una coincidencia que la mayoría de las estrategias que peor funcionaron en el torneo de Axelrod fueron las que no estaban diseñadas para responder a las elecciones de otros jugadores. Contra ese tipo de jugador, la mejor estrategia es desertar siempre, ya que nunca se puede estar seguro de establecer una cooperación mutua fiable.

2.3. El aprendizaje en la cooperación y la interacción social

Después de analizar estos experimentos y sus resultados, es natural suponer que los cambios en las estrategias cooperativas y los procesos de aprendizaje que experimentan los jugadores van de la mano. Incluso, puede afirmarse que el comportamiento MCC surge del proceso de aprendizaje experimentado por los jugadores [8]. Por ejemplo, en el experimento estudiado anteriormente con 100 iteraciones del Dilema del Prisionero [16], se llevaron a cabo análisis para comprobar los efectos del aprendizaje, encontrando fuertes evidencias que soportan la estrategia MCC. Esta relación nos lleva a pensar cómo puede un modelo de aprendizaje explicar matemáticamente comportamientos cooperativos utilizando la dinámica y los parámetros del Dilema del Prisionero.

2.3.1. El aprendizaje por refuerzo: Dinámica fundamental en la cooperación

Con el propósito de explicar los comportamientos de cooperación condicional, Cimini y Sánchez [8] proponen ir más allá de las estrategias de naturaleza imitativa (en la que los jugadores simplemente copian parámetros de determinados contrincantes) al estudiar dos dinamícas evolutivas innovadoras, en el sentido que permiten introducir estrategias ausentes en la población (las estrategias imitativas no lo permiten).

La primera es la regla de *Mejor Respuesta*, llamada originalmente en inglés *Best Response*, que representa una situación en la cual cada jugador tiene las suficientes habilidades cognitivas para calcular una estrategia óptima conociendo lo que sus vecinos hicieron en la ronda previa; y la segunda dinámica evolutiva utilizada es *Reinforcemente Learning* [21], el cual encarna la condición de un jugador que usa su experiencia para elegir o evitar ciertas acciones basándose en sus consecuencias: las acciones que cumplieron o superaron las espectativas en el pasado, tienden a ser repetidas en el futuro, mientras que las acciones que dejaron experiencias insatisfactorias son evitadas. Nótese que ninguna de esas dos

reglas dependen de usar información de los pagos de los demás jugadores.

Para comenzar, se tuvieron en cuenta diferentes estructuras espaciales que determinan las interacciones entre los jugadores:

- El diseño simple para una población bien mezclada (modelada por un grafo aleatorio de grado medio $\overline{k} = m$ reorganizada después de cada ronda).
- Estructuras más complejas como la *red libre de escala*[8] de Barabási-Albert (con distribución de grado[9] $P(k) \sim k^{-3}$ y $\overline{k} = m$).
- Redes regulares con condiciones de frontera periódicas (donde cada nodo está conectado a sus $k = m$ vecinos más cercanos).

Las simulaciones experimentales fueron llevadas a cabo usando los siguientes parámetros: $c_0 = 0{,}5$ (fracción inicial de cooperadores), $R = 1$, $P = 0$, $S = -\frac{1}{2}$, $T = \frac{3}{2}$ (entradas de la matriz de pagos del Dilema del Prisionero, tales que $T > R$, $S < P$, $2R > T + S$), $N = 1000$ (nodos de la red)y $m = 10$ (grado medio del grafo). Los parámetros de comportamiento MCC $\{q, p, r\}$ son todos fijados para cada jugador antes de la primera ronda del juego de una distribución uniforme $\mathcal{U}[0, 1]$, con la restricción adicional $p + r \leq 1$ para tener $0 \leq p_C(x) \leq 1$. Note que tanto la forma particular de la distribución inicial como la existencia de la restricción no afectan los resultados de los experimentos.

Para analizar la evolución del nivel de cooperación se hace uso entonces de la regla del *aprendizaje por refuerzo*. Se asume primero que el nivel de aspiración[10] A de los jugadores se mantiene fijo a lo largo del tiempo. También se considera el caso en el que los jugadores adaptan su nivel de aspiración después de cada ronda bajo la regla de actualización $A^{t+1} = (1 - h)A^t + h\pi^t/k$, donde h es la tasa de adaptación y $P < A^0 < R$. Los resultados basados en esta dinámica son presentados en la figura 2.5. Cuando A es un valor entre los pagos de castigo P y recompensa R ($P < A < R$), podemos observar un nivel de cooperación estacionario distinto de cero, de aproximadamente un 30 %, que no depende de la estructura de la población. El resultado más sobresaliente de esta dinámica es que, contrario a los otros procesos de actualización que se mencionaron, los valores de los parámetros de MCC $\{q, p, r\}$ se concentran alrededor de algunos valores estacionarios distintos de cero, los cuales son independientes de la estructura y de las condiciones iniciales del sistema.

Así, el aprendizaje por refuerzo resultó siendo el único mecanismo (sobre los otros mencionados en el artículo) que permitió la manifestación del MCC evolutivamente estable y al mismo tiempo la reproducción del nivel de cooperación con ausencia de reciprocidad de red.

Es preciso mencionar otras dos presentaciones de estas dinámicas. Primero, se verificó que el valor de λ influye solo sobre la tasa de convergencia del sistema; sin embargo, si los jugadores aprenden muy rápido ($\lambda \sim 1$) entonces los parámetros cambian drástica y rápidamente para alcanzar los valores estacionarios. Segundo, si introducimos em el sistema una fracción d de jugadores que siempre desertan, pasa que el nivel de cooperación final cambia: baja a 25 % para $d = 0{,}2$ y a 20 % para $d = 0{,}4$; pero las distribuciones estacionarias de los parámetros del MCC no son afectadas. Esto significa que el aprendizaje por refuerzo explica la heterogeneidad de los comportamientos observados en las poblaciones experimentales, la cual es consistente con el hecho de que con esta regla de actualización no se toman en cuenta ni los pagos ni las acciones del resto de jugadores.

[8] En una red libre de escala, algunos nodos están altamente conectados, es decir, poseen un gran número de enlaces a otros nodos, aunque el grado de conexión de casi todos los nodos es bastante bajo.

[9] El grado de un vértice en una red es el número de conexiones asociadas a un vértice, si se hiciera un recuento en una red del número de nodos por cada grado se tendría una **distribución de grado**.

[10] El pago al cual aspira cada jugador, el cual generalmente no es tan alto como la tentación T ni tan bajo como el pago S del cooperador que enfrenta a un desertor.

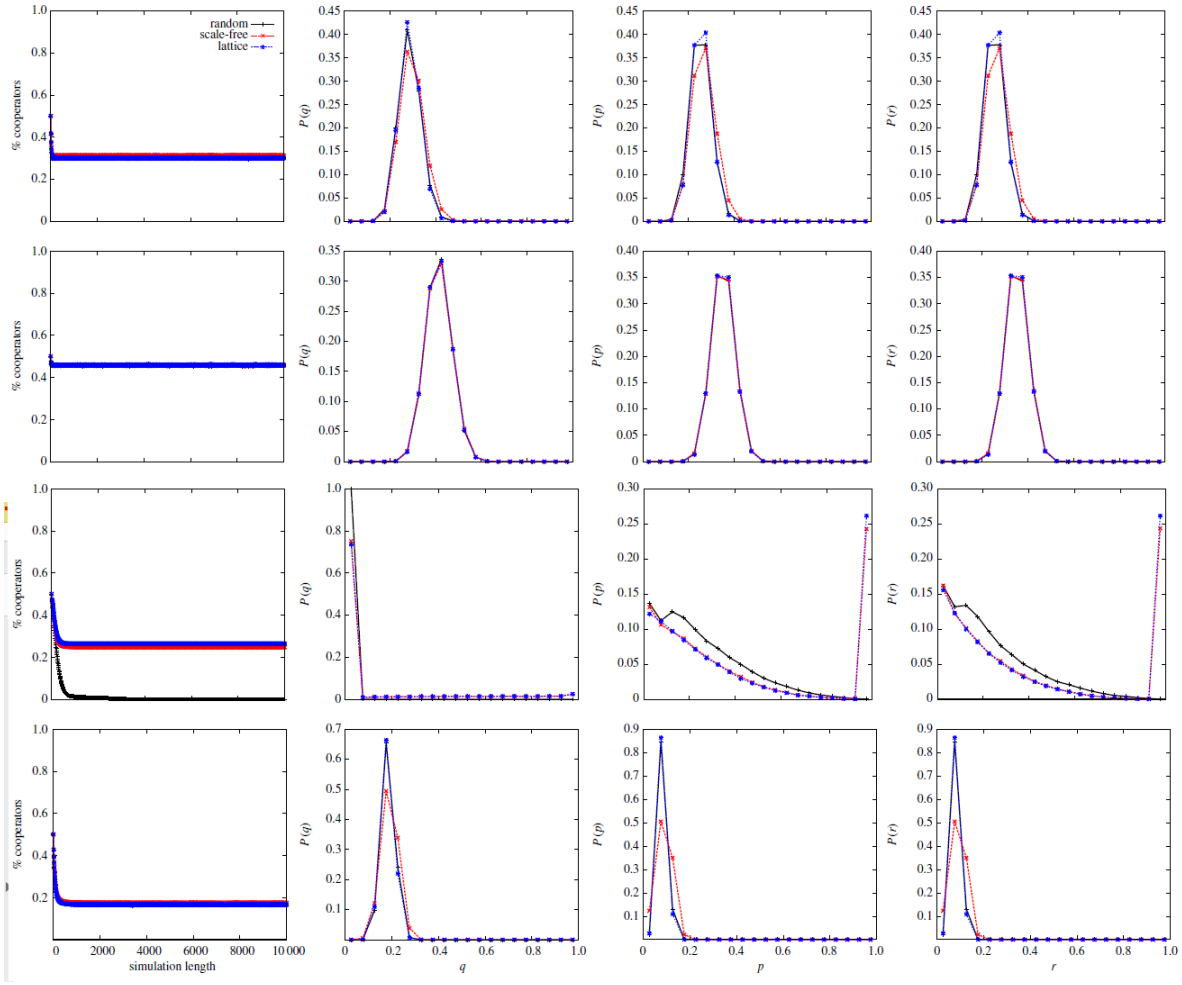

Figura 2.5: Evolución del nivel de cooperación (columna izquierda) y distribuciones estacionarias de los parámetros de MCC (de izquierda a derecha: q, p, r) cuando la dinámicas evolutiva es el aprendizaje por refuerzo, con una taza de aprendizaje de $\lambda = 10^{-1}$. De arriba a abajo: $A = \frac{1}{2}$, $A = \frac{5}{4}$, $A = -\frac{1}{4}$ y A adaptativa con $h = 0{,}2$ y $A^0 = \frac{1}{2}$. Los resultados son promediados sobre 100 realizaciones independientes, de acuerdo a la notación y métodos de Cimini y Sánchez [8].

La investigación mostró que el aprendizaje por refuerzo con un amplio rango de tasas de aprendizaje es el único mecanismo que permite explicar la estabilidad evolutiva del MCC, llevando a situaciones que son acordes con las observaciones experimentales en términos del nivel estacionario de la cooperación lograda, valores promedio y distribuciones estacionarias los parámetros del MCC y la ausencia de la reciprocidad en red. Queda clara así la relevancia los conceptos teóricos derivados del aprendizaje por refuerzo para el estudio de juegos sobre redes.

Una validación final importante del aprendizaje por refuerzo viene de estudiar el modelo de actualización de aprendizaje basado en la atracción causada y ponderada por la experiencia, conocido en inglés como *Experience Weighted Attraction* (EWA), el cual representa una dinámica evolutiva que combina aspectos de los modelos de aprendizaje de creencias (a los que *Mejor Respuesta* pertenece) y del aprendizaje por refuerzo. Con los resultados obtenidos para esta elección de esquema de actualización, Cimini y Sánchez afirman que el aprendizaje por refuerzo es la contribución determinante que lleva a situaciones que coinciden con los resultados empíricos.

2.3.2. El Aprendizaje de Atracción Causada y Ponderada por la Experiencia (EWA): ¿Cómo puede explicar el comportamiento humano?

Aprendizaje EWA en juegos en forma normal

Teniendo en cuenta dos enfoques distintos de aprendizaje en juegos y considerando importantes datos experimentales, los investigadores en ciencias económicas y toma de decisiones Colin Camerer y Teck-Hua Ho propusieron un modelo general de aprendizaje con el fin de responder a la pregunta "¿Qué modelos describen mejor el comportamiento humano?". El nombre de este modelo de aprendizaje puede traducirse del inglés como Atracción Causada y Ponderada por la Experiencia, ya que originalmente es llamado *Experience-weighted Attraction*, pero lo denotaremos aquí solamente como EWA. La palabra "ponderada" (weighted) en este sentido hace referencia al peso de importancia y atracción que un jugador da a una determinada acción o estrategia de juego dependiendo de la experiencia alcanzada. Veamos cómo lo describieron Camerer y Ho [5].

El modelo EWA combina elementos de dos enfoques aparentemente diferentes, y los incluye como casos especiales del modelo general. Uno de los enfoques es el de los modelos basados en creencias, los cuales inician con la premisa de que los jugadores hacen un seguimiento de la historia de las jugadas previas de los otros jugadores y forma alguna creencia acerca de lo que los otros harán en el futuro basándose en observaciones pasadas. Entonces ellos tienden a elegir la *mejor respuesta*, una estrategia que maximiza sus pagos esperados según las creencias que crearon.

Por otra parte, el enfoque del refuerzo de elección, asume que las estrategias son reforzadas por sus pagos previos, y serán más o menos propensos a elegir una determinada estrategia dependiendo de cierta manera de su nivel de refuerzo. Los jugadores que aprenden por refuerzo generalmente no tienen creencias acerca de lo que los otros jugadores harán. Ellos solamente tienen en cuenta los pagos que las estrategias les arrojaron en el pasado, pero no la historia de las jugadas que originaron esos pagos.

La notación

En el aprendizaje EWA las estrategias tienen atractivos que reflejan las predisposiciones iniciales, se actualizan basadas en la experiencia de pago y se determinan las probabilidades de elección de acuerdo a alguna norma específica. La notación usada por Camerer y Ho es la siguiente [5]:

Se estudian juegos formales de n personas. Los jugadores están indizados por i $(i = 1, \ldots, n)$ y el

espacio de estrategias del jugador i, S_i, consiste de m_i opciones discretas, esto es

$$S_i = \{s_i^1, s_i^2, \dots s_i^j, \dots, s_i^{m_i-1}, s_i^{m_i}\}.$$

El conjunto

$$S = S_1 \times \cdots \times S_n$$

es el producto cartesiano de los espacios de estrategias de los individuos y es el espacio de estrategias del del juego. $s_i \in S_i$ denota una estrategia del jugador i, y es, por lo tanto, un elemento de S_i.

El elemento de S

$$s = (s_1, \dots, s_n)$$

es una combinación de estrategias, y consiste de n estrategias, una por cada jugador.

$$s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$$

es una combinación de estrategias de todos los jugadores a excepción del jugador i. S_i tiene cardinalidad

$$m_{-i} = \prod_{k=1, k\neq i}^{n} m_k.$$

La función de pagos de valor escalar para el jugador i es $\pi_i(s_i, s_{-i})$. Se denota la estrategia actual escogida por el jugador i en la ronda t por $s_i(t)$, y la estrategia (vector) escogida por los demás jugadores se denota por $s_{-i}(t)$. El pago para el jugador i en la ronda t por $\pi_i(s_i(t), s_{-i}(t))$.

La esencia del modelo EWA son dos variables, que son actualizadas después de cada ronda. La primera variables es $N(t)$, la cual interpretamos como el número de equivalencias de observación de experiencias pasadas. La segunda variable es $A_i^j(t)$, la atracción del jugador i hacia la estrategia s_i^j después de que se ha jugado la ronda t.

Las variables $N(t)$ y $A_i^j(t)$ comienzan con algunos valores previos, $N(0)$ y $A_i^j(0)$. Esos valores pueden ser entendidos como resultado de una *experiencia pre-juego*. Las actualizaciones son regidas por dos reglas. Primero,

$$N(t) = \rho \cdot N(t-1) + 1, \quad t \geq 1. \tag{2.3}$$

El parámetro ρ es una tasa de depreciación o factor de descuento retrospectivo que mide el impacto fraccionario de la experiencia previa, en comparación con una nueva ronda.

La segunda regla de actualiza el nivel de atracción. Un componente clave de la actualización es el pago que una estrategia produjo o pudo haber producido en una determinada ronda. A través de un parámetro δ el modelo pesa los pagos hipotéticos que las estrategias que no fueron escogidas habrían arrojado, y pesa los pagos recibidos en realidad a partir de una estrategia $s_i(t)$ agregando un $1-\delta$ (así, reciben un total de 1). Usando una función indicadora $I(x, y)$, que es igual a 1 si $x = y$ y 0 si $x \neq y$, el pago ponderado puede ser escrito como $[\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi(s_i^j, s_{-i}(t))$.

La regla para las actualizaciones de la atracción toman $A_i^j(t)$ como la suma de una atracción previa de experiencia, ponderada y depreciada $A_i^j(t-1)$ más el pago (ponderado) de la ronda t, normalizado por el peso de experiencia actualizado:

$$A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi(s_i^j, s_{-i}(t))}{N(t)}. \tag{2.4}$$

El factor ϕ es un factor de descuento o tasa de decaimiento, que deprecia la atracción previa.

Una característica clave es el parámetro δ que pesa la fortaleza del refuerzo hipotético de las estrategias que no fueron elegidas de acuerdo con el pago que hubieran obtenido, en relación con el refuerzo de las estrategias elegidas de acuerdo con los pagos recibidos. Las otras características clave son dos tasas de descuento, ϕ y ρ, que respectivamente descuentan atracciones anteriores y peso de experiencia. El aprendizaje EWA incluye aprendizaje por refuerzo y jugadas ponderadas fabricadas (aprendizaje por creencias) como casos especiales, e hibrida sus elementos claves. Cuando $\delta = 0$ y $\rho = 0$, resulta el refuerzo acumulado de elección. Cuando $\delta = 1$ y $\rho = \phi$, los niveles de refuerzo de estrategias son exactamente los mismos que los de los pagos esperados, dadas ciertas creencias de jugadas ponderadas y fabricadas.

Las probabilidades

Las atracciones deben determinar de alguna manera las probabilidades de elegir las estrategias. Para la probabilidad $P_i^j(t)$ que tiene el jugador i de escoger la estrategia s_i^j en la ronda t, existen tres formas usadas hasta ahora: exponencial (logit), por potencia y normal (probit).

La forma exponencial está dada por

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}. \tag{2.5}$$

El parámetro λ mide la sensibilidad de los jugadores hacia las atracciones. Esta sensibilidad podría variar debido a estados psicofísicos de percepción o a los niveles de motivación de los sujetos. En esta función de probabilidad, al querer calcular la probabilidad de elección de la estrategia s_i^j, el exponente $\lambda \cdot A_i^j(t)$ en el numerador es solo el efecto de peso de atracción hacia la estrategia s_i^j.

La forma de probabilidad por potencia está dada por

$$P_i^j(t+1) = \frac{(A_i^j(t))^\lambda}{\sum_{k=1}^{m_i} (A_i^k(t))^\lambda}. \tag{2.6}$$

Fundamentalmente, es un problema abierto saber cuándo las formas logit, probit o de potencia funcionan mejor. Estudios anteriores han determinado que la forma logit muestra mejores ajustes que la forma por potencia. Por esto se hizo uso de la probabilidad exponencial para obtener los datos discutidos a continuación.

El experimento Camerer-Ho basado en EWA

Uno de los resultados importantes, que vale la pena mencionar aquí, obtenidos por Camerer y Ho se logró haciendo una serie de comparaciones de predicciones de los modelos de aprendizaje EWA, de refuerzo y de creencias, sobre el comportamiento de individuos en un juego de *acción media*, cuyas caracterísricas detallamos a continuación.

En los juegos coordinativos de acción mediana estadísticamente ordenados, el pago grupal depende de la mediana de todas las acciones de los jugadores. La matriz de pagos usada es la representada en la tabla de la figura 2.6. Los pagos incrementan en la mediana y disminuyen en la desviación de la mediana. Los juegos de acción mediana capturan aquellas situaciones en las que las presiones de conformidad inducen a las personas a comportarse como los demás, pero cada uno prefiere que el grupo elija una mediana alta.

		Mediana $\{X_i\}$						
		7	6	5	4	3	2	1
	7	1.30	1.15	0.90	0.55	0.10	-0.45	-1.10
	6	1.25	1.20	1.05	0.80	0.45	0.00	-0.55
	5	1.10	1.15	1.10	0.95	0.70	0.35	-0.10
X_i	4	0.85	1.00	1.05	1.00	0.85	0.60	0.25
	3	0.50	0.75	0.90	0.95	0.90	0.75	0.50
	2	0.05	0.40	0.65	0.80	0.85	0.80	0.65
	1	-0.50	-0.05	0.30	0.55	0.70	0.75	0.70

Figura 2.6: Matriz de pagos en el juego de acción mediana [5].

Se estiman los modelos EWA, de refuerzo de elecciones y de creencias usando 6 sesiones, cada una con 9 individuos que juegan 10 periodos juntos (la muestra es de 54 individuos). Las 6 sesiones se observan mancomunadamente para obtener frecuencias generales por ronda de las decisiones tomadas por los grupos de 9 individuos. En cada ronda cada jugador escoge un entero del 1 al 7. Al final de cada ronda se anuncia la mediana de los enteros escogidos y los jugadores calculan sus pagos. Como los grupos son grandes y los jugadores no saben cómo cambiaría la mediana si su decisión hubiera sido diferente, se asume que los jugadores forman creencias sobre la mediana de todos los jugadores, ignorando su propia influencia en la mediana y considerando al grupo como la composición de un solo jugador.

La figura 2.7 muestra las frecuencias a lo largo de las seis sesiones, integradas como una sola. Las elecciones iniciales están concentradas alrededor de 4 y 5, con una depresión en 6 y picos bajos en 3 y 7. Las elecciones posteriores se mueven bruscamente hacia las medianas iniciales, que fueron siempre 4 ó 5. Una característica sorprendente, que está oculta por la integración de las sesiones, es que la mediana de la décima ronda en cada sesión era igual a la mediana de la primera ronda.

En tres sesiones la mediana empezó siendo 4 y se mantuvo allí; en las otras tres sesiones la mediana empezó siendo 5 y se mantuvo allí. La figura 2.7 muestra tres características claves de la información que cualquier modelo de aprendizaje debería incluir: Los picos iniciales en 4 y 5 aproximadamente doblan su tamaño (mientras los jugadores converjan completamente hacia ellas); las elecciones de desequilibrio 3 y 7 son rápidamente extinguidas después de la primera ronda, y hay una depresión en las elecciones iniciales del el 6 (menos jugadores eligieron 6, en contraste con las estrategias próximas 5 y 7).

Desde el punto de vista del aprendizaje, los juegos de acción mediana son interesantes porque la pena por la desviación es bastante pequeña si los jugadores están cerca a un equilibrio. Sin embargo, la convergencia aguda se produce dentro de un par de períodos. Los modelos de aprendizaje que asumen que las opciones se refuerzan deben explicar por qué los jugadores se mueven rápidamente al equilibrio a pesar del gran refuerzo que tienen por estar cerca del equilibrio y a pesar de la pequeña ganancia extra por moverse al equilibrio. El modelo EWA puede explicar esta rápida convergencia si δ es cercano a uno, incorporando la estrategia de *mejor respuesta* inherente en el aprendizaje por creencia.
Aplicando las predicciones teóricas de los diferentes modelos, se obtuvieron los respectivos errores según los datos reales en la figura 2.7. Sobre esto, se observó que:

- Las atracciones iniciales estimadas de EWA reflejan básicamente el patrón de las frecuencias originales. El modelo EWA tuvo sus errores más notables en la predicción de la frecuencia real de 3 (0.6 por debajo), 6 (0.03 por encima) y 7 (0.01 por encima). Estas predicciones se observan en la figura 2.8.
- El modelo de refuerzo predijo muy por debajo las frecuencias de 3 y 7, aproximadamente 0.08 por debajo. Los jugadores que escogieron 7 en la primera ronda, cambiaron rápidamente a números

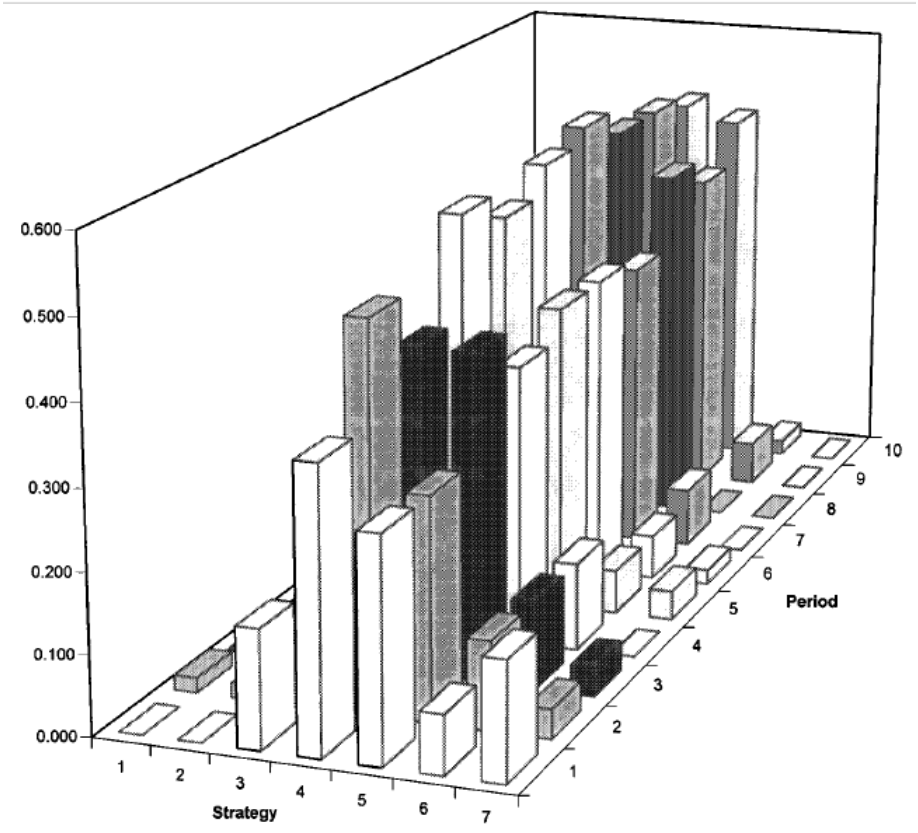

Figura 2.7: Frecuencias reales en el juego de acción mediana en el experimento Camerer-Ho [5].

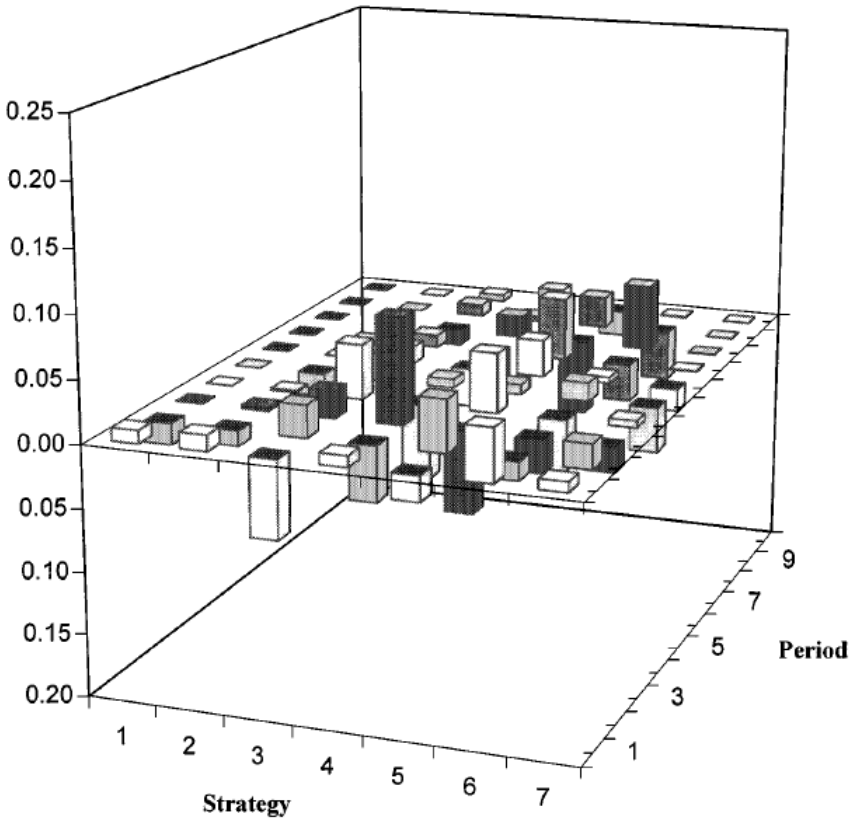

Figura 2.8: Errores de predicción del modelo EWA en el experimento Camerer-Ho [5].

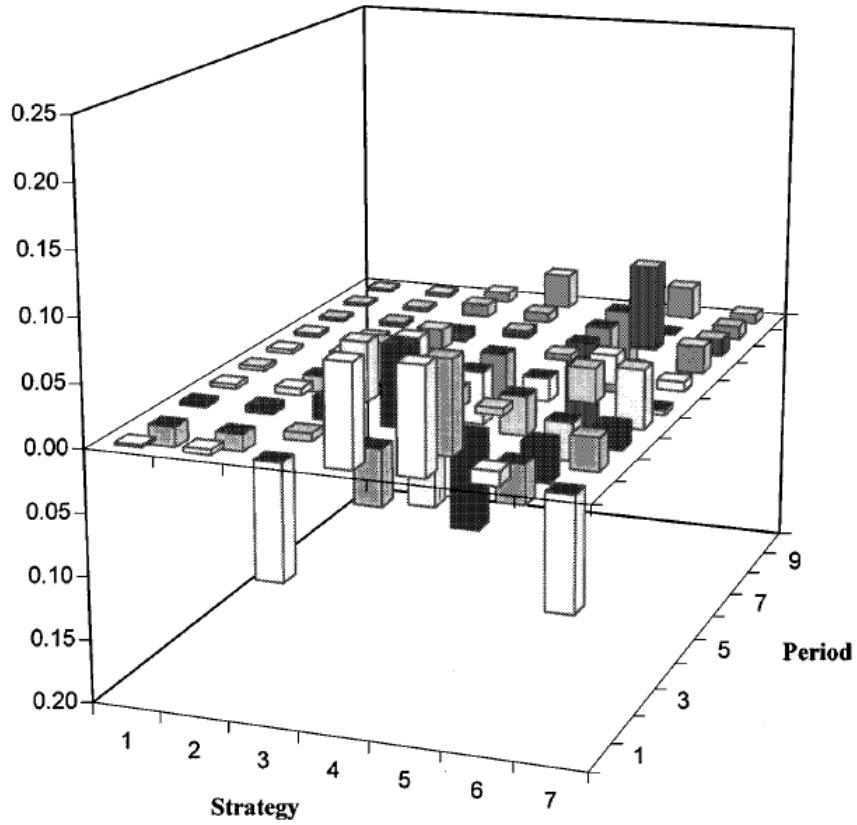

Figura 2.9: Errores en la predicción basada en el modelo de refuerzo de elecciones en el experimento Camerer-Ho [5].

más bajos en la segunda ronda. El aprendizaje por refuerzo no puede predecir qué tan rápido ocurre esta convergencia (ver figura 2.9).

- El modelo de creencias predijo muy por debajo las estrategias 3 y 7, pero por una razón diferente: En este modelo es difícil explicar por qué los jugadores escogerían la estrategia 6 menos que la 5 o la 7. El problema es que las creencias iniciales que dan una espectativa alta de pagos para 4 y 5 también dan una espectativa de pagos para 6 que es aproximadamente igual de alta, y más alta que la espectativa de pago en la estrategia 7. Así, es difícil encontrar un conjunto de creencias particulares que puedan explicar los picos en 4, 5 y 7 sin predecir también un pico en 6 (ver figura 2.10).

En resumen, las estimaciones de parámetros del modelo se graduaron con una parte de los datos y se usaron para predecir una muestra. Las estimaciones de δ son generalmente alrededor de 0,5, ϕ alrededor de 0,8 y 1; mientras que ρ varía de 0 a ϕ. Se encontró que los casos especiales de refuerzo y de aprendizaje por creencias generalmente son rechazados a favor de EWA, aunque los modelos por creencias mejoran en algunos juegos de suma constante.

También se concluye que el aprendizaje EWA es capaz de combinar las mejores características de los enfoques anteriores, permitiendo que las atracciones comiencen y crezcan con flexibilidad como lo hace el refuerzo de elecciones, pero reforzando estrategias no escogidas sustancialmente como los modelos basados en creencias lo hacen.

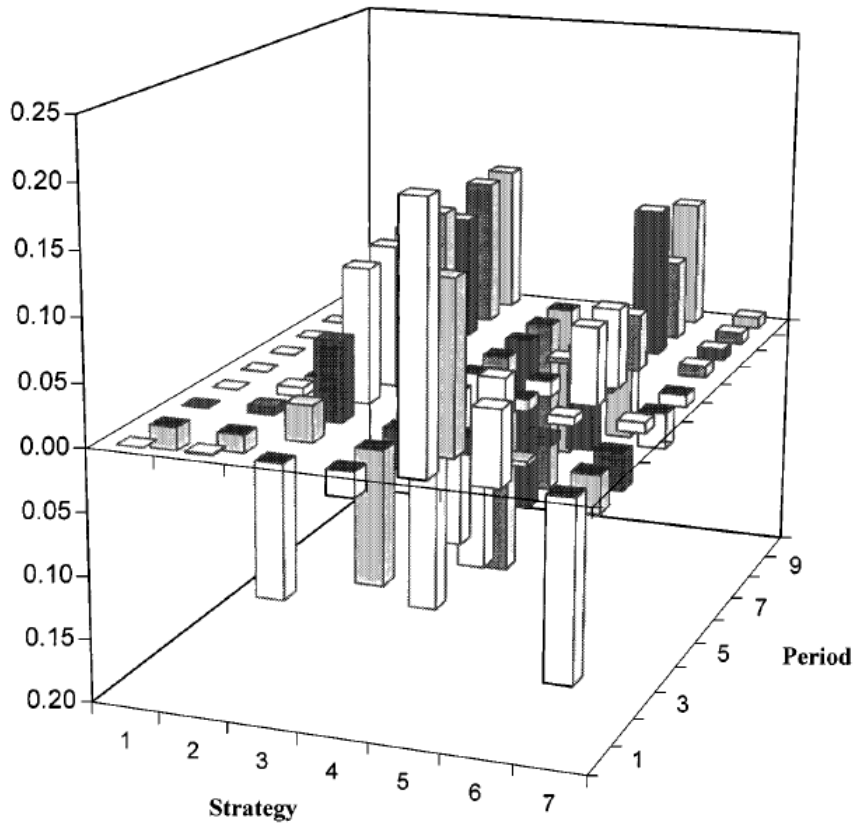

Figura 2.10: Errores de la predicción basada en el modelo de creencias en el experimento Camerer-Ho [5].

Camerer y Ho mencionan entre sus análisis finales [5] que el modelo EWA también tendrá que ser actualizado para hacer frente a tres desafíos de modelado para poder ser aplicado con más generalidad: Sofisticación, información imperfecta de pago y especificación de estrategias. Afirman que ajustar EWA a esas extensiones especiales podría permitir crear una manera unificada de predecir cómo las personas juegan en diseños de laboratorio y, eventualmente, cómo juegan en situaciones reales no prediseñadas.

Algunas extensiones del modelo EWA

incorporando *Sofisticación.* Para la extensión de EWA obtenida con la incorporación del análisis de situaciones en las que los jugadores solo disponen de información acerca de las elecciones pasadas de sus oponentes, ignorando información sobre los pagos de los otros (aprendizaje sofisticado) , los autores Camerer y Ho junto con el investigador en ciencias económicas Juin-Kuan Chong desarrolaron una investigación en la que proponen un modelo general que involucra la consideración de la sofisticación en las dinámicas del aprendizaje de los jugadores citecamererhochong2002.

En dicha investigación, los autores afirman que la mayoría de los modelos de aprendizaje asumen que los jugadores son adaptativos (es decir, responden sólo a su propia experiencia previa e ignoran la información de pagos de los demás) y que el comportamiento no es sensible a la forma en que los jugadores compiten. Muestran que la evidencia empírica sugiere lo contrario.

En el artículo [6], amplían el modelo de aprendizaje adaptativo EWA [5] para involucrar aprendizaje sofisticado y enseñanza estratégica en juegos repetidos. El modelo generalizado asume que hay una mezcla entre aprendices adaptativos y jugadores sofisticados. En el modelo generalizado se llega a que, al igual que antes, un aprendiz adaptativo ajusta su comportamiento de la manera EWA. Sin embargo, un jugador sofisticado no aprende y racionalmente responde mejor a sus pronósticos de todos los demás comportamientos. Un jugador sofisticado podría llamarse *previsor*. Un jugador previsor desarrolla pronósticos de los comportamientos de otros sobre rondas múltiples en vez de una sola ronda y elige "enseñar" a los demás jugadores al elegir un escenario de estrategia que le da el valor actual neto actualizado más elevado.

Estimaron el modelo usando datos de juegos equivalentes a la dinámica del juego del concurso de belleza[11]. En general, los resultados muestran que el modelo generalizado es mejor que el modelo EWA adaptativo para describir y predecir el comportamiento. La inclusión de la enseñanza también permite un enfoque empírico basado en el aprendizaje para la formación de la reputación que es al menos tan plausible como el enfoque basado en el tipo ahora estándar y es superior en el desempeño predictivo en diferentes formas.

EWA en juegos *complicados.* Basados también en la propuesta Camerer-Ho [5], se llevaron a cabo estudios recientes con consideraciones que iban más allá de los juegos con pocos jugadores y estrategias, bajo la autoría de Los investigadores Galla y Farmer [12]. Los autores afirman que tradicionalmente la teoría del juegos estudia el equilibrio de los juegos simples. Sin embargo, ¿esto es útil si el juego es complicado? y si no, ¿qué sería útil?.

Se llama juego complicado a aquel que consta de muchos movimientos posibles, y por lo tanto, muchos pagos posibles condicionados a esos movimientos. Se hicieron investigaciones sobre juegos de dos personas, donde los jugadores aprendieron basados en el modelo EWA [5]. Al generar juegos al azar, se caracterizaron las dinámicas de aprendizaje bajo EWA y se mostró que hay tres regímenes claramente separados:

[11] EL juego del concurso de belleza consiste en que cada jugador debe apuntar un número del 0 al 100, se calcula la media de los números que los jugadores han apuntado y gana el jugador que más se aproxime a 3/4 de la media.

1. Convergencia a un único punto fijo.
2. Una enorme multiplicidad de puntos fijos estacionario.
3. Comportamiento caótico.

En el tercer caso, la dimensión de los atractores caóticos puede ser muy alta, lo que implica que las dinámicas de aprendizaje son efectivamente aleatorias. En el régimen caótico, las ganancias totales fluctúan intermitentemente, mostrando ráfagas de cambio rápido pontuadas por períodos de inactividad, con pesadas colas similares a lo que se observa en la turbulencia fluida y los mercados financieros.

Los resultados sugieren que, al menos para algunos algoritmos de aprendizaje, existe un gran régimen de parámetros para el cual las interacciones estratégicas complicadas generan un comportamiento inherentemente impredecible que se describe mejor en el lenguaje de la teoría de sistemas dinámicos.

2.3.3. EWA en la cooperación: ¿Es compatible con la estrategia MCC?

Después de saber la aplicabilidad del aprendizaje EWA, las extensiones de su concepto recientemente propuestas y sus buenos ajustes a resultados experimentales, es importante verificar su compatibilidad con la cooperación, más aún, con la regla MCC [15], cuya relación con el aprendizaje vimos inicialmente en las investigaciones de Cimini y Sanchez [8], pero más adelante veremos los últimos resultados al respecto.

Respuestas halladas en un primer aporte a la relación del modelo EWA y la cooperación

Además de la propuesta de analizar resultados experimentales para dos estrategias no imitativas, a saber, *aprendizaje por refuerzo* (RL) y *mejor respuesta* (BR) (perteneciente al aprendizaje por creencias), dentro de la propuesta Cimini-Sanchez también se incluye un estudio de la funcionalidad del modelo EWA en la predicción de probabilidades teniendo en cuenta la regla MCC [8].

Se concluyó que a pesar de que la formulación original de EWA no puede ser generalizada trivialmente en el contexto MCC (debido a los múltiples parámetros y a que las acciones de los jugadores vecinos regulan la estrategia de cada individuo), es posible reproducir las características claves de la dinámica EWA con una simple combinación lineal de los aprendizajes por refuerzo y por mejor respuesta. En los resultados confirman que dicha formulación actualiza las estrategias exactamente como el modelo EWA original. No es necesaria la estipulación de atractores iniciales, ni la suposición de ninguna experiencia particular puesto que el objetivo no es hacer una reproducción cuantitativa de los resultados experimentales.

Se procede entonces de la siguiente manera: Para cada jugador i, en cada actualización t, se calcula el cambio en los parámetros MCC dados por las dinámicas del aprendizaje por refuerzo y mejor respuesta. Tales parámetros se denotan por $\{\delta q^t_{i(BR)}, \delta p^t_{i(BR)}, \delta r^t_{i(BR)}\}$ y $\{\delta q^t_{i(RL)}, \delta p^t_{i(RL)}, \delta r^t_{i(RL)}\}$, respectivamente. Entonces los parámetros son actualizados por

$$\begin{aligned} q_i^{t+1} &= q_i^t + \gamma\delta q^t_{i(RL)} + (1-\gamma)\delta q^t_{i(BR)} \\ p_i^{t+1} &= p_i^t + \gamma\delta p^t_{i(RL)} + (1-\gamma)\delta p^t_{i(BR)} \\ r_i^{t+1} &= r_i^t + \gamma\delta r^t_{i(RL)} + (1-\gamma)\delta r^t_{i(BR)} \end{aligned}$$

donde $\gamma \in (0,1)$ es el parámetro de mezcla de las dos estrategias en combinación lineal. Nótese que $\delta_{(BR)} = \delta$ (cantidad por la cual cambian los parámetros en cada actualización) y que $\delta \sim \lambda$ (parámetro

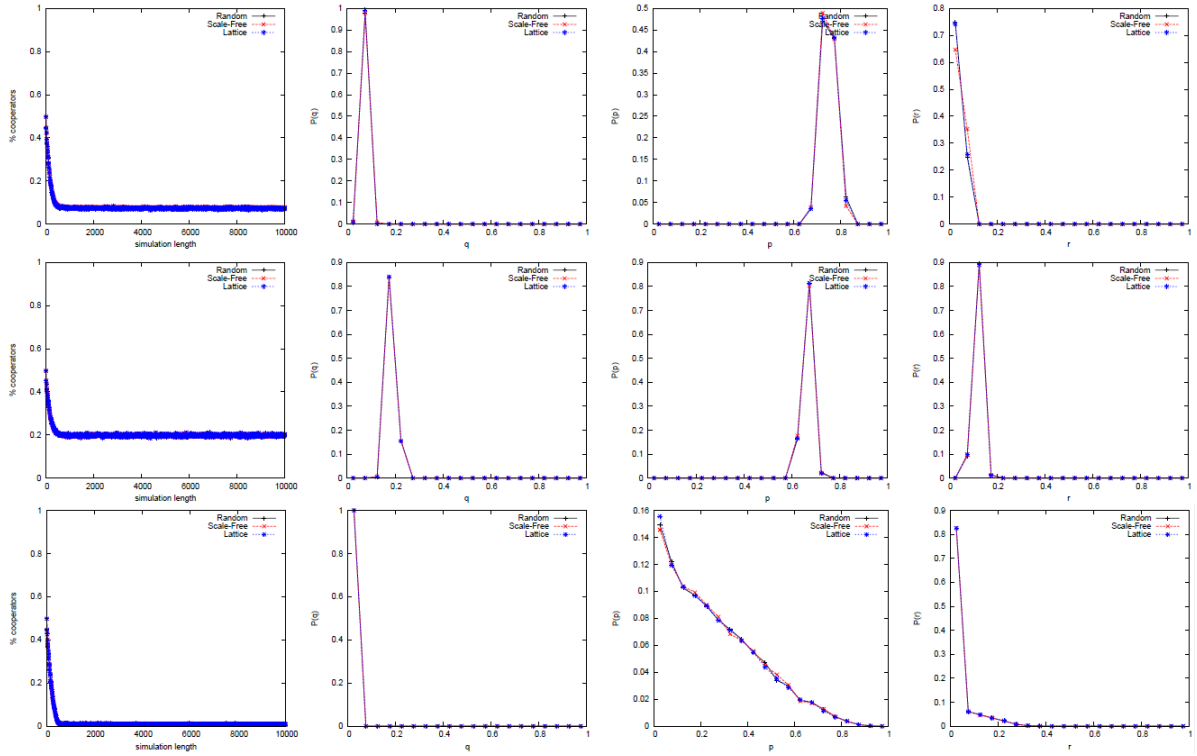

Figura 2.11: Evolución del nivel de cooperación (columna izquierda) y distribuciones estacionarias de los parámetros MCC (de derecha a izquierda: q, p y r) cuando la dinámica evolutiva es EWA, con $\delta = \lambda = 10^{-2}$ y $\gamma = \frac{3}{4}$. De arriba a abajo: $A = \frac{1}{2}$, $A = \frac{5}{4}$ y la aspiración adaptativa A, con $h = 0{,}2$, $A^0 = \frac{1}{2}$, de acuerdo a la notación y métodos de Cimini y Sánchez [8].

de intensidad de aprendizaje). Así, para que los términos sean comparables, usualmente se hace $\delta = \lambda$.

Los resultados para la actualización según EWA recopilados por la investigación, usando la notación y métodos propuestos por Cimini y Sánchez [8] (expuestos aquí en la sección 2.3.1) están representados en la figura 2.11. Se puede observar que cuando la tendencia a la deserción completa ocasionada por la estrategia Mejor Respuesta, no es dominante ($\gamma > \frac{1}{2}$), el nivel de cooperación se encuentra entre los niveles obtenidos separadamente por la estrategia Mejor Respuesta y el aprendizaje por refuerzo. Existen distribuciones estacionarias de parámetros de MCC, pero son más estrechas y, excepto para el parámetro p, se concentran en valores más pequeños que con el aprendizaje por refuerzo solo.

Respuestas halladas en el aporte más reciente al modelo EWA

Si bien los resultados de Cimini y Sánchez [8] no dan la claridad ni la precisión matemática requerida para la descripción de la cooperación mediante el modelo EWA, veremos que se ha demostrado recientemente que el modelo EWA, con sus descripciones originales, puede explicar directamente las dinámicas de cooperación en dilemas sociales.

Se trata de la investigación más reciente en lo que concierne a la funcionalidad del modelo EWA, llevada a cabo por los investigadores John Realpe, Javier Montoya (ambos vinculados a la Universidad de Cartagena), Giulia Andrighetto y Luis Gustavo Nardin [27]. En este artículo, la propuesta inicial

es estudiar la influencia de la cooperación en la resiliencia[12], esto es, el impacto que la cooperación ejerce sobre la capacidad de los seres humanos para asimilar cambios y aún así continuar.

Para llevar a cabo lo propuesto, los autores desarrollaron un modelo analíticamente manejable, lo cual llevaría también a explicar cómo la cooperación eleva los niveles de resiliencia en grupos de humanos. Se señala que las decisiones individuales acerca de cooperar o no, están basadas en el balance entre las motivaciones egoístas y las motivaciones prosociales[13], y esta característica es capaz de reproducir cuantitativamente los hallazgos de experimentos recientes a gran escala con seres humanos [15], incluso con mayor exactitud que modelos anteriores [8].

¿Cómo entra en acción el aprendizaje EWA en este contexto? Debido a que el comportamiento estratégico en humanos está basado en algoritmos de aprendizaje por refuerzo [5, 12], que están relacionados con los hábitos adquiridos a partir de experiencias pasadas y con los logros que los individuos esperan alcanzar en un futuro. Estas descripciones pueden ser capturadas por una forma simplificada de las ecuaciones del aprendizaje EWA [5].

$$x_i(t+1) = \frac{1}{1+e^{-\beta D_i(t+1)}}, \tag{2.7}$$

$$D_i(t+1) = (1-\alpha)D_i(t) + \Delta U_i(t), \tag{2.8}$$

donde $x_i(t+1)$ y $D_i(t+1)$ son, respectivamente, la probabilidad y la motivación (o *impulso*) de un individuo i para cooperar en una ronda $t+1$. Cuando el parámetro β es grande, el individuo tiende a cooperar si su *impulso* es positivo, y a desertar si su *impulso* es negativo. Cuando el impulso o β son nulos, se actúa aleatoriamente. El término $\Delta U_i(t)$ en la ecuación 2.8 es la diferencia de utilidades que resulta de las opciones de cooperar o desertar. Si $\Delta U(t)_i > 0$, el impulso o motivación del individuo i por cooperar crece; si $\Delta U_i(t) < 0$ su impulso decrece, mientras que si $\Delta U_i(t) = 0$ el impulso es constante. Finalmente, el parámetro α describe la pérdida de memoria: si $\alpha = 1$, el agente recuerda solo la ronda previa t, mientras que si $\alpha = 0$, se acumula la información de la historia total de juego.

El término $\Delta U_i(t)$ depende de una *componente individual* $\Delta I_i(t)$, que hace referencia a la motivación del individuo i para cooperar en la ronda t; y también depende de una *componente social* $\Delta N_i(t)$, que se refiere a la motivación del individuo por cumplir una determinada normatividad social. A través de esta componente social ya se introduce la regla MCC, por lo cual no es necesario recurrir a los parámetros de esta regla para llegar a explicar la cooperación.

La relevancia del aporte de esta reciente investigación a la funcionalidad del modelo EWA es que se muestra que este modelo puede ser utilizado directamente para explicar la cooperación en humanos, sin tener que utilizar este mecanismo de aprendizaje solo para recurrir primero a ciertas descripciones de los parámetros de la regla MCC y luego poder explicar la cooperación, como se ve en los desarrollos de Cimini y Sanchez [8].

2.4. Conclusiones

La cooperación es una de las principales fundamentos de los dilemas sociales actuales, de la teoría de juegos y de recientes estudios de redes de interacción social formadas por humanos. El estudio de las

[12] La resiliencia se define frecuentemente como la capacidad de los seres humanos para adaptarse positivamente a situaciones adversas.

[13] Se entiende por conducta prosocial toda conducta social positiva, no dañina para la colectividad, con o sin motivación altruísta. A su vez, se entiende por motivación altruísta el deseo de favorecer al otro con independencia del propio beneficio.

diferentes dinámicas y evoluciones de la cooperación a través del Dilema del Prisionero, conduce a un mundo abierto de interrelación de modelos matemáticos, físicos, económicos y sociales.

El aprendizaje juega un papel importante para la cooperación debido a los fuertes vínculos entre los cambios en las estrategias cooperativas y los procesos de aprendizaje que experimentan los jugadores [21, 19]. Se pudo ver que estudios y análisis para comprobar los efectos del aprendizaje han demostrado que, por ejemplo, el comportamiento MCC surge del proceso de aprendizaje experimentado por los jugadores [8, 16].

El análisis de los estudios observados en [26, 14, 17] muestran que los humanos frecuentemente actúan de forma más cooperativa de lo que dictaría el simple interés personal. Una posible razón de ello, es la forma iterada del dilema del prisionero. Se pudo ver por medio de resultados experimentales que esta variación del Dilema del Prisionero hace que la "no cooperación" se castigue de una manera más severa y que el caso contrario, es decir, la intención o acción de cooperar, se premie más de lo que podría sugerir el problema original.

Se pudo conocer aquí también que, según datos experimentales basados en teoría de juegos evolutivos, [11, 14, 13], los humanos presentan características especiales en la evolución y las dinámicas con las que actualizan sus estrategias a lo largo de diferentes iteraciones del Dilema del Prisionero; sugieriendo que, por ejemplo, la cooperación de cada individuo depende más de la cooperación que observa en el resto de juegadores que de los altos pagos que se ofrezcan, siendo más propensos a contrubuir en la medida que los jugadores vecinos lo hagan.

Es clara la aplicabilidad e importancia del comportamiento MCC a los dilemas sociales cooperativos, evidenciada a través de experimentos con el Dilema del Prisionero a gran escala [14, 13] y experimentos con el Dilema del Prisionero en modo multijugador [16].

Además de los experimentos con el Dilema del Prisionero, cabe resaltar otros resultados importantes basados en experimentos con el Juego del Bien Público [3, 10]. Con esto se ve además, que la tasa de disminución de la contribución y la sostenibilidad de la cooperación pueden estar relacionadas con factores y parámetros determinantes en la dinámica que describa el aprendizaje de los jugadores. Además, se pudo inferir que la amenaza del castigo eleva mucho la cooperación media de los jugadores.

Respecto a la relación de las dinámicas de aprendizaje con la cooperación, se ha evidenciado que el aprendizaje por refuerzo resultó siendo un mecanismo altamente compatible con la cooperación condicional, permitiendo la manifestación de la regla MCC evolutivamente estable explicando claramente los niveles de cooperación en experimentos sociales [8].

Llegando a nuestro punto central de interés, se encontró por medio de los resultados de Cimini y Sanchez [8], que el mecanismo propuesto allí, el cual utiliza parámetros que son obtenidos mediante el método de aprendizaje EWA, conduce a situaciones que son compatibles con resultados experimentales, pero bajo la condición de que la contribución del aprendizaje por refuerzo sea dominante. En estos resultados se señala también que hacer una adaptación precisa de la regla MCC bajo la parametrización completa del modelo EWA no es clara ni sencilla de describir. Sin embargo, gracias al aporte más reciente al modelo EWA [27], vemos que los resultados muestran una adecuada adaptación de las características de este mecanismo de aprendizaje a los comportamientos cooperativos y, en particular, a los parámetros de la regla MCC.

Quedando clara la compatibilidad de las ecuaciones de EWA con la regla MCC introducida en una función de pago normativa, es un hecho que el aprendizaje EWA puede explicar de forma matemática y directa tanto la cooperación como la regla Moody Conditional Cooperation observada en muchos dilemas sociales de la actualidad.

Bibliografía

[1] Abramson, G. *Introducción a la Teoría de Juegos.* Centro Atómico Bariloche, Instituto Balseiro y CONICET, 2006.

[2] Axelrod, R. *The evolution of cooperation.* Basic Books, 1984.

[3] Brañas, P., and Paz, M. Unraveling in public good games. *Games, volumen 2* (2011).

[4] Bradanovic, T. Teoría de juegos para dummies, 2012.

[5] Camerer, C., and Ho, T. Experienced-weighted attraction learning in normal form games. *Econometrica, volumen 67* (1999).

[6] Camerer, C., Ho, T.-H., and Chong, J. Sophisticated ewa learning and strategic teaching in repeated games. *Journal of Economic Theory* (2002).

[7] Camerer, C. F. Behavioral studies of strategic thinking in games. *California Institute of Technology* (2003).

[8] Cimini, G., and Sánchez, A. Learning dynamics explains human behaviour in prisioner's dilemma on networks. *Royal Society Interface* (2014).

[9] Dawes, R. Social dilemmas. *Annual Review of Psychology, volumen 31, páginas 169-193* (1980).

[10] Fehr, E., and Gächter, S. Altruistic punishment in humans. *Nature, volumen 415* (2002).

[11] Fisbacher, U., Gächter, S., and Fehr, E. Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters. Volumen 71, páginas 397-404* (2001).

[12] Galla, T., and Farmer, J. D. Complex dynamics in learning complicated games. *University of California* (2012).

[13] Gracia-Lázaro, C., Ferrer, A., Ruiz, G., Cuesta, J., Sánchez, A., and Moreno, Y. Heterogeneus networks do not promote cooperation when humans play a prisoner's dilemma. *Proceedings of National Academy of Science, USA* (2012).

[14] Grujić, J. *Models of social behaviour based on Game Theory.* PhD thesis, Universidad Carlos III de Madrid. Departamento de Matemáticas, 2012.

[15] Grujić, J., Cuesta, J., and Sánchez, A. On the coexistence of cooperators, defectors and conditional cooperators in the multiplayer iterated prisoner's dilemma. *Journal of Theoretical Biology* (2012).

[16] Grujić, J., Eke, B., Cabrales, A., Cuesta, J., and Sánchez, A. Three is a crowd in iterated prisoner's dilemmas: experimental evidence on reciprocal behavior. *Scientific Reports* (2012).

[17] Grujić, J., Gracia-Lázaro, C., Milinski, M., Semmann, D., Traulsen, A., Cuesta, J., Moreno, Y., and Sánchez, A. A comparative analysis of spatial prisoner's dilemma experiments: Conditional cooperation and payoff irrelevance. *Scientific Reports, volumen 4* (2014).

[18] Holland, J. Complex adaptive systems. *Daedalus, volumen 121, No. 1, A new Era in Computation, páginas 17-30* (1992).

[19] Horita, Y., Takezawa, M., Inukai, K., Kita, T., and Masuda, N. Reinforcement learning accounts for moody conditional cooperation behavior: Experimental results. *Scientific Reports, volumen 7* (2017).

[20] Huerta, A. La teoría de juegos en la modelación de sistemas adaptativos complejos. Sexto Seminario Pensamiento Sistémico y Análisis de Sistemas, Santa Fe Institute.

[21] Izquierdo, S., Izquierdo, L., and Gotts, N. Reinforcement learning dynamics in social dilemmas. *Journal of Artificial Societies and Social Simulation, volumen 11* (2008).

[22] Jackson, M. O. A brief introduction to the basics of game theory. *Stanford University* (2011).

[23] Leyton-Brown, K., and Shoham, Y. Essentials of game theory. *Synthesis lectures on artificial intelligence and machine learning* (2008).

[24] Milgrom, P., and Roberts, J. Adaptive and sophisticated learning in normal form games. *Games and Economic Behavior* (1991).

[25] Ohtsuki, H., Hauert, C., Lieberman, E., and Nowak, M. A simple rule for the evolution of cooperation on graphs and social networks. *Nature, volumen 441* (2006).

[26] Rand, D., and Nowak, M. A. Human cooperation. *Trends in Cognitive Sciences, volumen 17* (2013).

[27] Realpe, J., Andrighetto, G., Nardin, L. N., and Montoya, J. A. Balancing selfishness and prosociality can enhance the resilience of human groups. *arXiv, 1608.01291* (2016).

[28] Ricart, J. Una introducción a la teoría de juegos. *IESE Business School-Universidad de Navarra* (1988).

[29] Traulsen, A., Semmann, D., Sommerfeld, R., Krambeck, H., and Milinski, M. Human strategy updating in evolutionary games. *Proceedings of the National Academy of Science, USA* (2010).

[30] Velasco, C. S., and Pineda, F. V. *Teoría Conductual de la Elección: Decisiones Que Se Revierten*. Universidad Nacional Autónoma de México, 2004.