Suma Psicológica, Vol. 11, No. 1:
Marzo de 2004, 29-52,
ISSN 0121-4381

SIMULACIÓN DEL DILEMA DEL PRISIONERO A PARTIR DE MODELOS CONEXIONISTAS DE APRENDIZAJE POR REFORZAMIENTO

Julián Tejada H.,
Lina María Perilla R.,
Sara Victoria Serrato V.,
Andrés Felipe Reyes G.
Grupo de Neurocomputación [1]
Fundación Universitaria Konrad Lorenz, Colombia

ABSTRACT:

The development of computers has allowed the generation of models that let simulate the behavior of the alive organisms under controlled conditions, where the manipulation of the variables can be done in a precise way. Actually, the simulation models are based on the behavior of dynamic systems: as the Neural Networks, inside them arises one that is based on operating conditioning and it is named Reinforcement Learning. In the present investigation it was simulated through this model the Prisoner's Dilemma (PD), manipulating a variable that determines a motivational level of the organisms that make them to be cooperative. They were carried out around 187.800 essays in those who the digital organisms had to confront the PD, manipulating 6 motivational levels. The results allow to identify an intrinsic characteristic of the PD and it is that, under certain conditions the organisms opted not to confess in a consistent way without this reason we can affirm that they are being cooperative or self-controlled because in the simulation we decided that the organism did not have any knowledge of the existence of the other one, neither of the effects that their actions had on the consequences that their partner received.

Key words: *prisoner's dilemma, reinforcement learning, cooperation, self-control, connectionism, simulation.*

---

1 Correspondencia: Grupo de Neurocomputación FUKL, neurocomputacion@fukl.edu

RESUMEN

El desarrollo de los computadores ha permitido la generación de modelos que permiten simular el comportamiento de los organismos vivos en condiciones controladas donde la manipulación de las variables se puede hacer de manera precisa. En la actualidad, los modelos de simulación se basan en el comportamiento de sistemas dinámicos, como las Redes Neuronales. Dentro de estos modelos se destaca uno que se basa en el condicionamiento operante y se denomina Aprendizaje por Reforzamiento. En la presente investigación se simuló a través de este modelo el Dilema del Prisionero (DP), manipulando una variable que determinaba un nivel motivacional de los organismos que los incitaba a ser cooperativos. Se realizaron alrededor de 187.800 ensayos en los que los organismos digitales tenían que enfrentarse al DP manipulando 6 niveles de motivación. Los resultados permiten identificar una característica intrínseca al DP y es que bajo ciertas condiciones los organismos optaron por no confesar de manera consistente, sin que por esto se pueda afirmar que están siendo cooperativos o autocontrolados. Lo anterior se debe a que en la simulación se decidió que los organismos no iban a tener conocimiento de la existencia del otro ni del efecto que sus acciones tenían sobre las consecuencias que su compañero recibía.

Palabras Clave: *dilema del prisionero, aprendizaje por reforzamiento, cooperación, autocontrol, conexionismo, simulación.*

Hoy en día se están desarrollando nuevos modelos acerca del procesamiento que ocurre en el cerebro; estos modelos con los que se trabaja hoy en día utilizan el computador no como una metáfora de éste, sino como una herramienta de cómputo; las metáforas ahora hacen referencia a cualquier sistema que posea muchas unidades que al interactuar provoquen un comportamiento complejo en ausencia de algún mecanismo central que lo oriente. Por ejemplo, se sabe que la interacción de las células neuronales a partir de un comportamiento individual relativamente sencillo genera comportamientos complejos tan interesantes como la memoria o el aprendizaje (Lin & Lee, 1996; Haykin, 1999). Así mismo, existen modelos que se basan en el comportamiento de las colonias de las hormigas en el que se exhiben muchas unidades que interactúan y se organizan para realizar tareas complejas como la de construir un hormiguero, en el que no se requieren planos para construirlo ni arquitectos que dirijan su elaboración (Merkle & Middeendorf, 2002). Los modelos mencionados anteriormente comparten un origen común: la neurona artificial de McCulloch y Pitts (McCulloch & Pitts, 1943; Rumelhart, McClelland, & el grupo PDP, 1992); no obstante, han tomado direcciones tan amplias que difícilmente se pueden agrupar en lo que se denominan Re-

des Neuronales o conexionismo, y actualmente coexisten los modelos fundamentados en el comportamiento neuronal y el aprendizaje con otros modelos basados en los procesos de evolución genética, dentro de los que se encuentran los modelos que se basan en las ya mencionadas colonias de las hormigas (Merkle et al., 2002). Todos estos modelos han sido posibles de estudiar gracias a los computadores ya que están basados en procesos reiterativos, en los que una operación matemática sencilla se puede repetir cientos de veces, y utilizan un alto contenido de números aleatorios para lograr que sean estocásticos, es decir, que su comportamiento sea probabilístico y no determinista.

Dada la diversidad de los modelos, son pocas las cosas que tienen en común; aún así, se pueden identificar al menos tres: (a) unidades de procesamiento simple que realizan operaciones matemáticas sencillas, (b) un comportamiento individual que caracteriza lo que pueden hacer las unidades simples y (c) un comportamiento colectivo que determina la manera como estas unidades simples deben interactuar con otras unidades.

El interés que han generado estos modelos se podría atribuir a muchas razones dentro de las que se destaca su capacidad de aprendizaje, entendida como "cualquier cambio en un sistema que le permite resolver mejor una tarea por segunda vez u otra tarea similar" (Zhu, & Simon, 1987). Esto tiene diversas repercusiones no sólo en la psicología o las ciencias del comportamiento, sino también en las matemáticas y las ciencias de la computación. Para la psicología, dichos modelos se convierten en herramientas que le permiten simular y estudiar detalladamente fenómenos cognoscitivos o comportamentales, para la matemática y la computación son modelos que permiten optimizar funciones y realizar tareas de clasificación en situaciones en las que otros procedimientos no alcanzarían un resultado óptimo.

Actualmente, los modelos conexionistas son principalmente utilizados en tres grandes áreas: (a) en el modelamiento del sistema nervioso y del comportamiento; este enfoque tiene por objetivo la construcción de modelos que ayuden a entender un fenómeno específico; por ejemplo, simular el procesamiento cerebral que ocurre en una lesión cerebral como la esquizofrenia (Cheng, 1994). (b) Como procedimientos que controlan *hardware*, los cuales son una típica aplicación industrial, en la que los algoritmos basados en modelos conexionistas responden ante situaciones para las que los algoritmos tradicionales no pueden dar una respuesta adecuada debido a la naturaleza dinámica del problema, un ejemplo de este enfoque es lo que hizo Rivals (1995) al entrenar una Red Neuronal Artificial (RNA) para que controlara un vehículo Mercedes 4-WD mientras se desplaza longitudinalmente. Y (c) como métodos de análisis de información (Price, 2000, p. 41). En este apartado están todas aquellas aplicaciones en las que se utiliza las RNA como algoritmos de

búsqueda y clasificación; un ejemplo, de este apartado es la utilización de RNA para hacer predicciones financieras en situaciones de alto riesgo, como la predicción de los precios del próximo día para la US Treasury (O'Rourke, 1999).

Aparte de los detalles puramente técnicos, la aplicación de estos modelos requiere un análisis cuidadoso del problema que se pretende abordar para lograr determinar cuál es el modelo más apropiado. El presente artículo escrito por psicólogos es un resumen detallado que pretende orientar la aplicación de este tipo de modelos a un problema en concreto, y tiene como objetivo ayudar a la popularización de estos modelos demostrando su pertinencia y sobre todo demostrando que son modelos factibles de aplicar.

EL DILEMA DEL PRISIONERO

Como en cualquier aplicación ilustrativa, la elección del problema a abordar fue uno de los asuntos más importantes, dada su pertinencia y relevancia en el contexto psicológico (primordiales para la justificación del presente trabajo); por esta razón, se seleccionó un problema ampliamente investigado, al menos, en el área de economía (Reyes, 2003): el Dilema del Prisionero (DP).

La teoría de juegos ha propuesto un análisis de la interacción entre individuos que actúan proceden de con reglas específicas. A través de los años, un juego conocido como el DP ha atraído la atención de las áreas de conocimiento más diversas que han tenido algún interés en entender el comportamiento entre los organismos (Reyes, 2003), entre las que se encuentran la economía, la biología evolutiva, la zoología, la psicología social y la experimental, las matemáticas, la física, las ciencias de la computación, la sociología, la filosofía, la política, las relaciones internacionales y demás áreas enfocadas a la resolución de conflictos, justicia y persuasión (se recomienda revisar a Axelrod, 1984; Caporael, Dawes, Orbell, y Van de Kragt, 1989; Rachlin, 2002, quienes tratan extensamente el tema).

El DP ejemplifica el conflicto entre los intereses individuales y colectivos de quienes toman decisiones disyuntivas, las cuales son igualmente factibles y defendibles (Barrios s.f.), entendiéndose conflicto como una clase especial de frustración que ocurre cuando una meta es bloqueada por una meta competitiva (Beck, 1990). Su versión original involucra a dos jugadores que deben escoger entre dos alternativas de respuesta, generalmente llamadas cooperar y desertar, donde los individuos tienen sólo una oportunidad de responder y no saben cuál será la elección del otro. Cooperar es elegir una alternativa que trae para ambos jugadores una recompensa más alta que la que obtendrían con la otra alternativa que ofrece beneficios individuales; en otras palabras, es adoptar una estrategia que beneficie a todos los jugadores. Desertar es elegir una alternativa que le ofrece una recompensa alta al jugador que la elige, y que generalmente sólo lo beneficia a él.

Dependiendo de la combinación de las decisiones tomadas por ambos jugadores, cada uno recibe una de cuatro recompensas posibles a menudo denotadas por *T* (*temptation*), que representa la recompensa más alta posible y la tentación para desertar; *S* (*sucker*), representa la peor recompensa posible y resulta cuando uno de los dos jugadores ha cooperado y el otro le deserta; *P* (*punishment*), que significa el castigo por la deserción mutua, y *R* (*reward*) es la recompensa por la cooperación recíproca. La asignación de estos valores obedece a una regla que ha sido referida por varios autores: $S<P<R<T$ (Green, Price & Hamburguer, 1995; Rosen & Haaga, 1998; Axelrod, 1984), en la que se observa una relación entre los valores de las recompensas en la que siempre debe ser mayor la recompensa que se obtiene cuando alguno de los prisioneros deserta y el otro coopera, inclusive por encima de la situación en la que ambos deciden cooperar. Sin embargo, se recomienda que el valor de se obtenga a partir de la fórmula $R>(T+S)/2$ (Green, Price & Hamburguer, 1995, Rosen & Haaga, 1998 y Axelrod, 1984), con el fin de que la alternativa de cooperar sea también atractiva. Axelrod explica lo anterior de la siguiente manera: una parte de la definición del dilema del prisionero es que los jugadores no pueden librarse (salir, escapar) de su dilema explotando al otro en cada turno, como sí lo harían en un juego sencillo, porque saben que aunque les puede ir mal, ésta sigue siendo la mejor opción. Asumir esto significa que si en cada oportunidad deserto y soy desertado, la recompensa para cada jugador no es tan buena como la recompensa por la cooperación mutua. Por esto, se deduce que el promedio de la recompensa por la reciprocidad en la cooperación es mayor que el promedio de la recompensa por la tentación (*T*) y la paga por actuar de la forma (*S*). Lo anterior, junto al *ranking* de las recompensas, define al Dilema del Prisionero.

Los valores de recompensa se pueden ver más claramente en la matriz de la figura 1. De acuerdo con la teoría de juegos, en la versión de un solo ensayo, es decir, cuando hay un solo enfrentamiento entre los jugadores, la mejor elección es desertar, porque esta acción maximiza la recompensa sin tener en cuenta lo que su oponente haga (Lloyd, 1995); pero el dilema radica en que si ambos desertan cada uno recibe una recompensa de menor utilidad de la que tendrían si ambos hubieran cooperado, como se puede ver en la figura 1. Sin embargo, las variaciones al DP o su complejidad dependen del interés del investigador y del área en que trabaje.

Existe una versión iterada del juego en la que los participantes se enfrentan más de una vez y donde, a diferencia del juego sencillo, la mejor opción podría ser cooperar porque el problema que enfrentan los jugadores es distinto ya que la recompensa final no es la inmediata sino que debe pensarse como una serie de consecuencias que hay que tratar de maximizar. Cooperar podría ser la mejor opción para alcanzar el objetivo final en esta

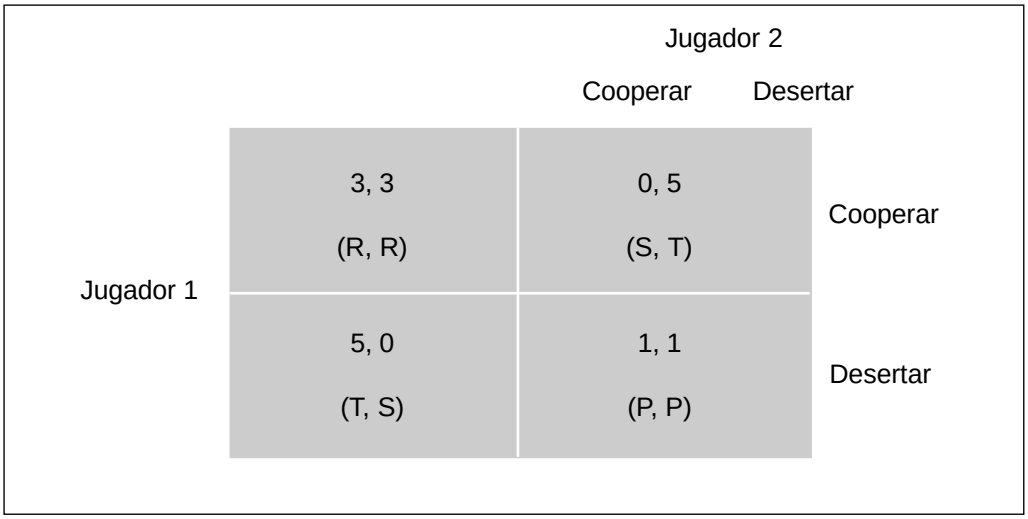

FIGURA 1. Matriz de recompensas para el ensayo del dilema del prisionero. En cualquiera de las cuatro celdas, el número de la izquierda es el resultado para el jugador 1 y el número de la derecha es el resultado para el jugador 2.

variación del juego si se considera que el bienestar común puede traducirse en bienestar individual. Sin embargo, hay juegos de este tipo en los que cooperar es la peor estrategia; la definición más clara de lo que puede ser la respuesta racional en este juego, es que mi respuesta racional (si racional implica ser la más óptima) depende de la estrategia del otro jugador, porque como lo menciona Hargreaves y Varoufakis (1996), aunque no confesar se puede ver como "la mejor" alternativa, el hacer creer al otro que no confesaremos podría ser aún mejor (para profundizar en este análisis, véase Rachlin, 2000; Hargreaves, et al., 1996, quienes proponen versiones del DP muy interesantes). En esta versión del dilema del prisionero se pueden inducir comportamientos de reciprocidad, pueden enviarse mensajes de "no agresión", pueden desarrollarse estrategias de juego y/o desarrollar una reputación (Shafir & Tversky, 1992).

Como Wilke, Messick, & Rutte afirman: "una alternativa más implica que en cada uno de los enfrentamientos los individuos pueden tomar decisiones cooperativas que maximizan las ganancias grupales mientras sacrifican algunas que son mejores a corto plazo a nivel individual, o pueden desertar para tomar una decisión no cooperativa que maximiza inmediatamente las ganancias individuales mientras que limita las ganancias del grupo (Camac, 1986). Cuando todos los individuos en el grupo actúan para maximizar sus ganancias individuales el grupo obtiene ganancias deficientes" (Wilke, et

al., 1986, citado por Rosen, Haaga, 1998, p. 143).

Esta versión iterada del DP es la más apropiada para abordar el problema de la cooperación, dado que los participantes deberán estar expuestos a la situación de elección más de una vez; de esta manera, la experiencia de interacción afecta significativamente la decisión que toma cada uno de los individuos. Por esta razón nuestro interés se centró en el DP iterado, y el problema de investigación es lograr determinar si esta versión del DP se podía simular a través de modelos conexionistas.

MÉTODO

El modelado a través de sistemas conexionistas involucra una serie de pasos específicos para determinar cuál de todos los modelos es el más apropiado para abordar el problema. Entonces se describirá inicialmente la ruta que nos llevó a tomar la decisión de utilizar el Aprendizaje por Reforzamiento, para después sí describir detalladamente el procedimiento de la investigación específica.

PROCEDIMIENTO PARA LA ELECCIÓN DEL MODELO

El primer paso, y tal vez el más importante, es el de seleccionar el modelo de aprendizaje a utilizar, ya que de su elección depende el enfoque que se va a adoptar. Inicialmente, se tomó la decisión de utilizar modelos conexio-nistas basados en el aprendizaje, ya que anteriormente se habían utilizado para abordar el mismo problema (Burgos, 1999a; 199b; 2001). Actualmente se puede afirmar que existen tres enfoques diferentes en los modelos basados en el aprendizaje, cada uno de ellos aplicable a ciertos problemas según las características de cada modelo. El primero de ellos es el *aprendizaje supervisado* (Lin et al., 1996), que se caracteriza porque el sistema no sólo debe alcanzar un objetivo, sino que dicho objetivo está claramente determinado. Un buen ejemplo del tipo de problemas que se pueden abordar con este modelo es la toma de decisiones clínicas (Price, Spitznagel, Downey, Risk & El-Ghazzawy, 2000), ya que en este problema hay una respuesta correcta y el sistema debe encargarse de clasificar adecuadamente al paciente dentro de un diagnóstico específico. El más común de los modelos de aprendizaje supervisado es el de la Retropropagación del Error (para encontrar más información acerca de ese modelo se recomienda revisar Russel & Norving, 1995; Lin et al., 1996; Haykin, 1999; O'Reilly & Munakata, 2000).

El segundo tipo de modelos se suele denominar *aprendizaje no supervisado* (Lin et al., 1996). En éste, el sistema se enfrenta con un problema que no tiene respuesta correcta y su objetivo es encontrar regularidad entre los datos presentados para clasificarlos. Es necesario aclarar que aunque el objetivo sea clasificar, como en el anterior modelo, la clasificación no se construye a partir de criterios externos que

retroalimenten el desempeño del sistema, sino que surge de las regularidades que presenten los datos.

El tercer modelo es el *aprendizaje por reforzamiento* (AR) (Lin et al., 1996) y (Haykin, 1999), en el que también existe una meta que el sistema debe alcanzar, lo que no implica que haya una única manera de lograrlo. Es un modelo muy apropiado en problemas en los que el objetivo sea la optimización de un proceso con condiciones cambiantes (Sutton & Barton, 1998).

En el proceso de decidir cuál modelo era el más adecuado, descartamos el *aprendizaje supervisado* por varias razones, la primera de ellas es porque es un modelo de aprendizaje fuera de línea (*off-line*) en el que los sistemas son entrenados para luego ponerlos en la situación que deben enfrentar. Este modelo no permite evaluar la aparición espontánea de un comportamiento específico no entrenado, para el caso la cooperación, por lo que bajo estos modelos las explicaciones serían tautológicas. Así, si nuestro interés fuera evaluar la emergencia espontánea, del comportamiento de cooperación bajo un modelo de aprendizaje supervisado, no podríamos decir que la aparición de este comportamiento es una emergencia espontánea puesto que el agente habría sido entrenado intencionalmente para ser cooperador. Necesitábamos, por tanto, un modelo de aprendizaje en línea, en el que las respuestas del individuo dependieran directamente de la situación que deben enfrentar sin un entrenamiento previo que precondicionara sus respuestas. Dadas estas condiciones, la mejor elección es el modelo de AR, que presenta un proceso de aprendizaje en línea que responde adaptándose directamente a las contingencias del ambiente; además, curiosamente es un modelo que utiliza la misma terminología del Análisis Experimental, por lo que su adaptación resultó más fácil de lo esperado.

El modelo de AR posee unas características distintas a las ya mencionadas: aunque son modelos conexionistas, no se basan en el comportamiento neuronal, por lo que sus elementos y la lógica dentro del modelo difieren significativamente. Los modelos de AR se caracterizan porque la unidad mínima de procesamiento se denomina "agente" (Sutton & Barton, 1998), el cual emite una serie de acciones. Dicho agente es introducido en un ambiente que responde ante sus acciones cambiando su configuración. Cada configuración se denomina "estado", que es la respuesta del ambiente ante las acciones del agente. El objetivo del agente es alcanzar la máxima recompensa posible y para lograrlo debe, a través de la interacción con el ambiente, identificar cuál o cuáles son las acciones que le ofrecen los mejores rendimientos; para esto cuenta con una función matemática que le permite estimar el valor de cada acción con el fin de poder compararlas y elegir la que le permita maximizar su ganancia. En este proceso de interacción agente-ambiente entran en juego varios ele-

mentos adicionales: se puede inducir cierta inclinación hacia alguna acción específica e incluso se podría generar en el agente un factor emocional que le facilite u obstaculice el análisis que el agente hace de sus acciones. A este elemento se le denomina 'política' (Sutton et al., 1998).

Una vez se han definido los elementos del modelo, el siguiente paso es identificar claramente los componentes del problema para poder traducirlos en los términos de los elementos del modelo. En el dilema del prisionero iterado podemos considerar a cada uno de los sospechosos como agentes. Las acciones que tales agentes pueden realizar son delatar o encubrir y cada una de éstas se convierte en el estado del otro agente, de tal manera que los estados del ambiente son haber sido delatado o haber sido encubierto.

En el problema del dilema del prisionero, el resultado de la elección de cada agente da lugar a una recompensa, situación que se ajusta perfectamente al modelo de AR que depende tanto de la elección que ha hecho el sospechoso como de la elección que hizo su compañero. Generalmente, las recompensas que se otorgan guardan una relación (Axelrod, R, 1984) en la que la elección de delatar obtiene mayores recompensas que la elección de encubrir; sin embargo, esto está condicionado a la elección que haya realizado su compañero.

En esta investigación se tuvieron en cuenta los valores de recompensa determinados por Axelrod (1984) mencionados en la parte inicial del artículo, en los que se puede ver que el agente al tomar la acción de encubrir encontrándose en un estado de delatado, sería retroalimentado con un puntaje de 0; la segunda forma de retroalimentación otorgaría un puntaje de 1 si tanto la acción como el estado fueran delatar, a diferencia del caso en el que el agente tomara la acción de delatar y fuera encubierto (caso en el que recibiría 5); para completar las alternativas de recompensas, si la acción y el estado se encuentran en la alternativa de encubrir, se retroalimentaría con un 3.

El último elemento necesario a traducir es la política, que podría representar alguna inclinación de los sospechosos por delatar o por encubrir, es decir, por una inclinación a ser cooperador o desertor. Dicha inclinación podría representar algo de la historia de vida de estos sospechosos o algún factor emocional que los induzca hacia estos comportamientos. En la presente investigación se decidió no incluir la política debido a que si el objetivo era observar si la conducta cooperativa se manifestaba, incluirla hubiese provocado que termináramos induciendo la respuesta que esperábamos encontrar.

Una vez se tiene claridad sobre los elementos del problema descritos en términos del modelo, el siguiente paso es interrelacionarlos a través de la función de valor (*Vx*). Dicha función establece la relación que hay entre las acciones, los estados, las recompensas y la política; se define como $Vx = \sum_i P^{a_i}_{ss'} R^{a_i}_{ss'}$ (Sutton et al., 1998), donde $P^{a_i}_{ss'}$ es la probabilidad de que al

realizar la acción a_i, encontrándose en el estado se pase al estado *s'* y $R^{a_i}_{ss'}$ es la recompensa que alcanza al realizar la acción a_i, encontrándose en el estado *s* y quedando en el estado *s'*. Es necesario aclarar que a esta fórmula le hemos suprimido el término $\gamma\, v^{\pi}(s')$ que corresponde a la política, que como hemos anotado antes no fue tenida en cuenta. Como se puede ver en la fórmula, el valor de una acción depende directamente de la experiencia que el agente haya obtenido con ésta, en términos de la cantidad de veces que ha tomado dicha decisión y las consecuencias que le ha acarreado.

Lo anterior puede ejemplificarse de una manera más sencilla si consideramos a los "Agentes" como dos organismos electrónicos que poseen un repertorio de comportamientos bastante sencillo: sólo saben hacer dos cosas, encubrir o delatar. Estos organismos son puestos en un ambiente en el cual recibirán una recompensa por sus acciones; dichas acciones tienen además un efecto sobre la recompensa que su compañero recibirá, siguiendo los valores que generalmente se utilizan en el DP. Estos organismos poseen una motivación que los motiva a alcanzar la máxima recompensa posible; para ello utilizan su experiencia anterior, que es representada por la cantidad de recompensa que obtuvieron cuando tomaron una determinada decisión. Este aspecto es controlado por el término $R^{a_i}_{ss'}$ de la función de valor que le indica al organismo cuánta recompensa ha obtenido al tomar cada una de las dos posibles decisiones. Adicionalmente, el organismo también tiene información sobre el comportamiento de su compañero y conoce cuál es la probabilidad de que él tome una determinada decisión; esto está consignado en el término $P^{a_i}_{ss'}$. Sin embargo, es necesario aclarar que, aunque las decisiones de cada organismo influenciaban la recompensa que alcanzaba su compañero, ninguno de los dos organismos tenía conocimiento de la presencia del otro; para ellos no había otro compañero, sólo había un ambiente con el que interactuaban.

Aparte de todos estos elementos era necesario incluir uno adicional que determina la forma como el organismo toma sus decisiones y que nuevamente es necesario contextualizar para poder explicarlo. Imaginemos a los dos organismos en el momento en que son enfrentados con el ambiente, ninguno de ellos sabe cuál es la decisión que más recompensa le puede ofrecer, ni siquiera conocen cuánto recibirán con cada decisión que tomen. En este momento los organismos tienen dos opciones: o dedican tiempo a conocer su ambiente, o se aferran a una alternativa de decisión que aparentemente les ofrezca una buena recompensa. Este conflicto se denomina el dilema entre exploración y explotación.

Debido a la naturaleza "en línea" de este tipo de modelos en los que generalmente no hay una etapa de entrenamiento o adaptación previa a la situación experimental, el agente u organismo se enfrenta a su ambiente

sin conocerlo y debe dedicar algún tiempo a esta tarea; sin embargo, debe saber administrar muy bien su tiempo, dado que la meta es lograr la máxima recompensa posible en un periodo de tiempo limitado, para garantizar que no sólo va a explorar, sino que logrará también explotar alguna o algunas de las alternativas que ha encontrado como las "mejores" o las que mejor rendimiento le ofrecen. Esta dicotomía exploración-explotación es uno de los elementos que consideramos primordiales dentro de la presente investigación y es el único valor que manipularemos, por lo que puede considerarse como la variable independiente; la intención es evaluar su efecto sobre las decisiones que tomen los agentes y las consecuencias que éstas les proporcionen, es decir, nuestras variables dependientes fueron el promedio de recompensas alcanzadas y el número de veces que fue elegida cada acción.

SUJETOS

A partir de las características del AR, los sujetos utilizados en la investigación fueron dos 'agentes electrónicos'. Ellos tenían la capacidad de elegir entre delatar y encubrir evaluando su decisión a partir de la función de valor descrita anteriormente.

INSTRUMENTOS

Se utilizó un programa de computadora desarrollado con Borland Delphi © en el lenguaje orientado a objetos Object Pascal ©, ya que es un ambiente de desarrollo fácil que permite generar aplicaciones para sistemas operativos Win32[2].

PROCEDIMIENTO

Ya definida la manera como sería modelado el DP a través del AR, nos dedicamos a la implementación del modelo (lo que generalmente se denomina "correr" el programa). Esta etapa involucró la precisión de algunos aspectos que no son lo suficientemente claros en los libros, y que consideramos uno de los principales aportes del presente artículo por cuanto es una orientación de la manera como se deben interpretar algunos de los aspectos más importantes a la hora de "correr" el programa.

El primero de tales aspectos es la integración de la fórmula matemática del valor con la del concepto de exploración-explotación, pues se debe tener en cuenta que la fórmula del valor arroja una información que indica cuál es la acción que más valor le ofrece al agente y la probabilidad de seleccionarla, que debería ser más alta si está en un periodo de explotación que si se encuentra en un periodo de exploración. Aquí es importante recalcar que consideramos que el comportamiento de los agentes u organismos no podía ser determinista, por lo que era necesario incluir algún mecanismo que evaluara el resultado de la fun-

2 Son todas aquellas aplicaciones que hacen uso de la memoria en bloques de 32 bits, y que corren en Microsoft Windows 95 © o versiones posteriores.

ción de valor otorgándole un componente estocástico al comportamiento de los agentes.

Interpretando la explotación como una probabilidad, decidimos definirla como una alta probabilidad de seleccionar la acción que más valor ofrece; específicamente, si este valor es alto, el agente se encuentra en una etapa de explotación en la que la mayoría de las veces optará por la alternativa que más recompensa le ofrece, lo que generalmente implica que el agente se dedique a realizar una sola acción. Si este porcentaje es bajo, el agente se encuentra en un periodo de exploración en el que la mayoría de las veces optará por la alternativa que le ofrece menor recompensa, para lograr de esta manera un conocimiento más amplio de lo que sucede en su entorno.

Para representar esto se utiliza una fórmula matemática que compara el porcentaje de explotación con el valor resultante de la generación de un número aleatorio con distribución uniforme que oscile entre 0 y 1; si este número es inferior al porcentaje de explotación, el agente realizará la acción que más recompensa le ofrece, de lo contrario elegirá la acción opuesta.

Lo más interesante de la probabilidad de explotación es que se puede equiparar con la motivación, de tal manera que cuando los organismos se encuentran en una explotación baja, podemos afirmar que son arriesgados; esto implicaría que el organismo estaría desechando la opción que más recompensa le otorga a corto plazo en búsqueda de otras alternativas. Por el contrario, cuando tienen un porcentaje de explotación alto se puede considerar que son organismos ambiciosos; esto quiere decir que optan siempre por la alternativa que les ofrezca la mejor recompensa a corto plazo.

A partir de todos estos elementos se generaron los primeros resultados; en esta etapa todo se hizo a lápiz y papel, con lo que se buscaba probar la efectividad de las fórmulas antes de implementarlas como parte de un programa de computador; a partir de estos resultados se diseñó la primera versión del programa en la que se tenían en cuenta todos los factores anteriormente descritos.

Sin embargo, ésta no fue la única versión que se realizó del programa, ya que a partir de los resultados encontrados en la primera, se decidió incorporar un nuevo elemento que permitiera que el nivel de explotación variara con el tiempo; a este elemento lo denominamos Explotación Dinámica (ED), en contraste con el valor de Explotación Estática (EE) manejado en la primera versión. La ED consiste en representar la explotación a través de una función sigmoide de fórmula

$$P_i(\theta) = li + \frac{ls}{1+e^{a(\theta - b_i)}} \quad i=1,2,3....,n^3,$$

donde $P_i(\theta)$ representa el nivel de explotación en el momento i, li es el mí-

3 Esta fórmula es una adaptación de la función característica de un ítem que se describe en (Herrera, Sánchez y Jimenez, 2001. p. 319)

FIGURA *2*. Resultados de un juego de 200 iteraciones con las siguientes condiciones iniciales: el agente 1 comenzó desde una posición de encubrir, lo mismo que el agente 2; EEA = 0.9 y semilla de números aleatorios = 846.

nimo nivel de explotación que tendrá el agente a lo largo de todos los ensayos, *ls* es el máximo nivel de explotación que alcanzará el agente durante todos los ensayos, *a* es una variable que determina qué tan suave será el cambio en el nivel de explotación en el transcurso de la totalidad de los ensayos, *e* es la base de los algoritmos neperianos y tiene un valor aproximado de 2.71 y, por último, está *b,* que es la variable que determina qué tan rápido pasa el agente de un nivel de exploración a un nivel de explotación.

Primera versión

Su diseño busca ser lo más versátil posible, permitiéndole al usuario configurar todas las variables que intervienen en el dilema del prisionero: las condiciones iniciales, el porcentaje de explotación y el número de iteraciones que ocurrirán en un juego. Adicionalmente, para poder correr la aplicación, es necesario configurar un valor denominado 'semilla', que determina la condición inicial a partir de la cual se generaran los números seudoaleatorios, esto es fundamental para garantizar que cualquier juego pueda ser replicado.

La aplicación genera los resultados en un formato gráfico y también en uno numérico con el fin de facilitar su interpretación, en la figura 2 se muestra un ejemplo de dichos resultados, en el que se observa que el agente 1 alcanza una recompensa mucho mayor que la que alcanza el agente 2, par-

FIGURA 3. Resultados de un juego de 200 iteraciones con las siguientes condiciones iniciales: el agente 1 comenzó desde una posición de delatar y el agente 2 desde una posición de encubrir, constantes para fórmula de explotación dinámica = 9, = 1.6, = 0.8, = 0.1 y la semilla de números aleatorios = 108.

tiendo de la misma condición inicial. La línea más oscura de la gráfica representa el comportamiento del agente 1 y la línea más clara el comportamiento del agente 2. Las líneas toman el valor de uno (1) cuando el respectivo agente toma la decisión de encubrir, y 0 en el caso contrario. La línea continua que cruza por encima de las otras líneas representa el Porcentaje de Explotación, que para este caso es Estático, por lo que se observa una línea completamente horizontal que no cambia con el tiempo, como sí ocurre cuando se configura una ED donde esta línea toma una forma de "s" (véase Figura 3).

Con la primera versión del programa se realizó un total de 456 juegos que constaban de 200 iteraciones, es decir, cada juego constaba de 200 situaciones en las que ambos agentes debían tomar una decisión de encubrir o delatar. Los 456 juegos representan un total de 91.200 decisiones por agente.

A lo largo de estos juegos se manipularon 3 valores diferentes de Explotación (véase tabla 1), que corresponden a tres comportamientos de explotación; el primero es una Explotación Estática Baja (EEB), en la que la probabilidad de que los agentes opten por la decisión que más valor les proporciona en un momento determinado es de tan sólo el 10%; el segundo es una Explotación Estática Media

TABLA 1. Diferentes valores de EE que se utilizaron para generar los resultados con la primera versión del programa

Valor de EE	No. de juegos
0.1	152
0.5	152
0.9	152

TABLA 2. Diferentes valores del término *b* de la ED que se utilizaron para generar los resultados con la segunda versión del programa

Valor de EE *b*	No. de juegos
3.5	162
6	156
9	152

(EEM), en la que dicha probabilidad asciende al 50%, y el tercer valor es una Explotación Estática Alta (EEA), en la que el porcentaje de explotación asciende al 90%. Se tomaron estos tres valores por considerar que representaban tres situaciones diferentes. En la primera de ellas, las decisiones de los agentes están regidas por la búsqueda de alternativas en su ambiente; en la segunda, se representaba un equilibrio entre estas dos alternativas (y los agentes explotaron y exploraron en la misma medida) y, en la tercera alternativa, los agentes se dedicaron a explorar en la mayoría de las ocasiones.

Cada grupo de 152 juegos estaba dividido en 4 subgrupos que se relacionaban con las decisiones iniciales de cada agente. Estas decisiones iniciales representan un factor muy importante que puede influir en el comportamiento del agente a lo largo de las iteraciones. Debido a que los agentes no cuentan con ninguna información para tomar la decisión inicial, se decidió incluirla dentro de las variables que se deben especificar en la configuración inicial de cada juego, de tal manera que le informe a la aplicación cuál es la decisión inicial de cada agente. Dado que las posibles combinaciones de decisiones que pueden tomar los agentes son sólo 4, se dividieron los 152 juegos en 4 grupos de 38; en los primeros, ambos agentes comenzaban con una decisión inicial de encubrir, en el segundo, ambos agentes tomaban como decisión inicial delatar, en el tercer grupo, de juegos el agente 1 empezaba delatando mientras que el agente 2 encubriendo, y en el cuarto grupo, el agente 1 comenzaba encubriendo y el 2 delatando.

Segunda versión

Esta aplicación mantuvo las mismas características de la inicial; adicionalmente permitía configurar una ED. Con esta versión se realizaron 473 juegos, cada uno de ellos de 200 iteraciones que representan 96.600 decisiones tomadas por cada agente. Al igual que en la primera versión, se manipularon tres valores diferentes de explotación a través de la utilización de tres valores diferentes para el término *b*. Dichos valores se resumen en

TABLA 3. Resumen de los promedios y desviaciones estándar de las recompensas alcanzadas y de las veces que encubrió cada agente a partir de cada uno de los valores de EE que se manipularon.

		Agente 1		Agente 2	
		Promedio	Desviación estándar	Promedio	Desviación estándar
EEA	Recompensa	420.87	230.39	383.31	204.76
	Decisión de encubrir	79.40	63.72	86.95	69.43
EEM	Recompensa	446.77	29.04	449.24	29.86
	Decisión de encubrir	99.94	7.47	99.45	7.07
EEB	Recompensa	443.91	60.08	454.05	82.84
	Decisión de encubrir	108.94	29.05	106.91	24.55

la tabla 2 y corresponden a tres variaciones de la ED; en el primero de ellos ambos agentes comenzaban explorando su entorno durante unos pocos ensayos y rápidamente empezaban a explotar, por lo que lo denominamos Explotación Dinámica Rápida (EDR); con el segundo valor el proceso de exploración era un poco más largo y hacia la mitad de los ensayos empezaba la explotación, a éste lo denominamos Explotación Dinámica Media (EDM); y el último valor correspondía a una exploración bastante larga y una explotación muy corta al final de las iteraciones, debido a esto lo denominamos Explotación Dinámica Demorada (EDD).

RESULTADOS

Una vez realizados todos los juegos se obtuvieron los siguientes resultados: los promedios de recompensas obtenidos por ambos agentes son similares en los valores de EEM y EEB, pero se presentaron diferencias cuando se utilizó una EEA.

Se encontraron diferencias en las desviaciones estándar (véase tabla 3) entre los diferentes valores de la EE, es decir, resultaba una enorme dispersión observada en el grupo EEA. Con respecto al promedio de veces en que los agentes decidieron encubrir, se encuentran diferencias entre distintos valores de la EE; en los juegos EEB se observan promedios y desviaciones más altas de la decisión de encubrir, y dichos promedios descienden a medida que aumenta la EE.

El análisis gráfico de la figura 4 (pág. 45) permite identificar tres tipos de comportamientos diferentes en las recompensas alcanzadas por los agentes que coinciden con los valores asignados a la EE. En los juegos con EEB las recompensas alcanzadas por ambos agentes son muy parecidas con un índice de correlación alto (tabla 4, pág. 46), lo que indica que durante este período las decisiones que tomaron ambos agentes los condujo a obtener niveles altos de recompensas simul-

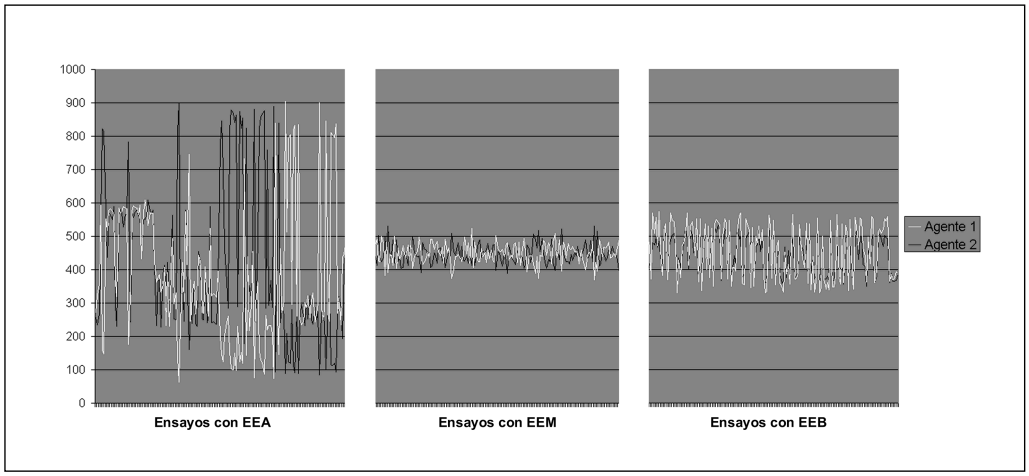

FIGURA 4. Gráfica que resume el comportamiento de las recompensas alcanzadas por los agentes a lo largo de los 473 juegos, en la que los primeros 162 corresponden con una EEA, los siguientes 156 con una EEM y los últimos 155 con una EEB.

táneamente; en la EEM se observa una correlación alta y negativa que indica que las decisiones tomadas por los agentes provocaba que cuando uno obtenía una recompensa alta su compañero obtenía lo contrario, y viceversa; por último, en la EEA se pueden observar las diferencias individuales más grandes y es posible identificar 4 comportamientos diferentes que se relacionan con las condiciones iniciales.

Adicionalmente, es importante recalcar que las condiciones iniciales no afectaron en las dos primeras condiciones, EEB y EEM, pero se volvieron un factor muy importante en la EEA, debido a que el promedio de recompensas alcanzados durante la EEA depende de las decisiones iniciales; por ejemplo, se observó que cuando ambos agentes empiezan encubriendo los promedios de las recompensas son mucho más altos que cuando ambos agentes empiezan delatando; así mismo, se puede identificar claramente que cuando ambos agentes parten de decisiones diferentes, el agente que empieza delatando va a ganar mejores recompensas que aquel que empieza encubriendo.

SEGUNDA VERSIÓN

En general, los resultados que se obtuvieron variaron un poco con los obtenidos en la primera versión; por ejemplo, se encontró que el promedio de recompensas conseguido por ambos agentes aumentaba a medida que se demoraba la explotación, lo que corresponde con el aumento del parámetro. Los promedios de recompensas más altos se encontraron con

TABLA 4. Resumen de las correlaciones entre las recompensas que alcanzaron los agentes durante cada uno de los tres períodos de EE

Período	Correlación
EEB	-0.4335
EEM	-0.7541
EED	0.9530

la EDD, así mismo, el número de veces que los agentes decidieron encubrir fue superior con este tipo de explotación. Individualmente se lograron los mejores y los peores resultados por cada ensayo cuando se manejaba una EDR. Los promedios de recompensas y el número de veces que encubrieron se resumen en la tabla 5 (pág. 47). También se observó una relación entre los diferentes valores de la ED y el promedio de veces que cada agente encubrió siguiendo el mismo patrón de los resultados obtenidos con la primera versión, es decir, a medida que aumenta el nivel de explotación disminuye el promedio de veces que deciden encubrir.

El análisis gráfico de los resultados arrojó algunos datos adicionales, uno de ellos es que progresivamente se pueden identificar 3 momentos diferentes en la gráfica (véase figura 5 pág 47), además se observa lo antes mencionado con respecto a las recompensas. Adicionalmente se realizaron correlaciones entre las recompensas alcanzadas por los agentes en los diferentes juegos, y se encontraron correlaciones negativas para los juegos con EDR y EDM, lo que significa que durante estos juegos, cuando uno de los dos agentes alcanzaba una recompensa alta, su compañero obtenía una recompensa pequeña y viceversa; por el contrario, en los juegos con EDD, las correlaciones entre las recompensas que obtuvieron los agentes fueron positivas y muy altas, lo que indica que durante este período cuando uno de los agentes lograba una recompensa alta, su compañero también lo hacía y viceversa. Los resultados de las correlaciones se reportan en la tabla 6 (pág 47).

Comparando los resultados obtenidos con la EE y la ED, se puede observar que en general la EE ofrece mejores recompensas en promedio que la ED, sólo equiparables cuando se utiliza una EDD. Comparando cada valor de explotación se puede observar que los valores de explotación más alta (EEA y EDR) corresponden a los valores de recompensas más altos individualmente, pero más bajos en promedio con índices de correlación altos y negativos; con respecto a los valores de explotación media (EEM y EDM), aunque presentaron un comportamiento similar, son muy diferentes en la dispersión de los datos (resultaron menos dispersos los de EEM). En cuanto a los índices de correlación, también son altos y negativos. Finalmente, los valores de explotación más baja coinciden con promedios altos y correlaciones altas y positivas.

DISCUSIÓN

El objetivo del presente artículo de investigación fue el de simular el com-

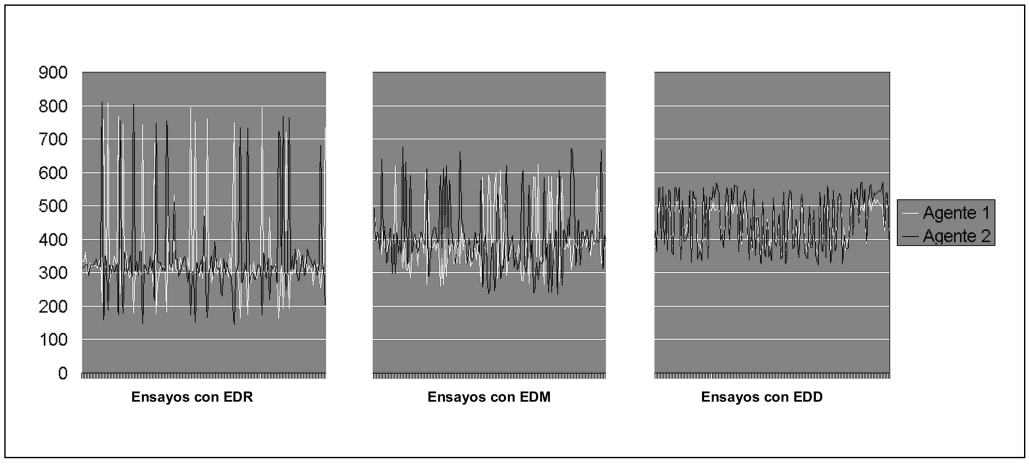

FIGURA 5. Gráfica que resume el comportamiento de las recompensas alcanzadas por los agentes a lo largo de los 473 juegos. Los primeros 162 corresponden con una EDR, los siguientes 156 con una EDM y los últimos 155 con una EDD.

TABLA 5. Resumen de los promedios de recompensas alcanzados por los agentes a partir de cada uno de los valores de Explotación Dinámica que se manipularon.

		Agente 1		Agente 2	
		Promedio	Desviación estándar	Promedio	Desviación estándar
EDR	Recompensa	344.02	130.21	339.44	125.97
	Decisión de encubrir	55.87	30.78	56.73	32.01
EDM	Recompensa	390.67	80.84	404.31	97.52
	Decisión de encubrir	83.15	25.15	80.5	20.31
EDD	Recompensa	441.35	54.2571	454.20	74.54
	Decisión de encubrir	107.451	23.90	104.99	18.57

Tabla 6. Resumen de las correlaciones entre las recompensas que alcanzaron los agentes durante cada uno de los tres períodos de ED

Período	Correlación
EDR	- 0.5258
EDM	- 0.5658
EDD	0.6192

portamiento de dos organismos sometidos al DP en búsqueda de la emergencia del comportamiento cooperativo. Lo que se pudo observar mediante la manipulación exclusiva de la variable del porcentaje de explotación fue que ambos organismos eligieron encubrir de manera consistente, generando de esta manera las recompensas en promedio más altas entre todas las alternativas. Esta variable corresponde a un elemento propio de los agentes y, como lo mencionamos, se podría relacionar con un factor motivacional, lo que nos permite visualizar que la motivación juega un papel fundamental en la manera como responden los organismos a situaciones en las que se ven enfrentados a un dilema (Rachlin, 2002), e incluso en la manera como los organismos deciden cooperar o desertar.

Para poder interpretar si estos resultados se pueden considerar argumentos que apoyen la emergencia del comportamiento cooperador es necesario evaluar si una constante decisión de encubrir se puede asumir como indicio suficiente para poder afirmar que el comportamiento está determinado por un interés de cooperación. Un punto de partida para este análisis es la evaluación de uno de los pormenores de la simulación, relacionado con el hecho de que ninguno de los dos agentes tenía conocimiento de la presencia del otro ni de las consecuencias que sus actos tenían sobre las recompensas que recibía el otro agente. Estas condiciones permiten afirmar que la conducta observada en los agentes no puede deberse a lo que se denomina cooperación, sino que debe ser producto de otras condiciones, debido a que la cooperación es un comportamiento que involucra la interacción de varios organismos en busca de un beneficio compartido.

En varias investigaciones con animales que fueron sometidos al dilema del prisionero (Green, et al., 1995 y Stephens, McLinn & Steven, 2003) se puede observar que en algunas ocasiones ellos también eligen simultáneamente la alternativa de encubrir, lo que nuevamente plantea el interrogante: ¿se puede considerar este comportamiento como un comportamiento de cooperación? Nosotros consideramos que esta pregunta tiene una respuesta sencilla: es posible siempre y cuando se garantice que ambos organismos conocen las consecuencias que sus actos tienen sobre las recompensas que obtendrán los otros. Si se ha de trabajar con animales, es necesario vislumbrar de manera clara una interacción social que no dé lugar a inferencias o interpretaciones de la conducta de los organismos, para que de esta forma la conducta de cooperar se vea en interacciones reales entre ambos organismos. Al respecto, varias investigaciones (Flood, Lendenmann, y Rapoport, 1983, citado por Green, et al, 1995) afirman que dado que ambos organismos se ven, se puede afirmar que conocen dichas consecuencias. Nuevamente, esto plantea un inconveniente: una reja a través de la cual los organismos ven, oyen y huelen a sus compañeros ¿es suficiente para

afirmar lo anterior? Al respecto, todos los autores (Green, et al., 1995 y Stephens et al., 2003) tienen muy claro que esto no es suficiente; sin embargo, no son claros los métodos que utilizan para garantizar dicha interacción. Esto queda de manifiesto en investigaciones como la de Green et al. (1995), en la que una paloma interactuaba con un computador en una situación del DP.

Así los animales puedan ver, oler u oír al otro organismo, difícilmente se puede considerar esto como una garantía de que el animal conoce el efecto que sus actos tienen sobre la conducta del otro. En el ámbito en el que nos encontramos no se puede equiparar lo que motiva a los animales a hacer lo que hacen, con lo que motiva a los seres humanos, porque de esta forma nos hallaríamos frente a una postura antropomórfica (Bruno, 1997). Nosotros consideramos que en la presente investigación no se puede hablar de la emergencia de un comportamiento cooperativo, dado que los organismos no tenían conocimiento de la existencia del otro. Esto es una conclusión que busca cuestionar las aplicaciones del DP en contextos donde no se pueda garantizar que los organismos interactúen entre sí y respondan conociendo las consecuencias de los actos de su compañero y los propios sobre las recompensas obtenidas.

Sin embargo, es necesario aclarar que aunque los organismos no tengan conocimiento de la existencia del otro y lo que esto implica, la situación a la que están sometidos sigue involucrando un dilema, ya no relacionado con cooperar o desertar, sino con recibir una "buena" recompensa en una situación en la que las contingencias ambientales siguen un patrón difícil de predecir. Por esto consideramos que la conducta de los organismos estaba controlada por un programa de refuerzo que determinaba cuál de las conductas era la que debía realizar. Al respecto, es posible que el autocontrol fuese una de las condiciones que podría determinar la conducta de los agentes si lo consideráramos como una selección deliberada de conductas del individuo en situaciones en las que obtiene consecuencias conflictivas (Kazdin, 1996 y Skinner, 1971), que, para el caso, serían una recompensa alta a corto plazo pero en promedio baja, o una recompensa baja a corto plazo pero en promedio alta.

Con respecto a este punto también surge un interrogante que afecta la explicación del autocontrol: los organismos desconocían la duración de los juegos y debido a esto no tenían suficiente información para evaluar si la conducta de preferir la decisión de menor recompensa, efectivamente les ofrecía mayores ganancias en promedio a largo plazo.

Por lo anterior se puede afirmar que el comportamiento observado no se ajusta a ninguno de los patrones que regularmente se encuentran como explicación al DP, por cuanto no se puede hablar de cooperación y tampoco de autocontrol. Aunque hay que tener en cuenta que en la simulación no están contenidos todos los factores

que comúnmente se incluyen en las investigaciones del DP, los resultados indican que las relaciones que están planteadas al interior del DP tienen una solución matemática sencilla: la mejor manera de alcanzar un promedio alto de recompensas es encubrir; esta relación determina el comportamiento de los organismos en ausencia de cualquier variable adicional; es decir, cuando no se tienen en cuenta factores que faciliten el engaño o la trampa, los organismos optarán necesariamente por un comportamiento de encubrir para poder alcanzar la mayor recompensa posible.

Finalmente, es importante recalcar la pertinencia de los modelos computacionales en la investigación del comportamiento, debido a que son modelos que permiten simularlo bajo condiciones de absoluto control, y que a la vez permiten investigaciones en las que se puede hacer una manipulación minuciosa de todas las variables que intervienen en un problema, facilitando que el investigador obtenga resultados de forma rápida y económica. Aunque consideramos que este tipo de investigación no reemplaza la investigación básica con animales o personas, sí puede servir de "filtro" que decante las preguntas que guían dichas investigaciones y que permita a su vez generar nuevos interrogantes.

REFERENCIAS

Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Barrios, A. (s.f.) Los dilemas morales en la clase de ética. En *Sociedad Andaluza de Educación Matemática.* Recuperado el 28 de noviembre de 2003, de http://thales.cica.es/rd/Recursos/rd98/Filosofia/03/filosofia-03.html#uno

Beck, R. (1990). *Motivation: Theories and principles. 3Ed.* New Jersey: Prentice Hall.

Brewer, M. (1989) "Ambivalent sociality: The human condition". En: Journal of Behavioral and Brain Sciences, 12(4), p. 699

Bruno, F. J. (1997). *Diccionario de términos psicológicos fundamentales.* Barcelona: Paidós Studio.

Burgos, J. (1999a). Cooperation as an Emergent Property of Selection by Reinforcement in Artificial Neural Networks. SQAB. Recuperado el 5 de noviembre de 2003 en http://sqab.psychology.org/abstracts-1999.html.

Burgos, J. (1999b). Economistas y Psicólogos encuentran nichos comunes. *El Ucabista.* Recuperado el 5 de noviembre de 2003 en htttp://www.ucab.edu.ve/prensa/ucabista/jul99/p05.htm.

Burgos, J. (2001). "A neural-network interpretation of selection in learning and behaviour". En: Journal of Behavioral and Brain Sciences, 24(3), pp. 531-533.

Caporael, L., Dawes, R., Orbell, J., Van de Kragt, A. (1989). "Selfishness Examined: Cooperation in the Absence of Egoistic Incentives". En: *Behavioral and Brain Sciences.*12, pp. 683-739.

Cheng, E. (1994). "A neural network model of cortical information processing in schizophrenia. I: Interaction between biological and social factors in symptom formation". En: Canadian Journal of Psychiatry, 39, pp. 362-367.

Green, L., Price, P. & Hamburguer, M. (1995). "Prisoner's Dilemma and the Pigeon: Control by Immediate Consequences". En: Journal of the Experimental Analysis of Behavior, 64, pp. 1-17.

Hargreaves, S. & Varoufakis, Y. (1996) *Game Theory: A critical introduction.* New Fetter Lane: Routledge.

Haykin, S. (1999). *Neural Networks; A Comprehensive Foundation.* New Jersey: Pretince Hall.

Herrera, A., Sánchez, N., & Jiménez, H. (2001). "De la Teoría Clásica de los Tests a la Teoría de Respuesta al Ítem". En: Aula Psicológica, 3, pp. 293-332.

Kazdin, A. (1996). Modificación de la conducta y sus aplicaciones prácticas. México D. F: Manual Moderno.

Lin, C. T. & Lee, C. S. (1996). *Neural Fuzzy Systems. A Neuro-Fuzzy Synergism to Intelligent Systems.* New Jersey: Pretince Hall.

Lloyd, A. (1995). Computing bouts of the prisoner's dilemma. Scientific American, 272 (6).

McCulloch, W. & Pitts, W. (1943). "A logical calculus of ideas immanent in nervous activity". En: Bulletin on Mathematical Biophysics, 5, pp. 115-133.

Merkle, D. & Middeendorf, M. (2002). "Modeling the Dynamics of Ant Colony Optimization". En: Evolutionary Computation, *10*(3), pp. 235-262.

O'Reilly, R. & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience. Understanding the Mind by Simulating the Brain.* Cambridge: Bradford Book The MIT Press.

O'Rourke, B. M. (1999). *Analyzing Financial Neural Networks Performance: Applying Fuzzy Clustering and Tree Classification.* PC AI, pp. 37-40.

Pack, L., Littman, M. & Moore, A. (1996). "Reinforcement learning: A survey". En: Journal of Artificial Intelligence Research.

Price, R. K., Spitznagel, E. L., Downey, T. J., Risk, N. K. & El-Ghazzawy, O. G. (2000). Applying Artificial Neural Network Models to Clinical Decision Making. *Psychological Assessment,* 12(1), pp. 40-51.

Rachlin, H. (2000). The Science of Self-Control. Cambridge: Harvard University Press.

Rachlin, H. (2002). Altruism and selfishness. *Behavioral and Brain Sciences.* 25(2), pp. 239-296.

Reyes, R. (2003) Dilema del Prisionero. En: *Diccionario Crítico de las Ciencias Sociales*. Recuperado el 5 de diciembre de 2003 en http://www.ucm.es/info/eurotheo/diccionario/P/prisionero_dilema.htm.

Rivals, I. (1995). *Modélisation et commande de processus par réseaux de neurones; application au pilotaje d' un véhicule autonome*. Université Paris 6.

Rosen, J. & Haaga, D. (1998). "Facilitating Cooperation in a Social Dilemma: A Persuasion Approach". En: Journal of Psychology, 132(2).

Rumelhart, D. E., McClelland, J. L. & el grupo PDP (Eds.). (1992). *Introducción al Procesamiento Distribuido en Paralelo*. Madrid: Alianza.

Russel, S. & Norving, P. (1995). *Artificial Intelligence. A Morden Approach*. New Jersey: Prentice Hall.

Shafir, E. & Tversky, A. (1992). "Thinking Through Uncertainty: Nonconsequential Reasoning and Choice". En: Cognitive Psychology, 24, pp. 449-474.

Skinner, B. F. (1971). *Ciencia y conducta humana*. Barcelona: Fontanella.

Stephens, D. W., McLinn, C. M. & Steven, J. R. (2002) Discounting and Reciprocity in an Iterated Prisoner´s Dilemma. *Science,* 298 (5601), pp. 2216-2219.

Sutton, R. & Barton, A. (1998). *Reinforcement Learning*. Cambridge: Bradford Book. The MIT Press.

Zhu, X. & Simon, H.A. (1987). "Learning mathematics from examples and by doing". En: Cognition and Instruction, 4, pp. 137-166.

Recibido el 16 de enero de 2004 y aceptado el 30 de enero de 2004