La Aplicación de Modelos de Consciencia Artificial en los Sistemas Multiagente

Raúl Arrabales Moreno y Araceli Sanchis de Miguel

Departamento de Informática
Universidad Carlos III de Madrid
raul.arrabales@alumnos.uc3m.es
masm@inf.uc3m.es

Resumen

Durante la última década han aparecido algunas implementaciones de modelos científicos de la consciencia basados en sistemas multiagente. El propósito de este artículo es recopilar y describir estos sistemas, determinando hasta que punto estas implementaciones satisfacen los modelos correspondientes, y analizando si proporcionan realmente las supuestas ventajas de usar consciencia artificial en la resolución de problemas. También se analizan en general las funciones de la consciencia y los beneficios que éstas pueden aportar en el rendimiento de los sistemas multiagente.

Palabras Clave

Consciencia artificial, sistemas multiagente, atención.

Introducción a la consciencia artificial

A lo largo de los siglos filósofos, neurocientíficos y psicólogos han desarrollado teorías acerca de la consciencia humana. A pesar de este gran esfuerzo histórico en la búsqueda de una explicación para la consciencia natural, se ha realizado relativamente poco esfuerzo durante las últimas décadas en el campo correspondiente de la inteligencia artificial. Los avances científicos logrados en el estudio de la consciencia durante los 80, que en gran medida aún siguen vigentes, han tenido una influencia modesta en los sistemas artificiales bio-inspirados. A menudo, la consciencia se define basándose en la relación existente en los humanos entre los siguientes procesos mentales: atención, razonamiento, reconocimiento y comportamiento (Kozma, 1997). Es decir, un ser consciente presenta la capacidad de atención hacia una cosa, y puede pensar acerca de ella, qué es, cómo es, por qué es así, etc., con el objetivo de reconocerla. Una vez que el objeto se identifica, el sujeto lo ha reconocido y entonces decide qué quiere hacer con él. Se dice que en los humanos todos estos mecanismos tienen lugar de forma consciente, son contenidos conscientes en nuestra mente. El paradigma de la consciencia artificial se inspira en estos procesos observados en los humanos y

otros mamíferos superiores con el objetivo de conseguir sistemas artificiales que presenten capacidades y funcionalidades análogas a las naturales.

Un problema clave en el proceso de modelado computacional de las teorías actuales viene determinado directamente de la propia naturaleza de estas teorías. Algunos de los paradigmas que se aplican para explicar los procesos conscientes, como la mecánica cuántica (Hameroff y Penrose, 1996), los efectos relativísticos (Rakovic, 1990), o la sincronización de las activaciones neuronales (Crick y Koch, 1990) son prácticamente imposibles de representar mediante un modelo plausible con agentes, o cualquier otra técnica software. Sin embargo, como explicamos más adelante, existen dimensiones de la consciencia que pueden explicarse con otro tipo de teorías que se prestan mejor a su aplicación y experimentación en sistemas artificiales. Es la dimensión fenomenológica de la consciencia la más incierta en los estudios científicos.

Existen teorías de la consciencia y la atención basadas en parte en aspectos funcionales cognitivos (Baars, 1988; Dennett, 1991; Searle, 1992; Block, 1995; Chalmers, 1997; Sun 2002), las cuales se podrían implementar de forma pragmática por medio de agentes software. Conviene, en cualquier caso, hacer una diferenciación inicial entre dos grandes dimensiones de los procesos conscientes. De esta forma se pueden establecer más claramente los aspectos de la consciencia que se pretenden emular en los sistemas artificiales, sin entrar en el vasto campo de la consciencia humana en toda su extensión e implicaciones. Por supuesto, la tarea de diferenciar entre tipos, clasificaciones y niveles de consciencia no es en absoluto trivial. De hecho, existen numerosas divisiones y clasificaciones (Edelman, 1992; Panksepp, 2005). Se habla de consciencia primaria, autoconsciencia, consciencia afectiva, propiocepción, intersubjetividad, etc. No obstante, existe una clasificación de más alto nivel que es útil para distinguir básicamente entre lo que actualmente podemos aspirar a reproducir en una máquina y lo que no. Se trata de diferenciar entre la consciencia fenoménica (CF) o intransitiva y la consciencia de acceso (CA) o proposicional (Villanueva, 2003). Muchos autores (Block, 1995; Chalmers, 1997; Sugiyama, 2000) sostienen que se puede distinguir entre estos dos tipos de consciencia, o el uso del término consciencia en dos situaciones diferentes. La CF es un tipo de experiencia subjetiva que el sujeto tiene por el hecho de ser consciente, mientras que la CA es de alguna manera la disponibilidad para el uso del razonamiento y la guía de las acciones y el habla. Por lo tanto, se distinguen dos formas de ver la consciencia, por un lado un sujeto es consciente cuando presta atención a un objeto del exterior, conociéndolo y comprendiéndolo mientras éste es foco de su atención; por otro lado, el sujeto puede percibir y sentir su propio interior al tener una experiencia consciente.

Los aspectos de acceso de la consciencia son muy interesantes en cuanto a su posible aplicación en sistemas artificiales. Por otro lado, los aspectos fenoménicos, cuyas características son de una naturaleza poco comprendida hasta el momento, se consideran fuera del ámbito del presente análisis. La consciencia puede ser considerada como una pasarela que proporciona acceso a casi cualquier contenido de la mente (Baars, 1988). Es en definitiva a lo que cada sujeto puede tener acceso sobre si mismo. En un momento dado hay un gran número de procesos neuronales inconscientes ejecutándose en paralelo; sin embargo, sólo ciertos contenidos se muestran a la consciencia en cada

momento. Es decir, la atención establece los contenidos de la mente que se perciben conscientemente. Por lo tanto, una de las características principales de los procesos conscientes, en relación a los procesos inconscientes, es que los primeros son mucho más limitados. Los mecanismos conscientes se basan en la memoria a corto plazo y la selección del foco de atención. Estos aspectos son claramente limitados, en el sentido de que no se pueden realizar simultáneamente dos acciones voluntarias (prestar atención a dos cosas a la vez) y la memoria de trabajo que se utiliza no puede manejar más de aproximadamente siete elementos separados al mismo tiempo, por ejemplo, números de teléfono (Miller, 1956).

Los conceptos vistos anteriormente se pueden emplear para crear modelos computacionales más eficientes, inspirados en el funcionamiento de la consciencia en los mamíferos superiores. En los siguientes apartados pretendemos presentar un breve repaso de las teorías y los modelos cognitivos de la consciencia más importantes, y su estado del arte en cuanto a la implementación y experimentación usando sistemas de agentes.

Principales teorías de la consciencia

Existen multitud de teorías sobre el funcionamiento, la evolución, la función y las características de la consciencia. Muchos distinguidos científicos y filósofos, como Nagel, Jackson, McGinn, Damasio, Crick, Dennett, Edelman, etc. han abordado el tema desde diferentes perspectivas. Todas ellas muy interesantes. Sin embargo, en el contexto del presente análisis, y desde un punto de vista pragmático centrado en la aplicación a modelos computacionales, nos centraremos en los trabajos basados en explicar los procesos cognitivos de la consciencia a nivel funcional. En definitiva, como hemos introducido anteriormente, se trata de comprender como se gestiona el acceso al conocimiento y el control de un vasto conjunto de complejos procesos paralelos (inconscientes) desde un único hilo secuencial (consciente). A continuación se describen algunas de las teorías más destacadas, aunque existen muchas más.

La dualidad de la mente

La mayoría de las teorías sobre la consciencia consideran que en la mente existen dos tipos distintos de procesos: conscientes e inconscientes. Desde el punto de vista de los contenidos con los que operan estos procesos, la dualidad se expresa en base a las diferentes formas de representar y procesar el conocimiento. Dependiendo de la naturaleza consciente o inconsciente de los procesos el conocimiento que utilizan puede ser declarativo o procedimental, localizado o distribuido, procesado en serie o en paralelo. Aunque también hay autores que defienden una naturaleza unitaria de lo consciente e inconsciente, como (Dennett, 1991), no existen desarrollos posteriores notorios que hayan desembocado en modelos computacionales. Las hipótesis que consideran la separación entre consciencia e inconsciencia, tienen que plantear los criterios de separación entre ambos dominios, así como el funcionamiento característicos de cada uno de ellos. Por ejemplo, Rosenbloom y Newell (1993) diseñan una arquitectura en la

que las tareas se realizan por encadenamiento de diferentes bloques funcionales. Cada bloque es una representación unitaria con un funcionamiento opaco, aunque sus entradas y sus salidas sí son accesibles. La consciencia se produce cuando una tarea se realiza por intervención de múltiples bloques. Si interviene un solo bloque es un proceso inconsciente. Otro criterio establecido por Mathis y Mozer (1996) es que la consciencia se caracteriza por estados temporalmente estables en una red de módulos computacionales especializados. Un criterio básico para la diferenciación entre procesos conscientes e inconscientes es la forma de acceso al conocimiento (Hadley, 1995; Clark y Toribio, 1992). La representación del conocimiento puede ser implícita o explícita. Los procesos conscientes usan información explícita directamente accesible, mientras que los procesos inconscientes manejan información implícita que no es accesible si no es a través de mecanismos interpretativos.

Sun (2002) argumenta que los procesos cognitivos se estructuran en dos niveles con mecanismos diferentes. Cada nivel codifica un conjunto completo de conocimiento para su procesamiento. Estos dos conjuntos de conocimiento se solapan en gran medida, por lo que los resultados de ambos han de combinarse. Según Sun se produce una sinergia entre el procesamiento implícito (inconsciente) y el procesamiento explícito (consciente). En algunos trabajos sobre la consciencia en robots, como por ejemplo (Kubota et al. 2001) se distingue entre comportamiento automático inconsciente, y comportamiento consciente, que requiere que el individuo tenga cierto grado de autoconsciencia. Denominan autoconsciencia a la consciencia que se tiene de uno mismo y su situación en el mundo. Estos autores establecen dos suposiciones acerca de la autoconsciencia: se activa cuando se produce un cambio relativamente grande en la información sensorial (percepción) y cuando la predicción acerca de la información sensorial es diferente de lo que en realidad se percibe.

La coherencia de la consciencia

Otro tipo de teorías se basan en el concepto de coherencia o coalición. En el caso de (Baars, 1988) una serie de procesadores especializados inconscientes proporcionan información a un espacio de trabajo común, que coordina a los procesadores mediante la selección de patrones coherentes de información (por su valor ilustrativo, analizaremos en más detalle el modelo de Baars a continuación). La noción de coherencia se emplea también a otros niveles, por ejemplo para denotar la activación neuronal sincronizada. En su búsqueda de las correlaciones neuronales de la consciencia, Crick y Koch (1990) llegaron a argumentar que la activación sincronizada a 40 Hz de coaliciones de neuronas era la base física de la consciencia. Aunque más tarde se retractaron al comprobar que estas activaciones entre 35 y 75 Hz en el cortex cerebral no tenían que estar necesariamente relacionadas con los procesos conscientes. También Damasio et al. (1990) hablan de coherencia en otro sentido similar. La reverberación en zonas neuronales de convergencia sensorial integra la información de cada sentido. A su vez, toda la información procedente de cada sentido se integra en una única zona de convergencia multimodal que daría lugar a los contenidos conscientes. De forma similar Schacter plantea la teoría de que múltiples módulos especializados mandan in-

formación a un único módulo consciente. En (Schacter et al. 2002) se analizan evidencias de la disociación de los diferentes tipos de conocimiento en el cerebro.

El teatro de la consciencia

Baars (1997) habla de un "teatro", en el que el foco de la consciencia se representa por el punto de luz sobre el escenario, que es dirigido por la atención. El escenario completo se corresponde con la memoria de trabajo, que es el sistema de memoria que almacena los contenidos conscientes. La información obtenida en el punto de luz se distribuye de forma global a través del teatro a dos clases de procesadores inconscientes: los que forman la audiencia reciben información del foco de luz; mientras, entre bastidores, los sistemas inconscientes contextuales dan forma a los sucesos que ocurren en el punto de luz. La metáfora del foco luminoso es también utilizada por Crick (1994) argumentando, acerca del procesamiento de la información visual, que fuera del punto de luz de la atención visual la información se procesa menos, de forma diferente o ni siquiera se procesa.

No hay que confundir esta metáfora del teatro que usa Baars con otra metáfora denominada "Teatro Cartesiano", que es en esencia opuesta a la defendida por Baars, ya que atribuye la consciencia a un punto concreto del cerebro, la glándula Pineal. Descartes pensaba que en esta glándula se localizaba el alma (Finger, 1995). Las teorías como esta que localizan la consciencia en un punto concreto del cerebro son mayoritariamente rechazadas por la comunidad científica. Si bien es cierto que los neurocientíficos buscan las correlaciones neuronales de la consciencia, no se cree que se localicen en un punto concreto, sino que posiblemente se formen a partir de coaliciones de neuronas (Crick y Koch, 2003).

Volviendo a la metáfora del teatro desarrollada por Baars, es importante resaltar que el "escenario" está compuesto por la memoria de trabajo. Donde los "actores" compiten por aparecer en el foco luminoso de la atención, en el cual aparecen como contenidos completamente conscientes. La selección del foco de atención se realiza en gran medida entre bastidores. Son los procesadores inconscientes los que llevan a cabo esta selección en base al contexto y a conjuntos de creencias (a menudo inconscientes) que determinan los pensamientos conscientes (la actuación en escena). Baars también indica que el foco luminoso de la consciencia es el instrumento que usa el "director" para tomar decisiones en el campo de la memoria de trabajo guiadas por la persecución de metas. Este director de la obra, también trabaja entre bastidores, lo que sugiere que en gran medida no tenemos acceso a las razones por las que hacemos las cosas. Este concepto encaja con el presentado por algunos autores (Rosenthal, 2000; Morin, 2002), que afirman que el *yo* consciente confabula para deducir las razones por las que el sujeto lleva a cabo sus acciones.

Según Baars, al vasto dominio inconsciente de conocimiento y control se puede acceder usando la consciencia. La consciencia se usa para el aprendizaje rápido y el reconocimiento preciso. También activa un gran número de rutinas automáticas que constituyen acciones específicas, proporcionando coordinación y control. Las experiencias

conscientes activan contextos inconscientes, que ayudan a interpretar sucesos conscientes futuros. En definitiva, la consciencia proporciona un marco para el acceso (función de búsqueda global) a los vastos contenidos inconscientes de la mente. Parece que las investigaciones realizadas con métodos de diagnóstico por imágenes (resonancia magnética funcional, tomografía por emisión de positrones, etc.) indican que esta hipótesis podría ser cierta (Baars, 2002; Baars et al., 2003); en cualquier caso, se necesitan más análisis neurológicos para confirmar o desmentir con seguridad las suposiciones de Baars.

La dimensión sentimental

Los sentimientos son el balance consciente de nuestra situación (Marina, 2002). Según diferentes teorías psicológicas, los sentimientos son la forma en que los seres humanos son capaces de sintetizar su situación en el mundo dentro del ámbito limitado de la consciencia. El comportamiento está condicionado por el estado emocional del individuo. Como se indica en (Franklin et al. 1998) los sentimientos o emociones proporcionan una valoración de lo bien que se están cumpliendo los objetivos del sistema. Un resumen extraordinariamente simplificado, sin entrar en el complejo mundo de los sentimientos, sería el siguiente: el sujeto sentiría alegría en caso de ver que sus objetivos se van cumpliendo de la forma prevista. En caso contrario, se sentiría frustrado. En los humanos los sentimientos influyen en la conducta, entre otras cosas, en base a un sistema de creencias. Por eso, bajo un estado de frustración unos individuos actúan desistiendo de sus objetivos originales por completo, mientras que otros optarán por intentar diferentes alternativas.

Aspectos funcionales de la consciencia

De las teorías sobre la consciencia analizadas hemos tratado de extraer una serie de funciones que creemos son la base de lo que hemos denominado consciencia de acceso (CA). En definitiva se trata de separar la parte funcional de la parte fenomenológica de la consciencia, e identificar en la primera las funciones y características que convierten a la consciencia en una ventaja evolutiva para los seres que la poseen. Hemos identificado las siguientes funciones básicas: (1) atención, (2) balance de situación, (3) búsqueda global, (4) Procesamiento de conocimiento implícito y explícito, (5) contextualización, (6) predicción sensorial, (7) memorización modal y multimodal y (8) autocoordinación. Nuestro planteamiento es que estas funciones deben estar integradas en un sistema de consciencia artificial para que éste presente las ventajas esperadas. En esta lista no hemos incluido otras funciones que se suponen necesarias en un sistema inteligente, pero que no están directamente relacionadas con la consciencia. Como por ejemplo, sistemas de creencias, razonamiento, reconocimiento, percepción, etc. Los detalles sobre la integración de los mecanismos de consciencia artificial con otros paradigmas de la inteligencia artificial están fuera del ámbito del presente análisis. A continuación se describen en más detalle las funciones de la consciencia identificadas:

(1) El mecanismo de atención proporciona al sujeto la capacidad de "prestar atención" a un determinado suceso u objeto, y de esta manera dirigir su aprendizaje y comportamiento.
(2) El balance de situación se refiere a que el sujeto sea capaz de mantener un resumen consciente de su estado. Los sentimientos juegan este papel.
(3) La capacidad de búsqueda global implica acceso a prácticamente todo el conocimiento que posee el sujeto. Esta función es necesaria para el acceso a las rutinas inconscientes de control y las diferentes memorias.
(4) La separación entre procesos conscientes e inconscientes ha de basarse en la diferenciación entre conocimiento explícito e implícito respectivamente. En cada uno de estos dominios debe existir capacidad de aprendizaje. Es decir, aprendizaje implícito inconsciente y aprendizaje explícito consciente. Ambos dominios deben coordinarse a través de mecanismos de control y acceso, como la atención y la búsqueda global.
(5) La contextualización es necesaria para el reclutamiento de procesadores inconscientes adecuados por parte del control consciente. Asimismo, para la resolución de problemas es necesario localizar el conocimiento relacionado con la cuestión que hay que resolver. La memoria asociativa es parte de los mecanismos de contextualización.
(6) La predicción sensorial se refiere a un constante proceso de monitorización y predicción inconsciente de la información obtenida por los sentidos. Cuando lo percibido es distinto de lo esperado, la información correspondiente debe hacerse consciente para poder tratar una situación imprevista.
(7) La memoria multimodal se corresponde con la memoria semántica y de trabajo, donde convergen temporalmente todos los contenidos conscientes. Las memorias modales mantienen indefinidamente contenidos de un tipo específico (por ejemplo, la memoria visual).
(8) La autocoordinación sustituiría en un sistema artificial al libre albedrío. Es decir, se encarga de coordinar las acciones para la consecución de las metas establecidas. Mecanismos como el habla interior y la introspección se incluyen en esta función como elementos de gestión de proyectos (entendiendo por proyecto el conjunto de tareas que se realizan para la consecución de una o varias metas).

Existen diversas sinergias entre las funciones descritas anteriormente, que en su conjunto dan lugar a lo podríamos denominar consciencia artificial. Con respecto a la función 4, hay que remarcar que las novedades requieren mayor participación de la consciencia para su aprendizaje. Es decir, interrelación con la función 1 y 6. Las funciones 5 y 6 representan el flujo de control de arriba abajo y de abajo arriba respectivamente. Por un lado la voluntad consciente invoca procesamientos inconscientes para llevar a cabo sus metas; por otro lado, los procesadores inconscientes que integran la información de los sentidos "llaman la atención" consciente en caso de encontrarse con una situación inesperada o novedosa. Creemos que las propiedades de coherencia o coalición de procesos expresadas por varias teorías se cubren con la función 5, ya que la asociación de procesadores constituye un tipo de coherencia a nivel funcional. La función 7 tiene que dar soporte a la "historia personal del sujeto", que es un aspecto de la consciencia indicado por varias teorías. Este concepto proporciona la unidad

necesaria que permite al individuo gestionar su propia experiencia e identidad. El mecanismo de coordinación indicado en la función 8 se relaciona con la función 2 (los sentimientos), ya que la selección de metas y acciones estará condicionada por el estado emocional.

Modelos de consciencia implementados con agentes software

Además del caso de las redes de neuronas artificiales, los sistemas multiagente parecen particularmente buenos candidatos para implementar modelos de consciencia porque presentan similitud con el estilo de funcionamiento del cerebro, en el que el trabajo se realiza de forma distribuida por grupos de neuronas especializadas, sin que exista un centro específico de control. Existen diversos sistemas artificiales bioinspirados en los mecanismos de la consciencia, por ejemplo: *Unified Model of Attention* (Hunt y Lansman, 1986), *ACT* (Anderson, 1993), *ACT-R* (Anderson, 1996), *IDA* (Franklin et al., 1998), *Reflection* (Sugiyama, 2000), *CLARION* (Sun, 1997; Sun, 2002), *CODAM* (Taylor, 2003), *Computational Agent Framework for Consciousness* (Moura y Bonzon, 2004), *SOAR* (Laird et al. 1987; Lehman et al. 2006).

De todos estos modelos, sólo IDA (*Intelligent Distribution Agent*), SOAR y CAFC (*Computacional Agent Framework for Consciousnes*) están implementados mediante sistemas multiagente. Los demás están basados en otro tipo de arquitecturas, como reglas de producción, redes semánticas, redes de neuronas, etc. Como hemos visto, una de las teorías de la consciencia más significativas en el marco de su posible aplicación a los sistemas multiagente es la teoría del Espacio de Trabajo Global (GWT - *Global Workspace Theory*) (Baars, 1988; Baars, 1997). De hecho, tanto IDA como CAFC están basados en esta teoría. Ambos modelos computacionales aplican la hipótesis de Baars acerca de que la consciencia es un suceso global que se produce en partes distribuidas del cerebro, lo cual encaja bien con el concepto de sistema multiagente. Agentes inteligentes independientes se envían mensajes unos a otros a través de un espacio de trabajo común. En este entorno la experiencia consciente emerge de la cooperación y la competición. IDA no es un sistema de propósito general sino que está específicamente diseñado para la optimización en la asignación de tareas a soldados de la marina de los EEUU. Por lo tanto la experimentación está limitada a la interacción que permiten sus interfaces específicos. CACF es potencialmente de propósito general, al igual que SOAR, pero implementa un modelo más limitado de consciencia. Por otro lado, SOAR no implementa directamente un modelo de consciencia, sino que está basado en modelos cognitivos clásicos. Considera los conceptos de meta, estado y operador para manejar el conocimiento.

Evaluación de las implementaciones de consciencia artificial

La consciencia en si misma se puede analizar desde muchas perspectivas, y consecuentemente los sistemas de consciencia artificial se pueden evaluar de formas muy diferentes dependiendo de los factores que se consideren relevantes. Básicamente, la

ventaja principal que teóricamente se puede obtener por el hecho de aplicar un modelo de consciencia, es el que sistema artificial sea capaz de manejar mejor situaciones nuevas y problemáticas (no esperadas). Uno de los objetivos del presente análisis es dilucidar si la presencia de las funciones de la consciencia mencionadas anteriormente influyen positivamente en el rendimiento del sistema, tal y como se predice en las teorías de la consciencia. Planteamos un método de evaluación basado en dos partes fundamentales:

- En primer lugar, para cada implementación considerada pretendemos determinar hasta que punto reproduce ésta la correspondiente teoría, analizando las posibles deficiencias. Concretamente, se analizan las funciones de la consciencia que se implementan en cada caso (sean o no consideradas por la teoría que inspira el modelo). Para ello utilizamos la lista de funciones que hemos identificado como clave para un sistema consciente.
- En segundo lugar, se analiza el rendimiento de las implementaciones de los modelos de consciencia. Este análisis debe estar orientado especialmente a determinar la capacidad de los sistemas a enfrentarse con situaciones inesperadas y aprender de las mismas.

La siguiente tabla resume la comparativa de los sistemas multiagente exclusivamente en base a la implementación de las funciones que consideramos clave en un sistema artificial consciente. Queda de manifiesto que ningún sistema abarca todas las funciones.

Función Implementada	**IDA**	**CAFC**	**SOAR**
(1) Atención	Sí	Sí	No
(2) Balance de situación	Sí	No	No
(3) Búsqueda global	Sí	Sí	Sí
(4) Procesamiento de conocimiento implícito y explícito	Sí	Sí	No
(5) Contextualización	Sí	Sí	Sí
(6) Predicción sensorial	No	No	No
(7) Memorización modal y multimodal	Sí	Sí	Sí
(8) Autocoordinación	No	No	No

Tabla 1. Comparación de implementaciones de modelos de consciencia con agentes.

En CAFC se consideran los conceptos de plan y condición, pero no existe un módulo de coordinación como tal que dirija la construcción de planes. Particularmente, como no existe la función de balance de situación, no es posible dirigir los planes de acuerdo al progreso que se está realizando. Hay que evaluar frente a diferentes problemas, para comprobar efectivamente la robustez respecto a problemas y situaciones nuevas así como la flexibilidad del sistema. Esto no se analiza enfrentándose siempre al mis-

mo dominio de problemas. Es necesario realizar una correlación entre funciones, grupos de funciones, y su impacto en el rendimiento.

Conclusiones

Las hipótesis barajadas en el presente artículo se basan principalmente en metáforas que simplemente ayudan a comprender de forma holística el funcionamiento de la mente humana. Si bien es cierto que una simple metáfora está muy lejos de constituir un cuerpo establecido de conocimiento científico, puede servir como herramienta para dirigir las investigaciones en diversos sentidos, que afirmen o desmienten las hipótesis planteadas. Desde el punto de vista de la implementación de modelos computacionales, el uso de estos esquemas simplificados de la consciencia tiene dos ventajas claras: la facilidad de implementación y comprensión del modelo, y por ende la posibilidad de experimentación con sistemas artificiales fácilmente observables, parametrizables y relativamente asequibles. Por supuesto en ningún caso la experimentación con sistemas artificiales puede en modo alguno sustituir a la experimentación con los verdaderos poseedores de consciencia natural. Sin embargo, como sabemos, la inteligencia artificial ha ofrecido retroalimentación útil durante los últimos 50 años a la psicología y viceversa, completando y mejorando los modelos de la mente en base a los resultados obtenidos en ambos dominios. Pensamos que en el campo de la consciencia, que no es más que una parte que ha de integrarse con las teorías existentes de la mente, ha de suceder lo mismo. Un modelo de consciencia no es suficiente, ya que la consciencia es un aspecto fundamental, pero no el único. Se requiere un modelo completo de la mente. En cualquier caso, conocer como funciona la consciencia y sus funciones asociadas es un buen comienzo para encajar el resto de piezas del puzzle. Aunque las teorías analizadas cubren un amplio rango de funcionalidades, uno de los conceptos que echamos en falta es el concepto de proyecto. Creemos que esta noción, entendida como la asociación de metas orientadas a conseguir un objetivo final, debe ser considerada como parte de la funcionalidad de autocoordinación.

La consciencia puede ser la forma que el sistema nervioso ha desarrollado evolutivamente para lidiar con sucesos novedosos e inesperados en el mundo (Franklin, 2005). Esta concepción supera los antiguos sistemas situacionales, como por ejemplo la arquitectura de subsunción (Brooks, 1990), en los que sin necesidad de controlar estados o conocimientos internos, un agente autónomo puede desenvolverse de manera exclusivamente reactiva. Un ejemplo de las carencias de este tipo de sistemas es (Arrabales et al., 2002), donde la necesidad de un mecanismo de atención se hace patente. El método de evaluación planteado, aunque se trata de una primera aproximación, proporciona medidas heurísticas acerca de que funciones de la consciencia son clave en el aumento de rendimiento (entendiendo por rendimiento, la mejor capacidad adaptativa del sistema). Ha de tenerse en cuenta la imposibilidad de realizar este tipo de evaluaciones con sistemas naturales. Por razones obvias no se pueden añadir y quitar funciones mentales a un sujeto natural (ya sea un ser humano u otro animal). Experimentar con sistemas diferentes también es un problema. Habría que probar con el mismo sistema añadiendo y quitando componentes (funciones). Esto se plantea como

un trabajo futuro: un sistema modular (tipo banco de pruebas) con el que se pueda evaluar mejor la aportación de todas y cada una de las funciones de la consciencia (y la correspondiente integración y sinergias entre diferentes funciones).

Referencias

ANDERSON, J.R. (1993). *Rules of Mind*. Hillsdale. Lawrence Erlbaum.

ANDERSON, J.R. (1996). *The Architecture of Cognition*. Lawrence Erlbaum.

ARRABALES, R. FLANAGAN, C. y TOAL, D. (2002). *An Adaptive Video Event Mining System for an Autonomous Underwater Vehicle*. Intelligent Engineering Systems through Artificial Neural Networks, Vol. 12, pp. 585-591. ASME Press.

BAARS, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

BAARS, B.J. (1997). *In the Theater of Consciousness. Global Workspace Theory, A Rigorous Scientific Theory of Consciousness*. Journal of Consciousness Studies, 4, No. 4, 1997, pp. 292-309.

BAARS, B.J. (2002). *The conscious access hypothesis: origins and recent evidence*. Trends in Cognitive Sciences, Vol. 6 No. 1 pp. 47-52.

BAARS, B.J. RAMSOY, T.Z. y LAUREYS, S. (2003). *Brain, conscious experience and the observing self*. Trends in Neurosciences. Vol. 26, No. 12, pp. 671-675.

BLOCK. N. (1995). *On a Confusion about a Function of Consciousness*. Behavioral and Brain Sciences 18, 227-87.

BROOKS, R.A. (1990). *Elephants Don't Play Chess*. Designing Autonomous Agents. MIT Press.

CHALMERS. D. (1997). *Availability: The Cognitive Basis of Experience*. The Nature of Consciousness, Edited by Block. N., Flanagan. O. and Guzeldere. G., MIT Press.

CLARK, A. y TORIBIO, J. (1994). *Doing without Representing?* Synthese, vol. 101, No. 3. Springer Science.

CRICK, F. (1994). *Astonishing Hypothesis: The Scientific Search for the Soul*. Scribner Book Company.

CRICK, F. y KOCH, C. (1990). *Toward a neurobiological theory of consciousness*. Seminars in the Neurosciences 2:263-275.

CRICK, F. y KOCH, C. (2003). *A framework for consciousness.* Nature Neuroscience, 6:119-126.

DAMASIO, A.R. DAMASIO, H. TRANEL, D. y BRANDT, J.P. (1990). *Neural Regionalization of knowledge access: preliminary evidence*. 55, pp. 1039-1047. Cold Spring Harbor Symposia on Quantitative Biology.

DENNETT, D.C. (1991). *Consciousness Explained*. Penguin.

EDELMAN, G.M. (1992). *Bright Air, Brilliant Fire. On the Matter of the Mind*. Basic Books.

FINGER, S. (1995). *Descartes and the pineal gland in animals: a frequent misinterpretation*. Journal of the history of the neurosciences. Sep-Dec; 4(3-4): 166-82.

FRANKLIN, S. (2005). *Evolutionary pressures and a stable world for animals and robots: A commentary on Merker*. Consciousness and Cognition 14, pp. 115-118.

FRANKLIN, S. KELEMEN, A. y McCAULEY, L. (1998). *IDA: A Cognitive Agent Architecture*. IEEE Conference on Systems, Man and Cybernetics. IEEE Press.

HADLEY. R.F. (1995). *The "explicit-implicit" distinction*. Minds and Machines, Vol. 5, No. 2. Springer Science.

HAMEROFF, S.R. y PENROSE, R. (1996). *Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness*. Toward a Science of Consciousness - Contributions from the Tucson Conference, MIT Press, Cambridge, MA.

HUNT, E. y LANSMAN, M. (1986). *Unified Model of Attention and Problem Solving*. Psychological Review, 4. pp. 446-461.

KOZMA, R. (1997). *On the conscious and subconscious components of knowledge representation in neural networks*. International Conference on Neural Networks. Vol. 4, pp. 2519-2523.

LAIRD, J.E. NEWELL, A. y ROSENBLOOM, P.S. (1987). *SOAR: an architecture for general intelligence.* Artificial Intelligence, vol. 33, pp. 1-64. Elsevier Science.

LEHMAN, J.F. LAIRD, J. ROSENBLOOM, P. (2006). *A Gentle Introduction to Soar, an Architecture for Human Cognition: 2006 Update*. Actualización del original Sternberg y Scarborough (1996).

MARINA, J.A. (2002). *El laberinto sentimental*. Editorial Anagrama.

MATHIS, D.W. y MOZER, M.C. (1996). *Conscious and Unconscious Perception: A Computational Theory*. Proceedings of the Eighteenth Conference of the Cognitive Science Society, pp. 324-328. Erlbaum.

MILLER, G.A. (1956). *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*. The Psychological Review, vol. 63, pp. 81-97.

MORIN, A. (2002). *Do you "self-reflect" or "self-ruminate"?* Science and Consciousness Review. Dec. No. 1.

MOURA, I. y BONZON, P. (2004). *A Computational Framework for Implementing Baars' Global Workspace Theory of Consciousness.* Brain Inspired Cognitive Systems.

PANKSEPP, J. (2005). *Affective Consciousness: Core emotional feelings in animals and humans*. Consciousness and Cognition 14 30-80.

RAKOVIC, D. (1990). *Neural Networks Vs. Brain Waves: Prospects for Cognitive Theory of Consciousness*. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 12, No. 3.

ROSENBLOOM, P.S. y NEWELL, A. (1993). *The chunking of goal hierarchies: a generalized model of practice*. The SOAR papers, vol. 1. MIT Press Series in Research In Integrated Intelligence.

ROSENTHAL, D.M. (2000). *Consciousness, Content, and Metacognitive Judgements.* Consciousness and Cognition 9, pp. 203-214

SCHACTER, D.L. REIMAN, E. UECKER, A. ROISTER, M.R. YUN, L.S. y COOPER, L.A. (2002). *Brain regions associated with retrieval of structurally coherent visual information*. Nature, 376, pp. 587-590.

SEARLE, J. (1992). *The rediscovery of the Mind.* The MIT Press.

SUGIYAMA, S. (2000). *Reflected Method for Having Consciousness*. IEEE International Conference on Systems, Man, and Cybernetics, 2000. Volume 5, 8-11 Oct. 2000 Page(s):3141 - 3146 vol.5.

SUN, R. (1997). *Learning, Action and Consciousness: A Hybrid Approach Toward Modelling Consciousness*. Neural Networks, Vol. 10, No. 7, pp. 1317-1331. Elsevier Science Ltd. 0893-6080/97.

SUN, R. (1999). *Computational Models of Consciousness: An Evaluation.* Journal of Intelligent Systems, 9. pp. 507-562.

SUN, R. (2002). *Duality of the mind. A bottom up approach toward cognition*. Lawrence Erlbaum Associates Publishers.

TAYLOR, J. (2003). *The CODAM model of Attention and Consciousness.* Proceedings of the IEEE International Joint Conference on Neural Networks, 2003, July 20-24.

TAYLOR, J.G. (1994). *The relational mind*. From Perception to Action Conference. IEEE.

VILLANUEVA, E. (2003). *¿Qué son las propiedades psicológicas? Metafísica de la psicología*. Instituto de Investigaciones Jurídicas. Universidad Nacional Autónoma de México.