Revista Latinoamericana de Psicología
2006, volumen 38, Nº 1, 137-148

EL PROYECTO COMPUTACIONAL CONEXIONISTA EN EL ANÁLISIS TEÓRICO DE LA ACTIVIDAD CONCIENTE*

FABIO ENRIQUE MARTÍNEZ
Universidad Nacional de Colombia

ABSTRACT

Consciousness, that experienced flow of subjective states, is one of the mysteries, and perhaps, the fundamental challenge of science until now. It is also a field of exploration specially active and fruitful, a field that has passed over the frontier of XIX and XX centuries, and recently arrived again with a strong impetus in the XXI century. However, there is a great controversy about the plausibility of a theoretical, analytical and formal (v.g.: computational) explanation of the phenomena that we associate with consciousness. Is it possible to establish a reductionist explanation of consciousness? Or in other words, is it possible to make a description of the conscious phenomena expressed in terms of functional and/or causal relationships? In this article I give some relevant elements to sketch the sufficiency of explanation of the connectionist computational paradigm, and how we could elucidate the formal principles embedded in the study of consciousness. The purpose of the present article is to suggest that the plausibility of the connectionist paradigm is supported by the following issues: (1) the level of fine-grained detail with which we define the representation and computability of conscious states, (2) the methodological and conceptual advances of brain sciences, and (3) the difference that we assume between the notions of simulation, modelling and computational representation of consciousness. With these ideas in mind, through the manuscript I will show a basic framework to understand why connectionism can be a plausible candidate to think about a formal theory of consciousness. Finally, in the light of the previous statements, I will point out some important issues to discuss the plausibility of a computational theory of consciousness.

Key words: consciousness, connectionism, computational representation, neural coding, formal theories of consciousness.

---

* El editor del presente artículo fue Andrés M. Pérez-Acosta, Editor Asociado.

1 Correspondencia: FABIO MARTÍNEZ. *E-mail:* faviolin@yahoo.com

RESUMEN

La actividad conciente, ese devenir que experimentamos como una serie de estados de subjetividad, es uno de los misterios, y quizás el desafío fundamental de la ciencia contemporánea. Es también un campo de exploración especialmente activo y fructífero que sobrevivió la transición entre los siglos XIX y XX, y nuevamente ha tomado un fuerte impulso ahora en el XXI. Existe, sin embargo, una gran controversia sobre la plausibilidad de una explicación teórica, analítica y formal (v.g.: computacional) de los fenómenos que asociamos a la actividad conciente. ¿Es factible formular una explicación reduccionista de la actividad conciente, es decir, una descripción del fenómeno expresada en términos de relaciones funcionales y/o causales? Los párrafos del presente ensayo escudriñan algunos elementos relevantes a fin de establecer la suficiencia explicativa que tiene el paradigma computacional conexionista para dilucidar los principios formales imbricados en el estudio de la actividad conciente. El propósito que subyace la elaboración siguiente es sugerir que la viabilidad del conexionismo y del proyecto computacional depende de los siguientes aspectos: (1) El grado de refinamiento con el que se defina la representación y la computabilidad de los estados concientes, (2) Los avances metodológicos y conceptuales de las ciencias del cerebro, y (3) La distinción que se haga entre simulación, modelamiento y representación computacional de la actividad conciente. Con estas ideas en mente, a través del capítulo mostraré un esquema conceptual básico para entender por qué el paradigma conexionista puede ser un candidato plausible para pensar una teoría formal de la actividad conciente. Finalmente, a la luz de los planteamientos presentados, señalaré algunos aspectos importantes para establecer la viabilidad de una teoría computacional de la actividad conciente.

Palabras clave: actividad conciente, conexionismo, representación computacional, codificación neuronal, teorías formales de la conciencia.

INTRODUCCIÓN

¿Es factible formular una explicación reduccionista de la actividad conciente, es decir, una descripción del fenómeno expresada en términos de relaciones funcionales y/o causales? Con el fin de responder a esta pregunta y establecer la suficiencia explicativa del paradigma computacional conexionista, el presente ensayo se ha organizado de la siguiente manera: en primer lugar expondré algunos elementos importantes que denotan el contexto de la definición de 'conciencia' que es asumido a lo largo del escrito. Luego mostraré una breve reseña de la controversia que proponen John Searle y David Chalmers con respecto al problema fundamental de la actividad conciente que debe enfrentar toda teoría de índole formal-computacional. Después presentaré las nociones de computabilidad y representación computacional (simbólica y subsimbólica) que retoma el conexionismo. De esta manera estableceré un esquema conceptual básico para entender por qué este paradigma puede ser un candidato plausible para pensar una teoría formal de la actividad conciente. En el siguiente apartado discutiré el estatus de la representación neuronal de los contenidos de la actividad conciente. Consideraré acto seguido uno de los problemas más importantes ligados a esta perspectiva de representación: el problema de la integración de los contenidos de conciencia. Finalmente, a la luz de los planteamientos presentados, señalaré algunos aspectos importantes para establecer la viabilidad de una teoría computacional de la actividad conciente.

Aspectos fundamentales de la actividad conciente

Antes de presentar los puntos relevantes para la discusión aquí planteada, es conveniente primero plantear algunos criterios para definir la conciencia[2]. Este vocablo, del latín 'conscientia', está asociado a tres significados diferentes pero íntimamente relacionados: (1) cortar o escindir, (2) hacer una distinción, y (3) conocer. Basándonos únicamente en esta definición etimológica parece válido afirmar, que aun el organismo unicelular más simple tendría una forma de conciencia, dado que puede establecer una distinción entre su medio interno y el medio externo que le circunda. Sin embargo, esta aseveración para algunos puede resultar sumamente incómoda, en especial cuando consideramos el grado de complejidad que alcanza la actividad conciente en los seres humanos.

En un nivel más abstracto y simbólico, son los conocimientos compartidos colectivamente a través del lenguaje, aquellos que conforman nuestro legado cultural, los que han hecho que en gran medida los seres humanos seamos capaces de establecer relaciones que van más allá de nuestros horizontes sensoriales, o en otras palabras, que seamos concientes de una realidad que va más allá de nuestros límites perceptuales, e incluso de nuestra propia experiencia (e historia) personal. La conciencia humana[3] ha creado modelos del mundo que trascienden sus propios límites perceptuales o sensitivos, accediendo de esa manera a universos y conceptos que no serían siquiera pensables de otro modo. Podemos apelar al conocimiento que nos proporciona la fotografía de la Torre Eiffel para saber de su existencia, a pesar de que no podemos percatarnos de esta manera de sus 320 metros de altura. Y de igual manera, tenemos a nuestra disposición el conocimiento avalado por la civilización occidental durante la última centuria para ser concientes de la enigmática organización de la materia a escala subatómica.

La actividad conciente tiene dos aspectos fundamentales: un componente pasivo y uno activo (Penrose, 1994). Según la caracterización que hace el matemático Roger Penrose, el primer aspecto corresponde a la percatación, o percepción conciente, y el aspecto activo corresponde al acto voluntario e intencional que mueve a la acción al sujeto conciente. En el vórtice de esta perspectiva dual de la actividad conciente se encuentra la propiedad más elusiva del fenómeno, esto es, la experiencia subjetiva que caracteriza los estados de conciencia, y de la cual tenemos un conocimiento directo y sensible. David Chalmers (1994b) ha puesto de manifiesto la dificultad de estudiar empíricamente este aspecto a partir de su conocida división entre los problemas fáciles y el problema 'duro' de la conciencia. Los primeros son aquellos que pueden ser descritos por una teoría de carácter funcional (p.ej., la aproximación computacional o la neurociencia experimental) puesto que los datos disponibles son observables por los métodos tradicionales de la ciencia. Estos corresponden a lo que Chalmers denomina datos de tercera persona (Chalmers, 1999, 2004). Ejemplos de este tipo de problemas son la discriminación y categorización de los

2 En inglés la palabra 'conciousness' pertenece a la misma clase de los vocablos 'awareness', 'happiness' y 'darkness', es decir, identifica una condición, un estado, atributo o cualidad de 'algo', y además, es un accidente, y por ende es una propiedad transitoria de ese algo. En español, el uso como sustantivo puede encubrir esta faceta del término, confundiendo la propiedad con el 'algo', lo cual, puede conducir al lector a la idea de que existe un 'homúnculo' que pilotea el pensamiento desde la pituitaria, o desde algún otro lugar al interior del organismo 'conciente'. Para evitar esta ambigüedad en los términos, en el texto se aludirá con preferencia a la expresión 'actividad conciente' en lugar de la expresión 'conciencia'.

3 El término 'conciencia' puede adquirir dos matices diferentes cuando se habla de los seres humanos. El más relacionado con la discusión que se sigue en este artículo es el sentido de percatación o el reconocimiento (darse cuenta) de una diferencia entre algo interior o exterior. El otro sentido se presenta en el contexto de las reflexiones sobre la conciencia moral: el conocimiento del bien y del mal. Una discusión acerca del libre albedrío, la libertad y su relación con la noción de 'pecado' se encuentra en Díaz (2004).

estímulos ambientales, la integración de información, la reportabilidad de estados mentales, el acceso a los estados internos propios y el control deliberado de la conducta.

El problema duro en la explicación de los estados concientes es la observabilidad de la experiencia conciente. El carácter fenoménico y subjetivo de ésta hace que los métodos tradicionales de investigación sean insuficientes pues los datos a los cuales tenemos acceso son de primera persona y por ende son impermeables a la observación empírica. El estudio de las propiedades subjetivas de la experiencia conciente nos ubica frente a un marco conceptual en donde el resultado de la observación es fruto de una interacción entre el observador y lo observado, ya que "aquello que se va a observar es la observación" (Botero, 2003). Capturar la medida o naturaleza exacta de este componente fenoménico, vivencial y subjetivo es el principal desafío que debe enfrentar una teoría con estatus científico cuya pretensión sea dar una explicación completa del fenómeno conciente. A continuación expondré en términos generales la controversia más importante suscitada con relación a la plausibilidad de una teoría computacional de la conciencia.

El problema de la conciencia: controversia Searle – Chalmers[4]

Frente a lo expuesto en las líneas previas, John Searle (1980, 1987, 1992) considera que los modelos cognitivos tradicionales de la inteligencia artificial y la ciencia computacional, nunca podrán re-crear el componente fenoménico de la actividad conciente puesto que la experiencia subjetiva es irreductible, es decir, no existe un lenguaje o una representación formal o simbólica adecuada que pueda describirla, y menos aún explicarla causal o funcionalmente. La objeción de Searle radica en que un sistema puede comportarse como si fuese conciente sin realmente serlo, en tanto no hay una conexión lógica y necesaria entre los estados internos subjetivos y la conducta externa públicamente observable[5]. En particular, a Searle le perturba la posibilidad de que a partir de la manipulación estrictamente sintáctica, un mecanismo artificial desarrolle la clase de contenidos y significados que asociamos con nuestra propia experiencia subjetiva. En palabras del mismo Searle: "ontológicamente hablando, el comportamiento, el rol funcional y las relaciones causales son irrelevantes a la existencia de los fenómenos mentales concientes" (Searle, 1992, p. 69). Según él, los modelos computacionales de la experiencia conciente tienen el mismo problema de interpretación que tiene la simulación asistida por computador de las tormentas de Londres. Por más detallado y refinado que sea el modelo, al ejecutar el programa nadie resulta 'empapado' por la lluvia creada digitalmente.

De la misma manera, simular algunas propiedades de los estados mentales —incluidos los estados concientes— en un computador no significa que éste en sí mismo posea algún tipo de experiencia conciente o fenoménica de la realidad, tal como la tenemos nosotros. Si esta afirmación es cierta, la posibilidad de ver en el futuro máquinas con 'conciencia artificial', o 'entes' concientes sin una base orgánica y/o biológica queda entonces restringida a la ciencia-ficción y a la fantasía de aquellos que rinden culto a los sistemas de inteligencia artificial (IA), y especialmente a la versión fuerte de ésta que asegura que las funciones cognitivas, incluyendo la percepción, el pensamiento, el lenguaje y la actividad conciente, son formas de computación de alta

4 La discusión original fue entre Searle y los esposos Paul y Patricia Churchland, pero Chalmers adoptó la misma posición que estos últimos. (ver artículo original del debate en el No. 1, de enero de 1990 de Scientific American. El debate se tituló "Artificial intelligence: A debate:").

5 Aunque esta aseveración socava también la viabilidad de la perspectiva conductista para dar cuenta de la actividad conciente, dados los objetivos del presente ensayo, en la discusión que sigue me concentraré específicamente en los argumentos que competen al paradigma computacional y a su derivación conceptual, la inteligencia artificial conexionista.

complejidad, y es sólo cuestión de tiempo para diseñar una máquina de cuyos cómputos emergerá la conciencia y las nociones de subjetividad e individualidad que caracterizan el comportamiento del ser humano. Para una fracción de los opositores de la IA fuerte, en la práctica las interacciones cerebrales que dan lugar a los seres concientes son tan complejas, que no será posible ver este tipo de organismos en un tiempo razonable para la escala humana. La posición de Searle, mucho más radical, sostiene que existe una imposibilidad de principio puesto que la vida mental y conciente tiene propiedades que no son computables (una postura similar se puede encontrar en Hodgson, 1991 y Penrose, 1989, 1994).

La crítica de Searle hacia la aproximación 'computacional' de la conciencia tiene su fundamento en el conocido problema mental de la 'habitación china' (Searle, 1980, 1987). Se trata de una habitación cerrada con sólo dos ventanas (una de entrada y otra de salida) en donde se encuentra un individuo *S* aislado del exterior que manipula un puñado de tarjetas escritas con ideogramas chinos. Algunas de ellas contienen las reglas de transformación que le permiten a *S* lanzar por la ventana de salida las tarjetas que corresponden a las que le fueron entregadas por la otra ventana. La metáfora con respecto al paradigma computacional nos ofrece un enigma: ¿Acaso entiende *S* lo que está haciendo, tal como lo entendería alguien que conoce el idioma chino y, por lo tanto, puede identificar el contenido de las tarjetas? La analogía, tanto como el cuestionamiento en ella anidado, pueden extenderse al considerar, en lugar de un individuo *S*, a una población completa de individuos (p.ej., los habitantes de China), o a un conglomerado de neuronas procesando información bioeléctrica. ¿Cómo puede surgir una mente unificada y conciente a partir de este orden estrictamente sintáctico que nos propone la teoría computacional?

Según el cuestionamiento de Searle, aún si tuviésemos la tecnología necesaria para hacer un análisis detallado de las propiedades de descarga que tienen las neuronas del cerebro de un individuo, o incluso si el conocimiento en las neurociencias fuese tan avanzado que nos permitiera disponer de los planos completos del patrón de organización causal que tiene el cableado interno del cerebro, las simulaciones que implementaren semejante mapa de conectividad neurocomputacional no podrían dar lugar al aspecto fenoménico de la actividad conciente: la experiencia subjetiva. Y esto se debe a que existe un tinte especial, hay algo intrínseco a ésta, algo de orden puramente semántico, que evade la descripción causal, mecanicista y sintáctica del algoritmo empleado para procesar la información. Bajo esta perspectiva no podemos decir que el procesamiento mismo sea en sí equivalente a esa comprensión de significado que tenemos los seres que, además de manipular y procesar símbolos, somos también concientes de su contenido semántico.

Por otra parte, David Chalmers (1994a) justifica el papel de la explicación computacional en el estudio de la cognición, y en particular de la actividad conciente, argumentando que es justamente, en virtud de la implementación computacional, que un sistema tiene 'propiedades mentales'. A diferencia de la tormenta de Londres, y otras funciones o fenómenos que al ser simulados computacionalmente pierden una parte esencial de su naturaleza (p. ej., la digestión, la oxidación, o la fotosíntesis), los estados mentales *son*, intrínsecamente funciones computables. Los estados mentales no dependen de los fundamentos físico-químicos del sistema en el cual se instalan, sino de la *organización causal* abstracta y formal que puede ser especificada computacionalmente. Para sustentar esta idea, Chalmers acude a un experimento mental desarrollado por el mismo Searle (Chalmers, 1994a, 1996; Searle, 1992), conocido como "el experimento del cerebro de silicona". Según éste, las neuronas de un individuo normal son reemplazadas gradualmente por chips de silicona de manera tal que si en cada estadio del reemplazo se conserva la organización y estructura funcional del cerebro, sería válido afirmar que el organismo de silicona resultante conserva las experiencias concientes a las cuales tenía acceso

el individuo con el cerebro 'orgánico' original. El principio de *invarianza organizacional* postulado por Chalmers (1996) establece que si algún sistema tiene experiencias concientes, entonces, cualquier sistema que tiene la misma organización causal de detalle fino tendrá cualitativamente experiencias idénticas (Chalmers, 1996). Si bien esta conclusión depende en gran medida de qué tan detallada sea esta 'equivalencia funcional' que preserva la organización causal (Hershfield, 2002), la idea principal formulada por Chalmers es suficiente para sembrar dudas sobre la postura adoptada por Searle.

En el apartado que sigue presentaré de manera sucinta algunas precisiones sobre la teoría de la computabilidad y los formatos de representación simbólica y subsimbólica con el fin de mostrar los aspectos que hacen que el paradigma conexionista y la aproximación teórica del Procesamiento Paralelo Distribuido sean candidatos plausibles para explorar computacionalmente esa organización causal sobre la cual se puede llegar a pensar una teoría formal de la actividad conciente.

Computabilidad y formatos de representación simbólica y subsimbólica

La teoría de la computabilidad formulada a partir del trabajo del británico Alan Turing (1937), introdujo los fundamentos teóricos necesarios para analizar formalmente los mecanismos del pensamiento, y en general, de los estados mentales. La concepción de la teoría clásica de la computabilidad, el legado de Turing y todos aquellos que desarrollaron sus ideas, llevó al límite la idea que plantea al cerebro como un procesador que funciona lógicamente al igual que lo hacen los mecanismos seriales, discretos y digitales que llamamos hoy en día 'computadores' (u ordenadores). Esta metáfora del cerebro como una computadora supone que aquel —al igual que todas las funciones que lo caracterizan como parte del aparato cognitivo— es un sistema que procesa información y opera aplicando reglas y manipulando símbolos (p.ej., cualquiera de las tarjetas escritas con ideogramas chinos del ejemplo de la sección anterior). En esta equivalencia funcional entre el cerebro y la lógica propia de las máquinas electrónicas, se presume la existencia de un código, o un formato de *representación,* cuyas unidades de análisis fundamentales son esos símbolos[6].

Smolensky (1988) muestra cómo la aproximación simbólica de la teoría computacional asume que los elementos computacionales, es decir, las unidades que portan la información sintáctica (p. ej., las reglas de transición en una máquina de Turing y las funciones lógicas binarias del álgebra de Boole), son también las unidades semánticas, es decir, las que tienen el significado de esa información que es expresada por los símbolos. En términos de Smolensky, éstos son a la vez (1) los elementos primitivos computacionales, las unidades de análisis necesarias para efectuar un cómputo, y (2) los primitivos representacionales, los elementos básicos a partir de los cuales se construye el significado, o el contenido semántico. Sin embargo, el conexionismo[7] de segunda generación que floreció hacia finales de la década de los 80's —lo que se conoce hoy en día como el enfoque del Procesamiento Paralelo Distribuido (PPD)— determinó un cambio en la forma de pensar con respecto a este formato clásico de *'representación computacional'* (Rumelhart & McClelland, 1986). Desde esta perspectiva, los fenómenos

6 En el modelo mental de la habitación china, el componente sintáctico de los símbolos es el único aspecto conocido por el individuo que se haya adentro, mientras que las propiedades semánticas asociadas al contexto parecen estar reservadas a quienes, fuera de la habitación y teniendo previo conocimiento del idioma chino, conocen el significado de cada ideograma escrito en las diferentes tarjetas que el individuo S ¿ingenuamente? manipula.

7 Término acuñado por Donald Olding Hebb (1949). Denomino 'conexionismo de primera generación' a esta forma original del paradigma conexionista.

cognitivos pueden ser explicados en función de la actividad de redes que se comportan como sistemas dinámicos no lineales enmarcados en una organización a gran escala del cerebro y del sistema nervioso.

El modelo teórico de PPD asume que el cerebro está constituido por un conjunto de elementos de cómputo (las neuronas) que efectúan operaciones simples e interactúan localmente a través de un conjunto de relaciones de conexión (uniones sinápticas) que pueden ser modificadas por mecanismos de aprendizaje y auto-organización. El cerebro es considerado como un sistema de procesamiento paralelo masivo que representa el conocimiento por medio de la actividad conjunta y distribuida de una población de neuronas. El hecho de que la información sea representada de esta manera implica dos cosas: (1) que la actividad de cada elemento contribuye en alguna medida a la representación de un atributo o categoría, y (2) que cada atributo es representado por el patrón global de activación de los diferentes elementos que conforman el sistema. A diferencia de los formatos de representación puramente simbólicos, esta aproximación conexionista plantea que los elementos básicos de significado o contenido semántico (los primitivos representacionales) son diferentes de las unidades computacionales, y están dados en forma distribuida a un nivel superior al de cómputo (Smolensky, 1988). Puesto que los primitivos computacionales (p. ej., las neuronas) ejecutan operaciones simples, en sí mismas carecen de contenido semántico. Pero es a través del patrón de actividad conjunta de estos elementos, y de su paso por una vasta matriz de conexiones, que se llegan a constituir las unidades de significado para el sistema cognitivo. Es por eso que Smolensky denomina *subsimbólica* a esta clase de computación, puesto que el nivel del cómputo opera a un nivel más fundamental que el de las representaciones.

Desde una perspectiva conexionista similar, sólidamente enraizada en la investigación experimental neurobiológica, Christof Koch (1996) nos plantea una definición de representación simbólica que es afín al interés científico por el estudio de la actividad conciente. Según este autor, existe una interpretación simbólica de la realidad cuando se presenta un patrón de disparo de una neurona (valor escalar), o de un vector de actividad de una agrupación de neuronas (vector de población), que tiene una fuerte correlación con una característica particular del mundo (p. ej., aspectos visuales de orientación, color, textura y movimiento). El significado de ese 'símbolo neuronal' depende del estatus en el que se encuentra ese selecto grupo de neuronas en una jerarquía de múltiples niveles. Es decir, que el contenido de un símbolo neuronal está determinado por las neuronas que proyectan hacia este grupo de células, su campo receptivo, y por aquellas que se ven afectadas por él, es decir, su campo de estimulación.

La caracterización de la representación subsimbólica que hace Smolensky y la noción de símbolo neuronal de Koch establecen los fundamentos conceptuales para formular una teoría de carácter formal-computacional de los estados mentales a un nivel más relacionado con la organización causal del cerebro, el único mecanismo que conocemos con la capacidad para dar lugar al fenómeno conciente —en contraste con el modelo computacional clásico basado en el simbolismo propuesto por la máquina de Turing. En particular, esta postura permite comprender el estatus de la representación neuronal de los contenidos de la actividad conciente, tema que procederé a discutir en el siguiente apartado.

Representación de los contenidos de la actividad conciente

Los contenidos específicos de un estado conciente son aquellos estados de gradación fina de la *experiencia subjetiva* en los cuales uno puede encontrarse en un momento dado (Chalmers, 2000). Un estado de contenido específico corresponde a esa imagen visual detallada que experimentamos al ver en una fotografía un conjunto de formas, colores y objetos dispuestos entre sí de acuerdo a patrones diferentes de

profundidad y distancia. Es a partir de estos elementos fragmentarios que surge la experiencia subjetiva distintiva de cada imagen que vemos, así como también existen experiencias particulares para cada secuencia de sonidos que conforman para nosotros una frase musical, o al igual que un zorro salvaje podría tener experiencias subjetivas diferentes de la multitud de olores que encuentra a diario en el bosque. Gran parte de los contenidos de las experiencias sensoriales que hacen parte de la actividad conciente pueden ser caracterizados según las categorías físicas equivalentes. Por ejemplo, en el sistema visual, la información puede analizarse en términos de la frecuencia de onda, la posición y la orientación en el espacio, el contraste brillo/sombra y el movimiento. Todas estas categorías son representadas en el cerebro según la actividad de una población de neuronas que se encuentran distribuidas espacialmente y que responden de una manera particular a cada una de estas dimensiones del espectro visual.

En la investigación experimental neurobiológica se presume que los contenidos de la experiencia conciente están correlacionados con la actividad exhibida por un sistema neuronal, cuando, ante un evento específico (por ejemplo, un estímulo visual consistente en una rejilla de barras horizontales) un individuo reporta verbalmente que efectivamente 've' el estímulo en cuestión, y además, en su cerebro se detecta (p. ej., a través de microelectrodos insertados en la superficie de la corteza cerebral) que hay un grupo particular de neuronas que dispara sistemáticamente con un patrón específico de potenciales de acción ante ese evento en particular. El contenido definido por los patrones neuronales de activación es concebido entonces como una *representación* del contenido aparente que tiene para el individuo la experiencia conciente de ver el estímulo. La naturaleza de esta '*representación neuronal*' depende tanto de las características de activación espacial como del patrón temporal de disparo del grupo de neuronas que están involucradas en la percepción conciente.

Un *sistema de representación neuronal* es un código —o un lenguaje— mediante el cual el cerebro traduce o representa la información acerca de las propiedades y de los contenidos específicos del estímulo presentado. El análisis de este código ha resultado una tarea bastante compleja puesto que, además de representar las propiedades del estímulo (p. ej., frecuencia de onda, posición y orientación), un sistema de representación neuronal debe tener en cuenta que éstas varían en función del tiempo (Abbott & Sejnowski, 1999; Gabbiani & Koch, 1999). Formalmente, un sistema de representación de esta naturaleza puede ser conceptualizado a partir de la misma clase de descripciones cuantitativas que ofrecen las ecuaciones de conducción nerviosa que propusieron Alan Hodgkin y Andrew Huxley, las cuales son el fundamento biofísico de los modelos computacionales de la neurociencia experimental que caracterizan la investigación contemporánea (Hodgkin & Huxley, 1952; Koch & Segev, 1998; O'Reilly & Munakata, 2000).

El desarrollo de estos modelos conexionistas ha puesto de manifiesto que la formalización matemática del procesamiento neuronal de la información puede contribuir a esclarecer, tanto las propiedades computacionales del funcionamiento cerebral, como las bases analíticas para formular las características de la representación neuronal de los contenidos de la experiencia conciente. Sin embargo, aún existen brechas conceptuales que deben ser superadas para establecer una teoría reduccionista de la conciencia.

La representación distribuida y subsimbólica que caracteriza la perspectiva conexionista ha superado muchos de los escollos que tenía inicialmente el paradigma clásico de la computación, pero trae consigo nuevos problemas que debe enfrentar el análisis formal de la actividad conciente. El problema de la conjunción o integración[8] es uno de esos desafíos conceptuales del conexionismo. Se presenta cuando diferen-

8 Versión en castellano de lo que se conoce en la literatura anglosajona como 'the binding problem'

tes características de un estímulo dado son procesadas de manera completamente separada por diferentes unidades, categorías o modalidades de representación. La cuestión, en breves términos es: ¿Cómo se integra la información que inicialmente ha sido representada de manera distribuida por billones de neuronas? O en otras palabras, ¿Cómo llega a ser experimentada concientemente como una sola unidad perceptual? Cuando vemos, por ejemplo, una manzana, la información que llega a la retina es procesada por grupos independientes de neuronas que se han especializado para detectar colores, orientación, movimiento, etc., y que se localizan en distintas áreas de la corteza visual. El problema consiste en comprender cómo los disparos individuales de estas neuronas pueden dar lugar a la 'unidad de la apercepción', es decir, a la sensación unificada de 'ver una manzana' (Churchland, 1994; McFadden, 2002).

Una de las hipótesis más destacadas en este sentido es la de Francis Crick (1994; Crick & Koch, 1990), quien enfatiza que el correlato neurofisiológico de la unificación de la información está relacionado con un patrón sincrónico de descarga a lo largo de una gran población de neuronas, así como también, las transformaciones que sufre al pasar por una vasta matriz de conexiones sinápticas (Churchland, 1996; Koch, 1996; Koch & Segev, 1998). Esta idea es el fundamento de la noción de 'símbolo neuronal' de Koch reseñada anteriormente. Un factor importante en esta propuesta es la función integradora de las oscilaciones de 40 Hertz que se propagan a lo largo del sistema de proyecciones que se conectan bidireccionalmente desde el tálamo hasta la corteza (Crick, 1994; Crick & Koch, 1990, 1995; Edelman, 1989; Llinas & Ribary, 1993; Llinas, Ribary, Joliot & Wang, 1994).

Además de la sincronía temporal de la onda de 40 Hertz, existen otras teorías, como la de Karl Popper, que intentan dar solución a este problema postulando que la unidad perceptual de la actividad conciente es una manifestación de un campo de fuerza (p. ej., electromagnético) capaz de enlazar e integrar la información que está distribuida en una población de neuronas (Hardcastle, 1994; Libet, 1994, 1996; McFadden, 2002; Popper, Lindahl & Arhem, 1993). La importancia de la propuesta de Crick radica en que aporta una base conceptual para entender de manera empírica y formal la naturaleza de la visión unificada del mundo que parece caracterizar el fenómeno conciente. El enfoque neurobiológico y computacional que subyace a esta hipótesis ha permitido develar una serie de estructuras y procesos cerebrales —los correlatos neuronales— implicados en algunas de las principales manifestaciones de los estados concientes (Chalmers, 2000, 2004; Churchland, 1994). Esta clase de análisis ha sido fructífera para analizar cuantitativa y cualitativamente los contenidos específicos de la percepción visual. En la parte final de este ensayo discutiré cómo pueden los elementos anteriormente señalados integrarse para establecer la viabilidad del proyecto computacional conexionista frente a la postulación de una teoría formal de la actividad conciente.

DISCUSIÓN

La controversia planteada por Searle acerca de la implausibilidad del proyecto computacional en la búsqueda de una explicación científica de los estados mentales, y en particular de la actividad conciente, se ha enfocado principalmente en la crítica de la teoría clásica de la computabilidad, es decir, la perspectiva simbólica de la inteligencia artificial. Sin embargo, a partir de los planteamientos presentados en los párrafos anteriores parece claro que las propuestas conexionistas pueden mostrar una forma alternativa de comprender la naturaleza de la mente y de los mecanismos del pensamiento que no solamente está íntimamente relacionada con el funcionamiento cerebral; también implica una reformulación de conceptos (p. ej., representación computacional) que hace necesario postular una clase de lógica alternativa al tipo de formalismo empleado por la corriente clásica de la inteligencia artificial. El nivel subsimbólico del conexionismo exige un análisis que está más ligado al paradigma de la

complejidad y la teoría matemática de los sistemas dinámicos que al esquema lógico-proposicional asumido por el modelo clásico de la computabilidad basado en la máquina de Turing. Pero este tipo de análisis puede implicar a su vez un desafío frente a la concepción que tenemos de la forma en que opera la mente.

El formalismo en la ciencia, indudablemente, ha contribuido a ampliar las fronteras del conocimiento que tenemos de nosotros mismos y del mundo, pero no sin un costo. Hemos tenido que cuestionar una serie de nociones que asumíamos como verdaderas, pero a partir de un análisis profundo, hemos reconocido que son ilusiones creadas por las limitaciones de nuestros sentidos y por las restricciones que tenemos para percatarnos de la realidad dada nuestra historia filogenética. Un ejemplo de revolución 'cognitiva' que obligó a replantear nuestra visión de la realidad fue el advenimiento de la teoría de la relatividad y la extraña lógica de la mecánica cuántica en la física. Estas construcciones conceptuales determinaron un cambio en la manera de pensar el universo y las leyes físicas que rigen el comportamiento de la materia y la energía. Así, no sería extraño que la formulación conexionista se llegase a apartar radicalmente de la forma tradicional en la que concebimos los procesos mentales, incluidos los estados concientes.

Ahora bien, ¿qué elementos aporta la formalización computacional del conexionismo al estudio teórico del fenómeno conciente? Asumiendo como cierta la idea de Chalmers acerca de que existe una organización causal que subyace a la actividad conciente, y que es susceptible de ser explicitada computacionalmente, la viabilidad del proyecto conexionista parece depender de manera crucial de los siguientes aspectos:

(1) El grado de refinamiento con el que se defina la representación y la computabilidad de los estados concientes. La noción de *computabilidad* implicada en este contexto trasciende el esquema conceptual basado en el paradigma de procesamiento simbólico clásico y se asemeja más a la definición de 'símbolo neuronal' que tiene Christof Koch (1996). Una '*representación computacional*', en este sentido adquiere un sentido matemático y filosófico más amplio, puesto que ya no se trata solamente del resultado dado por una serie de procedimientos de manipulación de símbolos, sino que se trata de una función o regla de transformación que expresa una realidad constituida por relaciones susceptibles de análisis cuantitativo y cualitativo que están gobernadas por las propiedades inherentes al procesamiento neuronal. Y sabemos que estas propiedades tienen un carácter dinámico y no lineal, es decir que varían de manera compleja en función del tiempo (p. ej., gracias a mecanismos de aprendizaje y auto-organización).

(2) Los avances metodológicos y conceptuales de las ciencias del cerebro. El lenguaje empleado en el campo de la neurociencia computacional hace uso de las unidades y variables biofísicas (p. ej., potencial de membrana, corriente, conductancia eléctrica) para expresar los patrones de organización causal —o conectividad funcional si se prefiere— del procesamiento neuronal. Este tipo de descripciones logran construir un puente entre los datos que surgen en el análisis de célula individual y los datos a gran escala de poblaciones de neuronas, incluyendo el procesamiento a nivel cognitivo y conciente. De esta forma, considerando que el cerebro es una máquina biológica que codifica la realidad física transformándola en patrones físico-químicos de organización inherentes a las células nerviosas, los datos de investigación empírica que generan estos estudios pueden hacer precisiones acerca de las bases neuronales que caracterizan, por ejemplo, los contenidos específicos de la experiencia conciente.

(3) La distinción que se haga entre simulación, modelamiento y representación computacional de la actividad conciente. La simulación puede ser entendida como una construcción de un modelo computarizado que pretende reflejar —más no equipararse a— la realidad a través de un len-

guaje o código representacional que sirve para determinar posibles circunstancias y soluciones de los hechos estudiados. Concebida de esta manera la simulación, es claro que la crítica de Searle a los modelos computacionales como un método o herramienta para simular el fenómeno conciente adquiere mayor sentido. No obstante, siguiendo el planteamiento de organización causal de Chalmers, la perspectiva computacional —como una formulación teórica de los estados mentales en lugar de una metodología para emularlos— tiene un fundamento sólido para determinar el tipo de relaciones abstractas imbricadas en la actividad conciente. Y el conexionismo proporciona un avance hacia la comprensión de índole causal y/o funcional de esas relaciones.

Trátese de una imposibilidad de principio, como asume Searle, o bien sea una cuestión de implementación, el análisis teórico-formal de la conciencia que se está desarrollando desde la perspectiva conexionista tiene elementos relevantes que vale la pena considerar seriamente. El más importante de todos quizás sea el hecho de que se trata de un paradigma que toma como modelo el funcionamiento cerebral. Y ese trozo de materia orgánica, hasta donde sabemos, incluso por experiencia propia, sigue siendo aún la última frontera por explorar a fin de entender el porqué de su interior emerge la actividad conciente.

Son los patrones de disparo neuronal que subyacen a cada pensamiento que diariamente experimentamos los únicos que hasta ahora pueden ser teñidos de esa coloratura que permea nuestros actos con una variedad de vivencias. Por eso, aún cuando los procesos de nivel biológico no sean los únicos capaces de constituir ese fenómeno de alto nivel que asociamos a la actividad conciente, clarificar las bases neuronales que la subyacen es una aventura que puede contribuir a establecer los cimientos de una explicación científica de la conciencia. Para Searle y aquellos que comparten con él su desilusión hacia la teoría de la computabilidad, este enfoque quizás no sea el mejor paradigma para comprender el fenómeno de la 'con-scientia', y de ninguna forma develará el mayor de sus grandes misterios: la experiencia conciente. Para estos pensadores, perspectivas como las del conexionismo y la neurociencia computacional están diseñando en máquinas de silicio 'tormentas' cuyos truenos no se podrán jamás escuchar.

REFERENCIAS

Abbott, L. & Sejnowski, T. J. (1999). *Neural codes and distributed representations: foundations of neural computation.* Cambridge: MIT Press.

Botero, J. J. (2003 Diciembre). La fenomenología y el estudio de la conciencia. Ponencia presentada en el encuentro Tiempo Realidad y Conciencia, por el Grupo de Estudios Contemporáneos en Conciencia y la Facultad de Ciencias Humanas de la Universidad Nacional. Tomado de http://www.humanas.unal.edu.co/conciencia/textos/mesaredonda_2.htm

Carruthers, P. (2001). Higher-Order theories of consciousness. En E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Tomado de http://plato.stanford.edu/entries/consciousness-higher/

Chalmers, D. J. (1994a) On implementing a computation. *Minds and Machines, 4,* 391-402.

Chalmers, D. J. (1994b). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3):200-19, 1995. Re-impreso en J. Heil (Ed.), *Philosophy of Mind: A Guide and Anthology.* Oxford University Press (2003). Tomado de http://www.u.arizona.edu/~chalmers/papers/facing.html

Chalmers, D. J. (1996). *The conscious mind.* Oxford University Press.

Chalmers, D. J. (1999). First-person methods in the study of consciousness. *Consciousness Bulletin,* University of Arizona. Tomado de http://www.u.arizona.edu/~chalmers/papers/firstperson.html

Chalmers, D. J. (2000). What is a neural correlate of consciousness? En T. Metzinger (Ed.), *Neural correlates of consciousness: Conceptual and empirical questions.* Cambridge: MIT Press. Tomado de http://www.u.arizona.edu/~chalmers/papers/ncc2.html

Chalmers, D. J. (2004). How can we construct a science of consciousness?. En M. Gazzaniga (Ed.), The cognitive Neurosciences III. Cambridge: MIT Press. Tomado de http://www.u.arizona.edu/~chalmers/papers/scicon.html

Churchland, P. S. (1994). «Can neurobiology teach us anything about consciousness?». Presentación en American Philosophical Association, Pacific Division. *Proceedings and Addresses of the American Philosophical Association, 67,* 4, 23-40. Tomado de http://cogsci.soton.ac.uk/~harnad/Papers/Py104/chuch.neuro.html

Churchland, P.M. (1996). Learning and conceptual change: The view from the neurons. En A. Clark & P.J.R. Millican (Eds.), *Connectionism, concepts and folk psychology: The legacy of Alan Turing* (Vol. 2). Oxford: Clarendon Press.

Crick, F. (1994). *The astonishing hypothesis*. Nueva York: Scribner's Sons.

Crick, F. & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in neurosciences, 2,* 263-275.

Crick, F. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex?. *Nature, 375,* 121-123.

Díaz, J. A. (2004). Libertad, determinismo y libre albedrío. Ponencia presentada en el encuentro Voluntad y Libre Albedrío: ¿Ficciones del cerebro? , realizado en Bogotá por el Grupo de Estudios Contemporáneos en Conciencia & la Facultad de Ciencias Humanas de la Universidad Nacional. Tomado de http://www.humanas.unal.edu.co/conciencia/textos/foro_mayo2004.pdf

Edelman, G.M. (1989). *The remembered present: A biological theory of consciousness*. Nueva York: Basic Books

Gabbiani, F. & Koch, C. (1999). Coding of time-varying signals in spike trains of integrate-and-fire neurons with random threshold. En L. Abbott & T. J. Sejnowski (Eds.), *Neural codes and distributed representations* (201-224). Cambridge: MIT Press.

Hardcastle,V.G. (1994). Psychology's binding problem and possible neurobiological solutions. *Journal of Consciousness Studies, 1,* 66-90.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Nueva York: Wiley.

Hershfield, J. (2002). A note on the possibility of silicon brains and fading qualia. *Journal of Consciousness Studies, 9, 7,* 25-31.

Hodgkin, A. L. & Huxley, A. F. (1952). Quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117,* 500-544.

Hodgson, D. (1991). *The mind matters: Consciousness and choice in a quantum world.* Oxford: Clarendon Press.

Koch, C. (1996). Toward the neuronal substrate of visual conciousness. En S. A. Hameroff, A. W. Kaszniak & A. C. Scott (Eds.), *Toward a science of consciousness* (pp. 247-258). cambridge: MIT Press.

Koch, C. & Segev, I. (1998). *Neural modeling: from ions to networks*. Cambridge: MIT Press.

Libet, B. (1994). A testable field theory of mind-brain interaction. *Journal of Consciousness Studies, 1,* 119-126.

Libet, B. (1996). Conscious mind as a field. *Journal of Theoretical Biololgy, 178,* 223-226.

Llinas, R. R. & Ribary, U. (1993). Coherent 40-Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Sciences, 90,* 2078-2081.

Llinas, R.R., Ribary, U., Joliot, M. & Wang, X. J. (1994). Content and context in temporal thalamocortical binding. En G. Buzsaki, R.R. Llinas & W. Singer (Eds.), *Temporal Coding in the Brain* (pp. 251-272). Berlin: Springer Verlag.

McFadden, J. (2002). Evidence for an electromagnetic field theory of consciousness. *Journal of Consciousness Studies, 9,* 23-50. Tomado de http://www.mindcontrolforums.com/electromagnetic-field-theory-of-consciousness.htm

Nagel, T. (1974). *What is it like to be a bat? Philosophical Review, 83*. Tomado de http://members.aol.com/NeoNoetics/Nagel_Bat.htmlNagel_Bat.html

O'Reilly, R. & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: understanding the mind by simulating the brain*. Cambridge: MIT Press.

Penrose, R. (1989). *The Emperor's new mind*. Inglaterra: Oxford University Press.

Penrose, R. (1994). *Shadows of mind: A search for the missing science of consciousness*. Inglaterra: Oxford University Press.

Popper, K.R., Lindahl, B.I. & Arhem,P. (1993). A discussion of the mind-brain problem. *Theoretical Medicine, 14,* 167-180.

Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vols. 1 y 2). Cambridge, MA: MIT Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3, 417-457.* Reimpreso En D. R. Hofstadter & D. C. Dennet (Eds.), The Mind's I. Harmondsworth: Penguin Books.

Searle, J. R. (1987). Minds and brains without programs. En C. Blakemore & S. Greenfield (Eds.), *Mindwaves* (pp. 209-2339. Oxford : Basil Blackwell.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge: MIT Press.

Searle, J. R. (1997). *The mistery of conciousness*. Nueva York: The New York Review of Books.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1-23.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungs problem. *Proceedings of the London Mathematics Society, 42,* 230-265.

Recepción: octubre de 2004

Aceptación final: julio de 2005