El juego de imitación de Turing y el pensamiento humano

Turing's imitation game and human thought

LEONARDO FRANCISCO BARÓN BIRCHENALL*
Universidad de Buenos Aires, Argentina

Abstract

In 1950, the English mathematician Alan Mathison Turing proposed the basis of what some authors consider the test that a machine must pass to establish that it can think. This test is basically a game; nevertheless, it has had great influence in the development of the theories of the mind performance. The game specifications and some of its repercussions in the conception of thinking, the consciousness and the human will, will be ramifications of the path that will take us through the beginning of the artificial intelligence, passing along some of its singular manifestations, to culminate in the posing of certain restrictions of its fundaments.

Key words: artificial intelligence; thought; simulation; Turing test.

Resumen

En 1950, el matemático inglés Alan Mathison Turing propuso los fundamentos de lo que algunos autores consideran la prueba que debería pasar una máquina para establecer que piensa. Esta prueba es básicamente un juego; sin embargo, ha tenido gran influencia en el desarrollo de las teorías sobre el funcionamiento de la mente. Las especificaciones del juego y algunas de sus repercusiones en la concepción del pensamiento, la conciencia y la voluntad humana, serán ramificaciones del camino que nos llevará a través de los inicios de la inteligencia artificial, pasando por algunas de sus singulares manifestaciones, a culminar en el planteamiento de ciertas restricciones de su fundamento.

Palabras clave: inteligencia artificial; pensamiento; simulación; prueba de Turing.

Solamente un ensayo, solamente una tentativa de humanidad.
Al principio hablaron, pero sus rostros se desecaron;
sus pies, sus manos, [eran] sin consistencia;
ni sangre, ni humores, ni humedad, ni grasa;
mejillas desecadas [eran] sus rostros; secos sus pies, sus manos; comprimida su carne.
Por tanto no [había] ninguna sabiduría en sus cabezas, ante sus Constructores,
sus Formadores, sus Procreadores, sus Animadores
Popol-Vuh**,** Anónimo.

El juego de Turing

En su artículo de 1950 en la revista *Mind*, titulado "Los aparatos de computación y la inteligencia", Alan Turing (1950/1983) plantea la posibilidad de pensamiento por parte de las máquinas; para esto, se sirve de un juego al que llama *Juego de la Imitación*. Este consta de tres participantes: un hombre (A), una mujer (B) y un interrogador (C) (que puede ser hombre o mujer). C se encuentra separado de los otros dos jugadores y no puede verlos ni escucharlos; solo los conoce como X e Y; además, solo puede comunicarse con ellos en forma escrita o mediante un mensajero (idealmente con una máquina, para evitar el reconocimiento caligráfico).

* Correspondencia: Leonardo Francisco Barón Birchenall. Calle 54 No. 9-15, Apto. 204, Bogotá, Colombia. Correo electrónico: laescaladesol@gmail.com.

El objetivo del interrogador es adivinar quién es el hombre y quién la mujer; el del hombre es inducir al interrogador a hacer una identificación errónea; y el de la mujer es colaborar con el interrogador para que este identifique correctamente quién es quién. El interrogador puede hacer preguntas del tipo *¿Puede decirme X de qué largo tiene el pelo?* (Turing, 1950, p. 70), a las que los otros dos jugadores pueden responder de la forma que consideren más conveniente y convincente para lograr su cometido; por lo demás, el entrevistador no puede exigir demostraciones prácticas de ningún tipo a los otros participantes. Estas son las reglas del juego; no se especifica tiempo límite ni otras restricciones.

Ya explicado el proceso, Turing plantea los siguientes cuestionamientos: "¿Qué sucederá cuando una máquina tome la parte de A en el juego? ¿Decidirá equivocadamente el interrogador con la misma frecuencia cuando se juega así el juego como ocurre cuando en él participan un hombre y una mujer?" (p. 70). Estas preguntas remplazan, finalmente, la pregunta original acerca de la posibilidad de pensamiento en las máquinas. La idea, entonces, es esta: *si en el juego, una máquina logra engañar a un interrogador, haciéndole creer que es una mujer o que el otro jugador es un hombre, una cantidad de veces equivalente a la que ocurriría si el juego se diera entre humanos y mayor a la que ocurriría por azar, podría decirse que la máquina en cuestión piensa y, por tanto, que las máquinas pueden pensar*. Es preciso aclarar que Turing no profundiza en las conclusiones que arrojaría el hecho de que una máquina pasara la prueba, por lo menos en el artículo que estamos refiriendo; en ese sentido, parece dejar las conclusiones a juicio del lector.

Téngase en cuenta también que la máquina expuesta a la prueba puede eludir preguntas, de cualquier forma, como negándose a responder, guardando silencio, repitiendo respuestas o contestando con otras preguntas; por ejemplo: la máquina podría evitar ser descubierta por su precisión matemática, contestando adrede mal a algunas preguntas de dicha índole. Para Turing, la mejor estrategia que podría adoptar la máquina para engañar al interrogador, sería "… intentar el logro de respuestas como las que naturalmente daría un hombre" (p. 72). Para alcanzar tal grado de sofisticación, aclara el matemático, se puede hacer uso de cualquier tipo de tecnología para la creación del aparato que ha de superar el reto.

El planteamiento de Turing acerca del juego, la parte fundamental del artículo al que hacemos referencia, ocupa unas pocas líneas, un par de carillas a lo más; el resto del texto está dedicado a responder posibles acotaciones al desafío (como convencionalmente se ha denominado a la prueba). Revisemos brevemente algunas de estas objeciones, para entender mejor la propuesta de Turing.

Según la *objeción teológica*, el pensamiento es una función del alma inmortal que Dios ha dado a los hombres, pero no a los animales o a las máquinas; por consiguiente, las máquinas no pueden pensar. A esto responde Turing que no podemos "incurrir en la irreverencia de usurpar el poder de Dios de crear almas" (p. 75), y al final desecha el argumento haciendo alusión a casos del pasado en que concepciones religiosas estuvieron en contra de un hecho científicamente validado más tarde; no se crea, sin embargo, que Turing cae en el embrollo metafísico de la repartición de almas; simplemente sucede que evita el terreno de la especulación pura.

Otra objeción considerada es la llamada *de cabezas hundidas en la arena*, según la cual, "las consecuencias de que las máquinas pensaran serían horrorosas" (p. 75); objeción esta, según Turing, posiblemente hecha por personas que postulan una superioridad humana basada en el pensamiento; en consecuencia, Turing deja ir la objeción, en sus palabras, con tan solo una "nota de consuelo" (es decir, la considera *no a lugar*).

Una réplica más científica es la *objeción matemática*, según la cual en cualquier sistema lógico se pueden formular afirmaciones imposibles de probar o refutar dentro de la lógica propia de dicho sistema, a menos que el sistema mismo no sea consistente; esta objeción estriba en el segundo teorema de la incompletitud de Kurt Gödel: ningún sistema consistente se puede usar para demostrarse a sí mismo; es decir, dentro de cualquier sistema lógico se pueden generar afirmaciones imposibles de probar o refutar dentro del propio sistema, a no ser que el mismo sea incoherente (esto es, que haga uso de un lenguaje que no le es propio), ya que

no todos los teoremas posibles se desprenderían lógicamente de los axiomas del sistema, obligando al uso de un lenguaje externo para establecer los elementos ajenos[1] (Hofstadter, 1983).

La respuesta de Turing a esta objeción es tan sencilla como categórica: si bien las máquinas pueden presentar errores de pensamiento y mostrar inconsistencias, los humanos también; en otras palabras, si se pudiesen evaluar todas las ideas de un individuo, estas no presentarían una total coherencia interna. De hecho, presentarían incoherencias, muchas incoherencias; no obstante, no es esto impedimento para que el sistema cognitivo humano funcione en concordancia con el medio. En el caso del humano, como en el de los sistemas de Gödel, existe la imposibilidad de explicarse *totalmente* a sí mismo, lo cual, nuevamente, no ha sido obstáculo para que los humanos, y los sistemas, sigan existiendo.

Otra acotación considerada es el argumento de *la conciencia de uno mismo*. Este plantea la imposibilidad de que las máquinas tengan auto-conciencia, es decir, que piensen sobre su pensamiento. Turing responde a esta objeción afirmando que *los estados de conciencia solo pueden inferirse mediante la observación de la conducta*, ya que es imposible comprobar la existencia de los procesos mentales ajenos. No hacerlo de este modo, equivaldría a un solipsismo, en el que solo se da por cierto el pensamiento propio. A estas y otras posibles objeciones contesta Turing; sin embargo, la refutación más conocida al reto del pensamiento maquinal es la del filósofo norteamericano John Searle, publicada en 1980 (cuando Turing ya había fallecido) y conocida como el *argumento de la habitación china*.

El argumento de la habitación china

En términos generales, este argumento consiste en la reducción del supuesto pensamiento de las máquinas, a un proceso de relacionamiento de símbolos basado en reglas impartidas por un operador. Este proceso no incluiría la comprensión de sentidos y significados, y funcionaría bajo principios meramente sintácticos (Searle, 1983).

Concretamente, este argumento consiste en un experimento mental, perteneciente a una clase de procesos experimentales virtuales de amplia aceptación en el ámbito de la filosofía de la mente. En dicho experimento, nuestro filósofo se imagina a sí mismo encerrado en un cuarto, en el que se le entrega un manuscrito en chino (idioma que no comprende en absoluto); luego, se le da otro manuscrito, también en chino, junto a una serie de instrucciones escritas en inglés que correlacionan el manuscrito entregado en primer término con el que se entregó después (los dos en chino); a continuación, se le entrega un tercer juego de símbolos en chino y otras instrucciones en inglés que permiten correlacionar este último manuscrito chino con los dos manuscritos anteriores (con base en la forma y obviando el significado).

Con toda esta documentación, el filósofo de la habitación china es capaz de *aparentar* el conocimiento de variados asuntos que le son consultados, así como el conocimiento del idioma chino (entregando determinado manuscrito cuando se le presenta otro); es decir, relacionando juegos de símbolos que no conoce, mediante instrucciones que sí conoce (ya que se dieron en inglés). El filósofo en la habitación hace parecer que conoce los temas que se le consultan, y el lenguaje en el cual le son consultados.

Veámoslo un poco más detalladamente: las instrucciones en inglés, perfectamente entendibles para Searle (quien, no olvidemos, se ha confinado el mismo en la habitación), equivaldrían al programa de un computador. Estas indicaciones le permitirían, al hombre o a la máquina, correlacionar lo que para un observador externo (hablante del chino) sería una serie de preguntas y respuestas lógicas, coherentes y acertadas (en forma de intercambio de manuscritos). El observador externo (quien no ve los escritos, la traducción ni los demás papeleos) no tiene forma de saber que quien da contestación tan acertada no tiene ni idea de qué está hablando (ya

[1] Para establecer los límites de un sistema no se puede decir, por ejemplo, que *Y es un elemento ajeno*, sin usar el término *Y*, que no pertenece al lenguaje del sistema; por tanto, se deben usar términos impropios en el proceso de definición, imposibilitándosele al conjunto llegar a ser auto-demostrado en su totalidad mediante su propio lenguaje.

que simplemente correlacionó símbolos, sin comprenderlos). Aun más, un conjunto de jueces que realizasen preguntas en inglés y chino, y recibieran respuestas correctas en ambos idiomas, podrían deducir que un mismo proceso de pensamiento subyace a ambas.

El *quid* del planteamiento de Searle en contra de la posibilidad de pensamiento en las máquinas, es la falta de *intencionalidad* del mismo. La intencionalidad se ha considerado tradicionalmente como un *rasgo definitorio de los procesos mentales*, y tal cual fue definida por el filósofo alemán Franz Brentano en el siglo XIX consiste en el hecho de *estar dirigido a algo*; es decir, *ser acerca de algo* (Acero, 1995). Las manipulaciones formales de símbolos realizadas por las máquinas no poseerían intencionalidad, ya que la base de su organización sería la forma, y no el significado; o lo que es igual, los procesos realizados por una máquina no estarían dirigidos a algo, no serían acerca de nada. Según Searle, los humanos, por el contrario, realizamos manipulaciones sintácticas de símbolos *y*, *además*, tenemos acceso a su significado. Para este filósofo de la mente, *los humanos somos máquinas exponentes de programas de computación, manipulamos símbolos formalmente y, aun así, podemos pensar.* Las computadoras no tienen intencionalidad y, por ende, no piensan. La intencionalidad de los humanos deviene de su biología, y las entrañas de nuestros aspirantes a seres pensantes no poseen tal característica.

A la defensa de Turing: el predicamento de las máquinas y los huracanes simulados

Iniciando la década de los 80, el matemático y físico estadounidense Douglas Hofstadter defendió la posibilidad de pensamiento en las máquinas, en un escrito en forma de conversación (virtual, como el argumento de la habitación china) titulado "Temas metamágicos bizantinos. El test de Turing: Conversación en un café" (1983). En esta conversación hipotética, toman parte tres personajes igualmente hipotéticos: Sandy, estudiante de Filosofía; Pat, estudiante de Biología; y Chris, estudiante de Física. Estos camaradas discurren alegremente entre sí, dejando de a poco emerger las ideas de Hofstadter en defensa del pensamiento de las máquinas, las cuales podemos caracterizar así:

Para Hofstadter (1983): *desear, pensar, intentar y esperar* (procesos considerados tradicionalmente como intencionales), *son características que emergerían de la complejización de las relaciones funcionales de las máquinas*. Refiere también, este teórico de la mente, que los computadores se han considerado tradicionalmente como objetos fríos y cuadrados, y que acaso, si esto cambiase, se facilitaría la concepción de la inteligencia artificial, y las máquinas podrían evocar en las mentes humanas "trazados de luz danzantes más bien que palas de vapor gigantescas" (p. 112).

Refiriéndose al estatus del pensamiento artificial como simulación del pensamiento humano, *Hofstadter entiende la simulación como equivalente a lo simulado*. Para establecer este punto, se vale como ejemplo de la simulación computarizada de un huracán, la cual considera como una suerte de huracán real que modifica las relaciones existentes dentro del programa, damnificando a los unos y ceros: habitantes binarios de la simulación. Se lee en el texto: "En el caso del huracán simulado, si observas la memoria de la computadora con la esperanza de ver cables rotos y demás, tendrás una desilusión. Pero mira el nivel correcto (…) Verás que se han roto algunos lazos abstractos, que han cambiado radicalmente algunos de los valores de las variables y muchas cosas más. Aquí tienes tu inundación…" (p. 97).

Evidentemente, el análisis de Hofstadter se da en un nivel abstracto, ya que si bien no afirma que el simulacro de un huracán es *idéntico* a un huracán real, sí ubica, usando una expresión propia del ámbito filosófico, la "huracaneidad", en los efectos del fenómeno y la coherencia de sus componentes; es decir, la esencia del huracán no estaría en sus componentes físicos, sino en sus efectos dentro de las restricciones especificas del marco en que se desarrolla. Se simularía pues la esencia de lo simulado, lo cual significa, extrapolando, que *la esencia del pensamiento habría de ser el proceso que lo subyace: el cómputo matemático*. Al final, no interesaría tanto el medio en el que, y mediante el cual, se realiza la operación, sino la operación misma.

Este tipo de razonamiento corresponde a una corriente filosófica conocida como *funcionalismo*, la cual fundamenta la ciencia cognitiva presentando los fenómenos mentales en función de sus roles causales, sin depender de un constituyente físico (Bechtel, 1991) (no confundir con el funcionalismo de James). Dicho de otra forma, no importa si se es un computador o un humano; para el funcionalismo, la esencia del pensamiento radica en el proceso del mismo y no en su sustrato físico. Son palabras del texto de Hofstadter: "Yo diría que el que tú hagas depender de mi cuerpo físico la evidencia de que soy un ser pensante es un poco superficial" (1983, p. 105).

Es evidente, en este postulado, una objeción fundamental al argumento de la habitación china (en el que la intencionalidad del pensamiento emerge del sustrato biológico del organismo). Para Hofstadter, al igual que para el filósofo de la mente Jerry Fodor (1997), *la intencionalidad del pensamiento existe; pero responde a la complejidad de las relaciones funcionales del ser pensante, sea este, máquina o humano*. Según la lectura que el lingüista Ray Jackendoff hace de Hofstadter, el planteamiento de este implica que si un computador puede alcanzar un alto grado de complejidad, la conciencia emergerá de alguna forma milagrosa ("*consciousness will somehow miraculously emerge*" [1987, p. 18]).

Al igual que para Turing, para Hofstadter, la forma de constatar que los otros piensan es mediante la observación de los hechos externos: sus acciones. Esta sería evidencia directa; el resto (preguntarles, por ejemplo), sería evidencia indirecta y por tanto sospechosa. De no confiar en el método de observación externa caeríamos en un solipsismo; en términos de Hofstadter: "… la gente acepta que el prójimo tiene conciencia tan sólo porque hay un monitoreo exterior constante sobre los otros, lo cual en sí se parece mucho al Test de Turing" (1983, p. 104). En consonancia y según Fodor: "… no podemos tener nunca razones para suponer que los predicados mentales se puedan aplicar a personas distintas de nosotros mismos." (1986, p. 94).

Como consecuencia de lo anterior, el conocimiento de los estados mentales de los otros es solo probable; la única forma de acercamiento es mediante la conducta. Una de las razones de esta tesis, es validar la existencia de pensamiento, sin la presencia de procesos "internos", basándose solo en la conducta observable. De esta forma, en el caso de la habitación china, los observadores externos podrían concluir que se está produciendo un proceso legítimo de pensamiento, tan solo considerando la naturaleza y relación de las preguntas y respuestas, sin importar qué clase de proceso se dé *dentro* de la habitación.

En cuanto a la ausencia de una estructura biológica, subyacente al pensamiento maquinal, se plantea Hofstadter la posibilidad transicional entre un nivel físico y uno biológico, en el sistema constituyente del organismo; en este sentido, los hombres seríamos máquinas y la base biológica de nuestra humanidad sería una serie de procesos físicos que bien podrían ser emulados por otra máquina; de esta forma, al igual que mediante la complejización de procesos formales, también podría emerger la intencionalidad. En todo caso, en el artículo que estamos refiriendo, Hofstadter no profundiza en las particularidades de su proceso de transición entre el nivel biológico y físico en los humanos, lo cual, junto a su planteamiento de las características conscientes-emergentes, y la posible identidad de la simulación y lo simulado, basada en presupuestos funcionalistas, constituye su defensa de la posibilidad del pensamiento artificial. No obstante, la posibilidad total, o la imposibilidad absoluta del pensamiento artificial, no son la única forma de contestar al desafío de Turing; pasemos ahora, a considerar otras opciones.

Posibles respuestas al desafío de Turing

¿Realmente, pueden pensar las máquinas?

Ángel Rivière, psicólogo madrileño, refiere cuatro respuestas al desafío de Turing, en sus *Objetos con mente* (1991). La primera es simple y llanamente: No, las máquinas no pueden pensar (que sería acorde al argumento de la habitación china); la segunda, considera totalmente posible el pensamiento en las máquinas y lo ubica en la misma categoría del pensamiento humano (consonante con el argumento de Hofstadter); la tercera, implica aceptar el desafío de Turing como una metáfora y, por

tanto, menguar de alguna forma la rigidez de una pretendida identidad entre el pensamiento humano y el maquinal, haciendo brumosa la frontera entre estos y aprovechando la analogía para el estudio de la mente humana. La cuarta respuesta corresponde, en palabras de Rivière (1988), a una posición matizada ante el desafío, en la cual se considera a la mente como un sistema de computo, pero no del tipo que propone Turing, sino como un sistema acorde a ciertas propiedades especificas del sistema nervioso humano. Consideremos un poco más de cerca las tres opciones que admiten la existencia de inteligencia artificial.

La aceptación total de la posibilidad de pensamiento artificial, v*ersión fuerte de la metáfora del ordenador* o *paradigma de cómputos sobre representaciones*, busca la explicación del conocimiento en general, profesando la identidad del pensamiento hombre-máquina y radicando su fundamento en operaciones formales sobre símbolos (de ahí la acepción de paradigma de cómputos…) (De Vega, 1984; Rivière, 1991). De acuerdo con esta línea teórica, el pensamiento podría identificarse con un tipo de estructura compleja algorítmica, que permitiría la resolución de problemas abstractos y cotidianos. Para alcanzar sus objetivos, los defensores de la versión fuerte de la metáfora del ordenador se valen del "modelado computacional", el cual consiste, *grosso modo*, en programar una máquina para que realice procesos de conocimiento comunes en las personas (Eysenck y Keane, 2000; Gardner, 1985); sin embargo, para lograr esto, no se circunscriben a las restricciones psicológicas características del sistema cognitivo humano, por lo cual, igualar o mejorar el proceso o el resultado mediante artilugios informáticos, les viene bien.

La aceptación restringida de la posibilidad de inteligencia artificial, *versión débil de la metáfora del ordenador* o *paradigma del procesamiento de información*, busca explicar el conocimiento psicológico en específico, sin la pretensión de generalizar su teoría a la cognición en general (De Vega, 1984; Rivière, 1991). Así como los teóricos de la metáfora fuerte hacen uso del modelado computacional, el paradigma del procesamiento de información se sirve de la *simulación*, la cual, si bien busca implementar procesos de conocimiento en máquinas para develar el funcionamiento cognitivo, respeta las restricciones psicológicas y se sirve de estudios de la actuación humana, esto es, de teorías psicológicas (método tal que es obviado por el paradigma de cómputos sobre representaciones) (Eysenck y Keane, 2000; Gardner, 1985). Esta forma de entender la inteligencia artificial se constituye como una influencia teórica contundente en la psicología cognitiva contemporánea.

La cuarta opción ante el desafío de Turing, que considera las especificidades del sistema nervioso humano en la investigación de los procesos de conocimiento, se conoce como *conexionismo*, y goza de buena reputación en el ámbito científico moderno. A diferencia del tradicional procesamiento de tipo serial (en serie), el conexionismo se ha caracterizado por postular un manejo de información simultáneo y paralelo, lo que implica la posibilidad de realizar varios procesos al mismo tiempo, e incluso varias fases de un mismo proceso en simultáneo. Diferenciándose aun más de las teorías anteriormente citadas, el conexionismo no se basa en representaciones simbólicas, sino en patrones de activación en redes de nodos (que equivaldrían en manera aún algo confusa, a las neuronas) (Haberlandt, 1997; Tienson, 1995).

De acuerdo con lo anterior, se busca una síntesis equilibrada entre procesos operacionalizables computacionalmente y características conocidas del *hardware* humano, es decir, su biología (para un acercamiento más sustancioso al conocimiento conexionista véase, por ejemplo, la recién citada *Introducción al conexionismo* de J. Tienson). Ya referidas algunas de las posibles respuestas al desafío de Turing, continuemos con los efectos que el reto del matemático causó en la comunidad académica.

Estibadores virtuales, paranoicos aparentes y otras particulares consecuencias del desafío de Turing en el desarrollo de la inteligencia artificial

En este punto nos es imposible proseguir, sin una definición de inteligencia artificial (I.A.), así que brindaremos la siguiente:

> la I.A., consiste en producir, en un ente no-humano, y ante un estimulo especifico, una respuesta que al ser dada por una persona, se consideraría inteligente. (Gardner, 1985; Simon y Kaplan, 1989)

Es esta definición sencilla, práctica y usual, la que nos hacía falta. En cuanto al albor de esta disciplina (la I.A.), señalaremos que estuvo nutrido por diversos saberes, entre los cuales resaltan la *teoría de la comunicación,* de Claude Shannon, que vio la luz pública en 1948; la *teoría cibernética,* de Norbert Wiener, de la década del 40; los estudios psicolingüísticos, liderados por Noam Chomsky, académico de creciente relevancia a partir de mediados del siglo pasado; y en gran medida, por el desafío de Turing; el cual, según algunos teóricos, constituye el nacimiento mismo de la inteligencia artificial (como Simon y Kaplan, 1989; de quienes tomamos también las referencias sobre las disciplinas que produjeron el nacimiento de la I.A.).

Ya sea que adoptasen la versión débil o fuerte de la metáfora del ordenador, varios científicos se vieron muy influenciados por el planteamiento de Turing, entre ellos, los pioneros de la inteligencia artificial: John McCarthy, Marvin Minsky, Allen Newell y Herbert Simon (todos investigadores norteamericanos), quienes se reunieron en 1956 en el *Simposio de Dartmouth* (que constituye un hito en la creación de la I.A.), En este encuentro se discutió sobre las bases de la nueva ciencia, y se acuño el término "inteligencia artificial" (Eysenck y Keane, 2000; Gardner, 1985). También en 1956, durante el *Simposio sobre la teoría de la información*, Newell y Simon presentaron su "máquina de la teoría lógica", bautizada *Johniac*,[2] la cual realizó con éxito la resolución de uno de los teoremas, que ya habían sido resueltos por Alfred Whitehead y Bertrand Russell; sin embargo, aunque el teorema fue resuelto en forma más *elegante* que la de Whitehead y Russell, según refiere Gardner (1985), no se aceptó su publicación, debido a la autoría robótica.

Con base en los planteamientos expuestos y las conclusiones alcanzadas durante estos encuentros, en épocas subsiguientes se creó *software* computacional que realizaba interesantes y curiosas labores, del cual se suele destacar: el programa *Eliza*, realizado en la década del 70 por el científico alemán, recientemente fallecido, Joseph Weizenbaum. Dicho programa consistía en la simulación (o parodia) de un terapeuta de corte rogeriano; el *SHRDLU*, de Terry Winograd, desarrollado por la misma época que el Eliza, se dirigía, mediante un reducido número de órdenes por escrito, a un acomodador de figuras geométricas, en un pequeño mundo virtual de bloques; y por supuesto, el inquietante *Parry*: paranoico aparente, profundamente consternado por la mafia y las carreras de caballos, escrito (es decir, creado como programa) en los tempranos 70, por el psiquiatra norteamericano Kenneth Colby (Copeland, 1996; Gardner, 1985; Simon y Kaplan,1989).

El funcionamiento de estos programas, Parry y Eliza en específico, no era en ningún sentido misterioso. Entre las acciones para las que estaban programados, conocidas como *proceso de comparación de patrones* (Copeland, 1996), estaba la de detectar palabras específicas en las oraciones, y contestar, tomando la primera de una lista de respuestas predeterminada; estas respuestas estaban asociadas en forma coherente con la palabra elegida, o consistían en una simple transformación sintáctica de la frase de entrada. La frase con que se respondía era ubicada luego al final de la lista, de forma tal que se agotase determinado número de respuestas, antes de repetirse. Parry, por ejemplo, emitía una respuesta acalorada cuando detectaba una increpación de paranoico en la charla de su interlocutor.

Estos programas también podían trasformar pronombres tales como *tú* o *mí*, en *yo* o *tú*, respectivamente, así como modificar la sintaxis de la oración entrante, para crear, mediante la oración

[2] En deferencia al matemático húngaro John von Neumann, quien tuvo gran influencia en la creación de los procesos mnémicos computacionales, el *software* auto-modificable, el procesamiento serial y los cómputos sobre símbolos aplicados a la computación (Simon y Kaplan, 1989).

de salida, una ilusión de entendimiento.[3] Eliza, por su parte, retenía frases encabezadas por *mío* o *mí*, escritas por el interlocutor, las etiquetaba, y las usaba luego, tomándolas como frases de contenido especialmente significativo para los pacientes (de acuerdo con la programación del operador, claro está). Para los casos en que no se detectaba un patrón específico al cual contestar, contaba Eliza con frases de cajón como: "¿Qué te hace pensar eso?" (Copeland, 1996) (si acaso el lector se ha visto interesado por una conversación de este tipo, no hace falta más que navegar en la red, en donde se encuentran disponibles al público diversas versiones de estos programas).

Aunque el proceso de funcionamiento de los primeros programas de I.A. no es ningún arcano, el desempeño de los mismos resulta desconcertante y atrayente, pero causa aun más desconcierto todo el asunto de la creatividad en las máquinas. Considérese que varios autores de la psicología de la creatividad, como Howard Gardner, Robert Weisberg y Margaret Boden, comparten algo conocido como la concepción "más de lo mismo", la cual consiste en el planteamiento de que no hay nada especial o místico en el trabajo creativo de la mente humana, sino más de lo mismo, de los procesos que utilizamos habitualmente (Romo, 1997). En este sentido, no es de extrañar que existan programas que realicen obras de arte pictóricas, musicales o literarias; programas que descubran leyes científicas, e incluso, uno que otro que publique un artículo (todo esto avalado por el ámbito al que pertenecen las creaciones) (Boden, 1994).

Para la muestra, un botón. He aquí las primeras líneas, respetando el idioma vernáculo, del libro de prosa y poesía, de Racter (abreviación de *raconteur*, término galo para narrador de cuentos): programa de inteligencia artificial creado por William Chamberlain y Thomas Etter, quienes le atribuyen la autoría del escrito: "*At all events my own essays and dissertations about love and its endless pain and perpetual pleasure will be known and understood by all of you who read this and talk or sing or chant about it to your worried friends or nervous enemies*" (Racter, 1984; sin numeración de pagina o puntuación en el original).

Con respecto a la creatividad en la inteligencia artificial, el investigador croata Mihály Csikszentmihalyi (1998) señala que a los computadores les es dada la información y las variables específicas por parte de los científicos; que esta información sirve como pábulo de sus creaciones, y que esto no sucede en la vida real. Por el contrario, algunos autores consideran que la creatividad es un proceso no-exclusivo de los humanos, e incluso consideran que las creaciones artísticas computacionales zanjan otro tanto la brecha que aleja a los hombres de las máquinas; al respecto, afirma la investigadora en inteligencia artificial Margaret Boden (1994) que la creatividad computacional no amenaza nuestro auto-respeto, ya que el hecho de compartir procesos no termina por igualar a los seres que los realizan.

En las teorías de la mente

La influencia del desafío de Turing no fue exclusivamente sobre el trabajo en la inteligencia artificial; se dio también en teóricos del funcionamiento de la mente humana, quienes propusieron ideas de suma relevancia, como la de *arquitectura mental*, del filósofo norteamericano Jerry Fodor (1968, 1986, 1997). Fodor, claro seguidor de la metáfora fuerte, propone la organización funcional de la mente, mediante la siguiente división: transductores sensoriales, sistemas de entrada y sistemas centrales. Los transductores tomarían, en forma pre-conceptual, información sobre los estímulos del ambiente, para pasarla luego a los sistemas de entrada. Estos serían perceptivos y tratarían la información algorítmicamente, enviándola luego a los sistemas centrales (el pensamiento y la resolución de problemas) (Fodor, 1986).

[3] Por ejemplo, en el caso de Eliza, una posible respuesta a "*Tú me* odias" sería "Te gusta pensar que *yo te* odio, no es cierto", transformando la oración que contiene: "*Tú me"* por una pre-establecida que contiene: "*Yo te",* modificando a su vez la conjugación del verbo (Copeland, 1996).

Lo más curioso de la tesis de Fodor es la definición que realiza de los sistemas de entrada y de los sistemas centrales: los primeros, responderían a la característica de ser modulares, lo que implica funcionar de forma algorítmica, automática y sumamente eficiente, en un dominio especifico y con un tipo de información que no se compartiría con otros módulos. Estos sistemas modulares de entrada reconocerían solamente los estímulos que les es propio tratar y aplicarían procesos formales sobre ellos, de forma súper rápida, eficiente, obligatoria y, de cierta forma, obtusa, ya que harían lo que tienen que hacer, siempre de la misma forma, y lo que es aun más importante, *sin que nuestra conciencia tuviese acceso a sus procedimientos, ni pudiese modificarlos; solo contemplar al resultado del cómputo de información, que llega del ambiente* (Fodor, 1986).

Los sistemas centrales, según Fodor encargados del pensamiento y la resolución de problemas, funcionarían bajo las características que este filósofo denomina *isotropía* y *quineanismo*, que básicamente se refieren al inconveniente de definir qué información es relevante para aplicar en una situación problemática, o para comprobar una hipótesis dada, y qué información se afecta y modifica después de un proceso de conocimiento (Fodor, 1986).

Estas supuestas características del funcionamiento del pensamiento humano (isotropía y quineanismo) guardan preocupante similitud con el llamado "problema del marco", propio del ámbito computacional, y que según refiere John Tienson (1995) fue una de las razones de la crisis de la *buena y anticuada inteligencia artificial* (BAIA, como la denomina Haugeland). El *problema del marco* consiste justamente en la extrema dificultad de que una máquina, o su operador, establezca el conjunto de información que se debe considerar antes y durante una acción específica, así como el conjunto de información que se ve modificada luego de algún proceso. En este punto no es difícil notar cómo el funcionamiento especifico de una máquina, y en este caso una restricción procedimental o de implementación, se extrapola de forma tal que *se le confieren a la mente humana propiedades y restricciones propias de las máquinas*.

La propuesta de Fodor ha dado paso a lo que se conoce como *arquitecturas mentales post-fodorianas* (Igoa, 2003), como las *inteligencias múltiples* de Gardner, la *mente computacional* de Jackendoff o la escuela de la *modularidad masiva*. Estas formas de explicación de la mente, basadas en la metáfora del ordenador, proponen una particular partición de la mente que contempla una gran variedad de componentes, entre los que podemos contar: de lenguaje, visión y musical; perceptivos; de inteligencias de diversos tipos (entre los más conservadores modularmente hablando); o módulos innatos de física, biología y psicología; e incluso, módulos a granel del tamaño de conceptos (Cosmides y Tooby, 2002; Fodor, 1986; Hirschfeld y Gelman, 2002; Jackendoff, 1987; Karmiloff-Smith, 1994; Pinker, 2001).

De estas arquitecturas, la propuesta taxonómica de la mente, realizada por Ray Jackendoff (1987), ha sido de particular importancia en la psicología moderna. Consiste, básicamente, en la diferenciación entre *mente fenomenológica* y *mente computacional*. En la primera residirían las ilusiones, sensaciones, imaginaciones y la conciencia de uno mismo, mientras que la segunda se encargaría del reconocimiento, comparación, análisis y demás procesos de conocimiento, los cuales *estarían por fuera del alcance consciente*. Reflexiones de esta índole, sumadas a numerosas investigaciones prácticas en psicología, han llevado a postular un supuesto muy común en las explicaciones contemporáneas sobre la mente, según el cual, *la mayoría de los procesos del conocimiento suceden a un nivel no consciente, en ocasiones llamado: inconsciente Cognitivo* (LeDoux, 1999).

Es así como tenemos que en las teorías actuales sobre la mente se evidencia, cada vez con mayor fuerza, una clara diferencia e independencia entre la mente que realiza los procesos de conocimiento y la mente que comporta la conciencia (Froufe, 2004). Esta última (la conciencia) ha ido perdiendo paulatinamente su relevancia en el terreno del pensamiento (en términos teóricos por supuesto); de hecho, existen teorías como la del filósofo americano Richard Rorty (1989), quien propone el abandono de la investigación y discusión de los concep-

tos relativos a la conciencia fenomenológica, para centrarse en el estudio neurofisiológico, buscando con ello cambiar la forma intencional de comprender la mente, por una centrada en su constituyente neural; incluso, propone Rorty que así pervivan los términos referentes a la mente fenomenológica (los términos mentalistas propiamente hablando, como: desear, querer o intuir), los estados del sistema nervioso serían suficientes para explicar y entender la actividad humana.

Consideraciones como las precedentes sobre el carácter computacional único de los procesos de conocimiento, la imposibilidad de comprender fenómenos que no sean formulados en términos de cómputos sintácticos sobre representaciones, o la futilidad de los estados conscientes, constituyen un ejemplo de las repercusiones de una extrapolación de concepciones sintáctico-computacionales a la explicación de la mente humana. Esto puede redundar (o degenerar) en teorías confusas sobre los procesos de pensamiento, la voluntad y la esencia del conocimiento; pero, un momento, tal parece que el sencillo juego planteado por Turing ha tenido algunas repercusiones inesperadas; acaso sea menester, hasta donde nos sea posible, plantear las condiciones del tipo de inteligencia que se toma hoy en día como modelo para la explicación de la mente. Ya que hemos llegado hasta aquí, ¿por qué no hacerlo?

Restricciones fundamentales de la inteligencia artificial

Ya expuestas las singularidades del tema en cuestión (quizá muy extensamente), mediante la presentación del desafío de Turing, sus consecuencias y posibles respuestas, el contraargumento de la habitación china, las razones de Hofstadter y la alusión al nacimiento y temprano desarrollo de la inteligencia artificial, estamos preparados para tratar el tema que va a englobar la reflexión buscada: las restricciones fundamentales de la inteligencia artificial. Proponemos aquí, las dos siguientes:

1. Imposibilidad ontológica de la identidad de la inteligencia artificial y el pensamiento humano, debido a génesis dispares.
2. Imposibilidad semántica de la identidad de la inteligencia artificial y el pensamiento humano, debido a categorías conceptuales dispares.

Consideraciones preliminares sobre el sentido y el engaño

En relación con el asunto del acceso al significado en la I.A., comenta Searle que "una de las afirmaciones de los propulsores de la IA fuerte [versión fuerte de la metáfora del ordenador] es que cuando yo comprendo una historia en inglés, lo que estoy haciendo es ni más ni menos lo mismo —por lo menos aproximadamente lo mismo— que lo que hacía al manipular los símbolos chinos" (1983, p. 458) (recuerda el lector el asunto de los manuscritos chinos, no es cierto). Este presupuesto de los defensores de la metáfora fuerte no es muy atinado, ya que *la "ruta" del proceso cognitivo, en el ejemplo de la habitación china, consiste en una conexión directa entre sistemas de entrada y salida, o una conexión que en cualquier caso, no pasa por el "almacén" de significados*; y es que ese es un atajo cognitivo no del todo impropio del humano.

Los investigadores en neuropsicología cognitiva Andrew Ellis y Andrew Young (1992) refieren posibles conexiones entre *lexicones* de entrada y salida que obvian el paso por el almacén semántico, ruta tal que si bien en ocasiones es funcional (para la repetición de palabras desconocidas o no-palabras, por ejemplo), puede constituir una patología, de ser el único medio usado para la comprensión o producción del lenguaje.[4] Para decirlo en forma sencilla: el recibir y responder preguntas relativas a una lluvia venidera, en forma coherente e incluso precisa, solo basándose en la estructura de las mismas y no en su significado, no implica considerar el uso de un paraguas en un futuro próximo para evadir el chaparrón. Estamos hablando pues de dos tipos de entendimiento distintos: asociativo y

[4] Recordemos que en el modelo de comprensión y producción del lenguaje presentado por estos académicos, el significado de las palabras se encuentra ubicado en un sitio distinto al de la conformación estructural-ortográfica de las mismas, ya sea en formato visual o auditivo, total que se pueden asociar palabras entre sí sin acceder a su significado (Ellis y Young, 1992).

comprensivo; el último abarca al primero, pero no al contrario. Esta cuestión de acceso al significado ha obstaculizado los avances en I.A. de tal forma que se ha considerado dotar a las máquinas de una semántica que analice imágenes y las relacione con estructuras sintácticas, almacenadas en la memoria (Sagal, 1982; Simon y Kaplan, 1989).

Existe también, en relación con la prueba de Turing, el problema de la máquina engañando, haciendo creer que es un humano; y es que este asunto del engaño no carece de profundidad, ya que presupone la capacidad del que engaña de comprender, en un nivel no necesariamente consciente, que los otros poseen representaciones mentales que son esencialmente distintas a la realidad, y que se basan en estas para conducirse; en este sentido, las representaciones del que es engañado pueden ser más o menos acordes a la situación, pudiendo por eso conducirlo a equivocaciones en su actuar.

Ángel Rivière refiere la diferencia entre un engaño, digamos automático, como el de algunos animales para despistar a sus enemigos y presas, y uno de tipo táctico (Rivière y Núñez, 1998). Este se define por su flexibilidad y su posibilidad de ser adaptado según la situación y las características del que se quiere engañar. El engaño no-táctico, entonces, sería inflexible, predeterminado y no intencional, y sucedería en seres que no poseen la capacidad de reconocer al otro como un ser de experiencia, falible y guiado por sus representaciones mentales del mundo y de los otros; seres que no poseen la capacidad que usualmente se denomina *teoría de la mente* (ToM). Sin intención de entrar en la polémica sobre la ToM, la semántica o el tipo de engaño que implicaría la táctica de una máquina para pasar el desafío de Turing, es obligado hacer notar que hace falta algo más que coherencia sintáctica para poseer pensamiento.

Imposibilidad ontológica de la identidad de la inteligencia artificial y el pensamiento humano debido a génesis dispares

El estudio de la mente y su análisis no se agotan en una visión centrada en las características actuales de la población, representada por unos pocos. Debe considerarse, además, entre otras cosas, la aproximación filogenética; es decir, el estudio de la mente desde presupuestos evolucionistas y de adaptabilidad y progresión como los que han sido defendidos por la escuela conocida como Nueva Síntesis (llamada así debido a su concreción de teorías cognitivas con la teoría evolucionista de Darwin) (Mithen, 1998; Pinker, 2001). De acuerdo con las investigaciones de esta escuela, la especificidad genética innata interactúa con las características ambientales haciendo posible la modificación modular (Fernández y Ruiz, 1990).

De acuerdo con lo anterior, la mente, incluso, puede haber venido mutando de tal forma que recursos que usamos hoy día en un proceso X, pudiesen ser modificaciones de otros usados originalmente en un proceso Y (Cosmides y Tooby, 2002). Esta transformación no solo seria filogenética, ya que podría también presentarse en forma progresiva en el transcurso de una vida, mediante, por ejemplo, el proceso que Karmiloff-Smith (1994) denomina *redescripción representacional*. Durante este proceso de cambio conceptual, los formatos representacionales, esto es, las formas en que está representado el conocimiento, van modificándose hasta alcanzar niveles cuyas características constituyentes son disímiles de las de sus predecesores.

Desde otro ámbito de la psicología, se afirma que, *dependiendo de la organización de un sistema y de la complejidad de sus relaciones, pueden emerger cualidades que no existen por sí solas en los elementos constituyentes de un conjunto* (Morin, 1999); es decir, las relaciones estructurales y funcionales de un conjunto son más que la suma de las características estructurales y funcionales de sus constituyentes. Cuando las cualidades de los componentes interactúan, crean fenómenos que responden específicamente a la dinámica del grupo en que se presentan. En sistemas complejos, es decir, con gran cantidad de componentes y relaciones, los fenómenos resultantes de la interacción de los elementos del grupo pueden generar características definitorias del propio sistema (es siguiendo esta lógica que Searle propone a la intencionalidad como propiedad emergente de la biología humana, y Hofstadter la propone como propiedad emergente de la conformación estructural del organismo).

Desde estos presupuestos (la modificación modular, la redescripción representacional y las propiedades emergentes), *el pensamiento podría considerarse como una propiedad emergente del sistema humano en su totalidad, que se ha constituido con el paso del tiempo en lo que es hoy: una cualidad* (por nombrarlo de algún modo) *que responde a las particularidades del hombre como especie y a las de la historia que ha transitado*. Para Hofstadter (1983), las cualidades intencionales serían características que emergerían en la máquina de la complejización de sus relaciones funcionales internas. En este sentido, es oportuno aclarar que si la emergencia de cualidades está basada en las relaciones generales que posee un sistema, la aparición de las características específicas presentes en el humano respondería a un largo y complejo proceso de interacción entre sus constituyentes biológicos, culturales y ambientales (entre tantos otros).

Asimismo, las características que pudiesen emerger de las relaciones de los componentes de las máquinas responderían a las especificidades de su *hardware*, su ambiente, su historia y, por supuesto, ¡sus creadores! Es en este sentido que el argumento de Hofstadter se queda corto, ya que aun aceptando que la interacción de características en la máquina redundara en la emergencia de propiedades novedosas, no hay ninguna razón para suponer que estas serían idénticas a las que han emergido en los humanos a través de su historia. Señala también Hofstadter (1983, pp. 115-116): "Lo que pasa además es que los diseñadores humanos van a acelerar el proceso evolucionario al dirigirse en forma deliberada a la meta de crear inteligencia…".

Este hecho implica que en la evolución y génesis de la máquina, a diferencia de la nuestra, la participación de los creadores es clara. Sería quizá insensato ignorar estos factores preponderantes en la aparición de las propiedades emergentes en las máquinas, que imposibilitan la identidad del pensamiento humano y la inteligencia artificial desde la constitución misma de su esencia. No significa esto que las características que pueda llegar a presentar una máquina, bien sea por su constitución o por la relación compleja de sus componentes, no puedan ser similares en sus efectos o procesos a las humanas; simplemente significa, que no son, o serán, idénticas. Para lograr un repertorio de características mentales iguales a las de los humanos, se requeriría la creación de un ser con características constituyentes iguales a las humanas, y este asunto, por métodos distintos al tradicional, acaso se nos complicase un poco.

Imposibilidad semántica de la identidad de la inteligencia artificial y el pensamiento humano debido a categorías conceptuales dispares

Para Turing, la mejor estrategia que pudiese usar una máquina sería "la de intentar el logro de respuestas como las que naturalmente daría un hombre" (1981, p. 72); es decir, la máquina debería *imitar* con precisión la conducta verbal humana (y eventualmente también las acciones). Sin embargo, *imitar implica la irrealidad de lo que imita frente a lo imitado*; veamos cómo.

De acuerdo con Chomsky (1980, 1988), existe cierta clase de categorías semánticas innatas, categorías conceptuales predeterminadas en el humano que se etiquetan por medio del lenguaje y significan lo mismo universalmente; aun siendo nombradas de forma distinta por distintas culturas, estas categorías implicarían la existencia de estructuras conceptuales similares en las distintas personas del mundo; de ahí lo que este autor denomina "verdades de significado" (1988, p. 35). Estas *verdades de significado* no dependerían de la experiencia para su constatación, es más, ni siquiera requerirían constatación, ya que oficiarían como pilares categoriales del proceso de adquisición de conocimiento.

Esta idea de conocimientos y categorías preestablecidas en la mente es también un presupuesto fundamental para los teóricos de la modularidad masiva, quienes refieren, con amplio apoyo empírico, la existencia de conocimiento (nociones o preferencias) de tipo físico, biológico, psicológico y de otras variadas índoles. Estos conocimientos estarían presentes en los recién nacidos (o en todo caso a tempranísima edad) y constituirían el fundamento para la consolidación del conocimiento. Estos autores afirman, además, que *una categoría de conocimiento innata al humano le permite dis-*

tinguir entre seres animados e inanimados o vivos y no-vivos (Cosmides y Tooby 2002; Pinker, 2001).

Con base en este principio y en las *verdades de significado* de Chomsky *podemos entender las categorías semánticas de simulación y simulado como dispares, por el hecho mismo, tan evidente que se obvia, de que la simulación por definición requiere lo simulado para existir; su esencia se basa en tratar de ser algo que no es: en aparentar*; en otras palabras, la simulación es irreal en el sentido de la realidad de lo que simula: desde estas premisas, *la inteligencia artificial representa una fantasmagoría*. Aclaremos que bajo ninguna circunstancia estamos en una discusión de términos de cómo calificar lingüísticamente la inteligencia artificial; el punto aquí recae en las diferentes esencias del pensamiento humano y el pensamiento artificial, ocupando distintas categorías conceptuales en la clasificación propia de la mente humana; esto debido, entre otras cosas, a que las génesis de ambas son claramente dispares.

Ahora, aunque esta negación de la identidad entre el pensamiento humano y la I.A. pueda parecer vana, no lo es; téngase en cuenta que pioneros de la I.A., como Allen Newell y Herbert Simon, "escriben que el tipo de conocimiento cuya existencia ellos afirman en las computadoras es exactamente el mismo que el de los seres humanos." (Searle, 1983, p. 461); y no se olvide tampoco de las repercusiones que puede traer, y de hecho ha traído, en la concepción contemporánea de la mente, el forzado parangón entre esta y una máquina computarizada. Así que si tomar el pensamiento humano por único, o como el pensamiento por antonomasia, resulta ingenuo, lo es también considerar que el pensamiento de las máquinas, en algún tramo de su periplo evolutivo, resulte ser idéntico al nuestro.

La inteligencia artificial llegara con seguridad, si es que no lo ha hecho ya, a superar la prueba de Turing; es de esperar también que se logre simular, de alguna forma, la intencionalidad y la conciencia en cerebros de silicio; pero, aun así, las propiedades de estas máquinas no serán las mismas que las nuestras, ya que entre la complejidad de la interacción de sus componentes constitutivos, se contarán factores distintos a los que cuentan en nosotros; el pensamiento de las máquinas, si quiere llamársele así, va a pertenecer, o pertenece, a categorías primordialmente distintas.

Referencias

Acero, J. J. (1995). Teorías del contenido mental. En F. Broncano (Ed.), *La mente humana* (pp. 175-206). Madrid: Trotta.

Bechtel, W. (1991). *Filosofía de la mente. Una panorámica para la ciencia cognitiva.* Madrid: Tecnos.

Boden, M. (1994). *La mente creativa: mitos y mecanismos*. Barcelona: Gedisa.

Chomsky, N. (1980). On cognitive structures and their development: A reply to Piaget. En M. Piattelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 35-54). Cambridge, MA: Harvard University Press.

Chomsky, N. (1988). *El lenguaje y los problemas del conocimiento. Conferencias de Managua I*. Madrid: Visor.

Copeland, J. (1996). *Inteligencia artificial: una introducción filosófica*. Madrid: Alianza.

Cosmides, L. y Tooby, J. (2002). Orígenes de la especificidad de dominio: la evolución de la organización funcional. En L. Hirschfeld y S. Gelman (Comps.), *Cartografía de la mente*: *la especificidad de dominio en la cognición y la cultura* (pp. 132-173). Barcelona: Gedisa.

Csikszentmihalyi, M. (1998). *Creatividad. El fluir y la psicología del descubrimiento y la invención*. Barcelona: Paidós.

de Vega, M. (1984). *Introducción a la psicología cognitiva*. Madrid: Alianza.

Ellis, A. W. y Young, A. W. (1992). *Neuropsicología cognitiva humana*. Barcelona: Masson.

Eysenck, M. W. y Keane, M. T. (2000). *Cognitive psychology: A student's handbook*. Hove, UK: Psychology Press.

Fernández, P. y Ruiz, M. (Eds.) (1990). *Cognición y modularidad*. Barcelona: PPU.

Fodor, J. (1968/1980). *La explicación psicológica. Introducción a la filosofía de la psicología*. Madrid: Cátedra.

Fodor, J. (1986). *La modularidad de la mente*. Madrid: Morata.

Fodor, J. (1997). *El olmo y el experto. El reino de la mente y su semántica*. Barcelona: Paidós.

Froufe, M. (2004). Disociaciones entre cognición y conciencia. En D. A. Duarte y E. A. Rabossi (Eds.), *Psicología Cognitiva y Filosofía de la Mente: Pensamiento, Representación y Conciencia* (pp. 285-307). Buenos Aires: Alianza.

Gardner, H. (1985/1988). *La nueva ciencia de la mente*. *Historia de la revolución cognitiva*. Barcelona: Paidós.

Haberlandt, K. (1997). *Cognitive psychology* (2nd. edition). Boston: Allyn and Bacon.

Hirschfeld, L. y Gelman, S. (2002). Hacia una topografía de la mente: una introducción a la especificidad de dominio. En L. Hirschfeld y S. Gelman (Comps.), *Cartografía de la Mente* (pp. 23-67). Barcelona: Gedisa.

Hofstadter, D. (1983). Temas metamágicos bizantinos. El test de Turing: conversación en un café. En D. Hofstadter y D. Dennett (Comps.), *El ojo de la mente: fantasías y reflexiones sobre el yo y el alma* (pp. 90-125). Buenos Aires: Sudamericana.

Igoa, J. M. (2003). Las paradojas de la modularidad. *Anuario de Psicología*, *34* (4), 529-536.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1994). *Más allá de la modularidad*. Madrid: Alianza.

LeDoux, J. (1999). *El cerebro emocional*. Buenos Aires: Planeta.

Mithen, S. (1998). *Arqueología de la mente*. Barcelona: Crítica.

Morin, E. (1995). *Introducción al pensamiento complejo*. Barcelona: Gedisa.

Pinker, S. (2001). *Cómo Funciona la Mente*. Barcelona: Destino.

Racter (1984). *The Policeman's beard is half constructed. Computer prose and poetry*. New York: Warner Books.

Rivière, A. (1991). *Objetos con mente*. Madrid: Alianza.

Rivière, A. y Núñez, M. (1998). *La mirada mental* (2ª ed.). Buenos Aires: Aique.

Romo, M. (1997). *Psicología de la creatividad*. Barcelona: Paidós.

Rorty, R. (1989). *La filosofía y el espejo de la naturaleza*. Madrid: Cátedra.

Sagal, P. T. (1982). *Mind, man, and machine*. Indianapolis: Hackett Publishing Company.

Searle, J. (1983). Mentes, cerebros y programas. En D. Hofstadter y D. Dennett (Comps.), *El ojo de la mente: fantasías y reflexiones sobre el yo y el alma* (pp. 454-493). Buenos Aires: Sudamericana.

Simon, H. y Kaplan, C. (1989). Foundations of Cognitive Science. En M. Posner (Comp.), *Foundations of Cognitive Science* (pp. 1-47). Cambridge, MA: MIT Press.

Tienson, J. (1995). Una introducción al conexionismo. En Rabossi, E. (Comp.), *Filosofía de la mente y ciencia cognitiva* (pp. 359-380). Barcelona: Paidós.

Turing, A. (1950/1983). Los aparatos de computación y la inteligencia. En D. Hofstadter y D. Dennett (Comps.), *El ojo de la mente: fantasías y reflexiones sobre el yo y el alma* (pp. 69-89). Buenos Aires: Sudamericana.

Fecha de recepción: abril de 2008
Fecha de aceptación: agosto de 2008