ISSN: 1576-3935

Monismo anómalo, intencionalidad, falacias mentales e inteligencia artificial

Camilo Andrés ORDÓÑEZ PINILLA

Universidad Nacional de Colombia

Recibido: 12/09/06
Aprobado: 20/12/06

Introducción

La Inteligencia Artificial (IA) es uno de los proyectos más importantes en los que la Filosofía actualmente tiene participación. Como lo han señalado muchos críticos, el carácter necesariamente intencional de lo mental parece ser un obstáculo insalvable para el éxito del proyecto de la IA. En el presente texto se buscará mostrar que a partir de la tesis del Monismo Anómalo de Donald Davidson es posible articular una propuesta que permita caracterizar de una manera intencional los sistemas de la IA y por tanto garantizar su éxito. Con este fin se buscará mostrar que la apelación al carácter intencional de la mente es un

elemento de tipo heurístico y se propondrá que la postulación de estados intencionales desde la perspectiva de primera persona corresponde a una *falacia mental* con la que el ser humano se auto-engaña, al creer que posee tales estados.

Mente y cuerpo

Una de las tareas centrales de la Filosofía analítica contemporánea ha sido responder las preguntas acerca de la mente. Los filósofos de la mente intentan entender, básicamente, qué es la mente y cómo funciona. Si bien la diversidad de posturas y opiniones acerca de la mente es amplia, un lugar común en la Filosofía de la Mente contemporánea ha sido el rechazo al dualismo cartesiano. Dentro de tal espíritu se enmarca la postura denominada *Monismo Anómalo*. Postulado por Donald Davidson, el Monismo Anómalo es una doctrina que afirma cierta clase de identidad entre los estados-procesos mentales y los estados-procesos cerebrales. Esto hace que se circunscriba dentro de la idea general de la *Teoría de la Identidad de la Mente*. El Monismo Anómalo afirma que si bien los eventos mentales *son* eventos físicos, cuando hay una descripción mental (psicológica, en tanto usa términos psicológicos) de un evento, tal descripción no puede ponerse dentro del *corpus* de una teoría estricta, en el sentido en que se usa el término 'teoría' en las Ciencias Naturales. Específicamente, el Monismo Anómalo afirma que las descripciones psicológicas no pueden ponerse dentro de una teoría determinista.

Inteligencia Artificial

El proyecto de la Inteligencia Artificial[1] es uno de los proyectos más ambiciosos con los que se ha comprometido la especie humana. Además de ambicioso, tal proyecto es también uno de los más interdisciplinarios de la historia. La Inteligencia Artificial se nutre de los conocimientos aportados por áreas del conocimiento tan distintas como la Lingüística, la Neurología, las Matemáticas y la Ingeniería. Uno de los campos académicos que realiza un aporte central al desarrollo de la Inteligencia Artificial es la Filosofía, especialmente la Filosofía de la Mente, en tanto ella busca explicar qué debe entenderse por la mente y los aspectos filosóficos de su funcionamiento. De las objeciones que se realizan al proyecto de la Inteligencia Artificial, tal vez las más interesantes son las que atacan sus fundamentos filosóficos. Esto, en tanto los defensores de la Inteligencia Artificial deben enfrentarse a desafíos conceptuales para garantizar el éxito de su proyecto.

Inteligencia Artificial y Monismo Anómalo

1 Un punto importante a este respecto es que no hay acuerdo acerca de cuál es el proyecto de la Inteligencia Artificial. Por un lado está la versión *tecnológica*, que considera a la Inteligencia Artificial como una disciplina que busca construir dispositivos que hagan las cosas que se supone hace la mente y por otro lado está la versión *teórica*, que considera a la Inteligencia Artificial como una disciplina que estudia la inteligencia en general, esto es, una disciplina que explica el fenómeno de la inteligencia y la intencionalidad, independientemente de que se busque o no construir un sistema del que puedan predicarse tales características. Dentro de la versión tecnológica hay, a su vez, dos vertientes. Tales vertientes son denominadas por John Searle [Searle, 1980] *Inteligencia Artificial débil*, la cuál busca construir dispositivos que hagan lo que hace la mente, bien utilizando los mecanismos de la mente (*simulación*) o mecanismos diferentes, e *Inteligencia Artificial fuerte*, en la que se buscan construir dispositivos que *sean* una mente (*duplicación*). Cuando aquí se hable de Inteligencia Artificial se hará en un sentido tecnológico, bien sea fuerte o débil; y entendiendo el término 'débil' sólo en el sentido en el que el dispositivo de Inteligencia Artificial busca *simular* el funcionamiento de una mente mediante la simulación de los mismos procesos que se dan en ella.

De los argumentos de carácter conceptual que se esgrimen en contra de la posibilidad de éxito de la Inteligencia Artificial, tal vez el más importante es el que apela al carácter necesariamente intencional de la mente y a la imposibilidad de reproducir esa característica en los sistemas que crea la Inteligencia Artificial. La tesis central que se intentará mostrar en el presente escrito será que utilizando la visión de la mente que se desprende del Monismo Anómalo puede darse una respuesta satisfactoria al argumento mencionado en contra de la Inteligencia Artificial. Específicamente se intentará mostrar que, a la luz de la doctrina del Monismo Anómalo, el carácter necesariamente intencional de la mente debe entenderse desde un punto de vista semántico y que interpretado de esa manera es posible proponer una manera plausible de predicar la intencionalidad de los sistemas de Inteligencia Artificial.

Para intentar argumentar tal tesis se propone el siguiente esquema de trabajo: Primero, se buscará exponer de una manera general la doctrina del Monismo Anómalo. Los puntos principales de la exposición serán la caracterización de tal doctrina en el contexto del Materialismo, la realización de un análisis somero del argumento que ofrece Davidson a favor de tal doctrina y la postulación de las tesis del Monismo y de la Anomalía, haciendo énfasis en que con tales tesis se separan los ámbitos ontológico y semántico acerca de lo mental. Segundo, se hará una exposición del argumento Standard en contra de la Inteligencia Artificial y se buscará mostrar que tal argumento debe entenderse en un plano puramente semántico. Tercero, se caracterizará de una manera un poco más precisa el argumento Standard en términos semánticos y se intentará mostrar en qué sentido la versión semántica del argumento podría dar razones en contra del éxito del proyecto de la Inteligencia Artificial. Cuarto, se buscará hacer una réplica a tal versión semántica del argumento. En tal réplica se intentará establecer que el argumento Standard depende de la tesis de la caracterización necesariamente intencional de lo mental y se buscará mostrar que la postulación de tal tesis obedece a una motivación de tipo heurístico; motivación que podría tenerse acerca de los sistemas de Inteligencia Artificial. Esto falsearía una de las premisas de la versión semántica del argumento Standard. Por último se bosquejará e intentará responder a una objeción a la réplica propuesta; objeción que parte del uso común del lenguaje.

Antes de empezar es necesario hacer una aclaración preliminar. El presente escrito busca, a partir de suponer la tesis del Monismo Anómalo, responder a una crítica Standard a la Inteligencia Artificial. La tesis se supone, puesto que sería imposible aquí (y tal vez en cualquier otro lado) mostrar concluyentemente que el Monismo Anómalo es la doctrina adecuada para interpretar la mente. Pero, podría sugerirse lo siguiente para aludir a que el Monismo Anómalo es un paradigma adecuado acerca de la mente. Por un lado existe un acuerdo en la comunidad académica (científica y filosófica) de que debe rechazarse el dualismo sustancialista cartesiano entre la mente y el cuerpo. Este debe ser un componente central de cualquier teoría acerca de la mente y el Monismo Anómalo lo posee. Pero por otro lado, y como se verá más adelante en la sección V, ciertos eventos humanos parecen recibir un tratamiento correcto sólo cuando se describen en términos intencionales (psicológicos). El Monismo Anómalo garantiza la posibilidad de tal tratamiento correcto de esos eventos. De esta manera, el que por un lado represente un rechazo al dualismo sustancialista cartesiano, pero el que también deje abierta la posibilidad de tratar ciertos eventos humanos considerándolos intencionales, eventos que no estarían bien caracterizados si no se hace de esa manera, hace que el Monismo Anómalo sea una doctrina plausible acerca de la mente.

Monismo anómalo[2]

Se había mencionado que la piedra de toque de la Filosofía de la Mente contemporánea era el rechazo al dualismo cartesiano entre lo mental y lo físico. Una de las posturas más importantes al respecto era la de las *Teorías de la Identidad de la Mente* (TIM). El punto en común que tienen las TIM es la postulación de la identidad entre estados-procesos mentales y estados-procesos cerebrales. Las TIM establecen que los estados-procesos mentales, como tener experiencias perceptuales, *son* procesos cerebrales. Esta es una identidad expresa, es decir, no debe entenderse en el sentido de los estados-procesos mentales se *correlacionan* con estados-procesos cerebrales, sino que *son* estados-procesos cerebrales.

En la tradición filosófica se ha planteado que el *Materialismo*[3] es la opción más adecuada para entender los estados-procesos mentales desde la perspectiva de una TIM. Para el materialismo los estados-procesos mentales son estados-procesos cerebrales *actuales*. Así, la perspectiva materialista podría entenderse como afirmando que en un estricto sentido ontológico *los estados mentales no existen*; en el mundo sólo existen estados cerebrales. De esta manera, el materialismo postularía que ciertas oraciones de identidad entre estados-procesos mentales y estados-procesos cerebrales son verdaderas. Estas oraciones tendrían la forma «Tal estado-proceso mental *es* tal estado-proceso cerebral». La posibilidad de la verdad de tales oraciones es argumentada por los materialistas, en la mayoría de los casos, apelando a una distinción análoga a la distinción sentido-referencia, que se encuentra en la teoría semántica de Gottlob Frege. De esta manera, los dos términos constituyentes de tales oraciones de identidad (el que se refiere al estado-proceso mental y el que se refiere al estado-proceso cerebral) tendrían la misma referencia, sólo que presentada de una manera diferente.

La doctrina de Donald Davidson en cuestión —*Monismo Anómalo* (MA) — se concibe en el contexto de las TIM, en tanto plantea cierta clase de identidad entre estados-procesos mentales y estados-procesos cerebrales. En lo siguiente se explicará con un poco más de detalle cómo se concibe la identidad en el MA. Con este fin, se buscará mostrar qué significa e implica la identidad y se intentará caracterizar el MA de una manera que sirva al propósito general del escrito.

Una buena estrategia explicativa del MA consistiría en hacer un bosquejo del argumento que ofrece Davidson a su favor en '*La psicología como Filosofía*'[4].

Argumento del monismo anómalo

(1) Hay relaciones causales entre ciertos eventos-procesos mentales[5] y eventos-procesos físicos del mundo y viceversa.

(2) Cuando hay relaciones causales entre eventos-procesos, entonces tales relaciones, descritas en el lenguaje adecuado, pueden integrarse en un sistema determinista cerrado.

(3) No hay leyes psicofísicas precisas.

(C) Los eventos-procesos mentales son eventos físicos. Pero, las propiedades psicológicas no son reducibles a propiedades físicas.

2 Para la siguiente exposición se seguirán principalmente las consideraciones hechas en Malpas 2003.

3 Si bien pueden plantearse diferencias importantes entre los dos términos, en el presente escrito se consideraran sinónimos los términos 'materialismo' y 'fisicalismo'.

4 Davidson [2001a]. Pág. 67-69.

5 Los términos 'mental' y 'psicológico' se usarán indistintamente.

Veamos un análisis de este argumento:

La premisa (1) se conoce como P*rincipio de interacción*. Tal principio establece que ciertos eventos[6] mentales causan eventos físicos y a su vez son causados por eventos físicos. Davidson asume esto sin justificarlo. Seria bueno aclarar un poco qué está entendiendo Davidson con la expresión 'eventos mentales' (estados y procesos mentales). Atendiendo a la cita:

> «La conclusión defendida en este trabajo […] podría formularse así: el estudio de la acción, las motivaciones, los deseos, las creencias, la memoria y el aprendizaje humanos, al menos en la medida en que se hallan lógicamente vinculados a las llamadas «actitudes proposicionales», no puede emplear los mismos métodos que, o reducirse a, las ciencias físicas, más precisas»[7].

Parece claro que para Davidson los eventos mentales son eventos describibles con las *actitudes proposicionales* –términos referidos a estados y eventos cuya descripción contiene verbos psicológicos, como creer o desear, a los que puede adscribírseles la cláusula "que" y en las que se relacionan sujetos con contenidos proposicionales. Esto, pues, según la cita, Davidson afirma que la conclusión de su artículo es que los eventos mentales que se vinculan a las actitudes proposicionales no pueden tener un tratamiento análogo al de los eventos del dominio de las Ciencias Naturales. Esta conclusión es equivalente a una parte de la tesis del MA (la parte de la tesis de la anomalía). De esta manera, la tesis del MA debe hablar de eventos mentales en el sentido de eventos relacionados con actitudes proposicionales y por tanto el argumento para probar tal tesis debe hablar de eventos mentales en el mismo sentido. Por tanto, en la presente exposición del argumento a favor de MA, cuando se hable de eventos mentales se estará hablando de los eventos mentales relacionados con las actitudes proposicionales. Cabe aclarar que las actitudes proposicionales tal vez no representan una clase exhaustiva del inventario de los eventos mentales.

Por otra parte, ¿qué son los eventos físicos?: En '*Mental Events*'[8] puede encontrarse una alusión a lo que entiende Davidson por el dominio de lo físico, cuando se afirma que el dominio de lo físico debe ser caracterizado por un lenguaje físico preciso en un sistema cerrado y en donde todo se subsuma bajo leyes estrictas de la naturaleza.

La premisa (2) se conoce como *Principio de causación legaliforme*. Este principio establece que toda interacción causal puede ponerse dentro de un sistema determinista de leyes estrictas. Esto también es asumido por Davidson sin justificación. Dejando a un lado todos los problemas que genera la postulación de tal principio relacionados con la noción de 'causa', lo que merecería un tratamiento aparte *in extenso*, hay involucradas dos nociones que merecen no ser pasadas por alto: la noción de ley estricta (formal) y de sistema cerrado. Por *sistema cerrado*, Davidson entiende algo similar a un sistema en el que se subsume una clase de eventos *E* tal que dado un evento e_1, todos los eventos que influyen en e_1 pertenecen a la clase *E*. La noción de ley estricta (formal) incluye tanto que tales leyes se den en sistemas deterministas cerrados, como que exista un criterio independiente del efecto que permita determinar si se dan o no las condiciones de aplicación. Para un ejemplo

6 De aquí en adelante, cuando se hable de eventos, tanto psicológicos (mentales) como físicos, debe entenderse que se está hablando de una noción análoga a la de estados-procesos.

7 Davidson. [2001a] Pág. 95.

8 Davidson. [2001b].

de esto último, tómese la oración "Todo cuerpo en caída libre en el vacío experimentará una aceleración de 9,8 Km. /s^{2}". Esta oración tiene forma de ley (de ley de la Física), en tanto puede analizarse en términos de una condición –"Si un cuerpo experimenta caída libre en el vacío"– y un efecto –"entonces caerá con una aceleración de 9,8 Km/s^{2}". Según la noción de ley estricta, tal oración será una oración que expresa una ley si y sólo si pueden establecerse criterios para saber si un cuerpo está cayendo en caída libre en el vacío antes de tener que medir que la velocidad de un cuerpo en caída libre en el vacío es de 9,8 Km. /s^{2}. Esto no pasa en oraciones que hagan correlaciones entre condiciones mentales y efectos conductuales, pues sólo la conducta es criterio para atribuir los estados mentales, por tanto, no es posible saber si se cumple la condición mental sino hasta que se tiene acceso a la conducta.

De poner en conjunción los principios dados por las premisas (1) y (2) se sigue lo siguiente: Dado que existe interacción causal entre eventos mentales y físicos y que tales interacciones siempre están cubiertas por leyes estrictas, se sigue que debe haber leyes estrictas que correlacionen los eventos mentales y físicos, es decir, *debe haber leyes psicofísicas*[9].

Ahora bien, tenemos la premisa (3) se conoce como *Principio de Anomalía de lo mental*. Tal principio establece que no existen leyes psicofísicas. Esto significa que no existen leyes estrictas en virtud de las cuales los eventos mentales puedan ser predichos, es decir, que los predicados psicológicos no pueden aparecer en leyes estrictas.

Así, tenemos que de las premisas (1) y (2) se sigue la existencia de leyes psicofísicas, mientras que la premisa (3) establece expresamente la inexistencia de tales leyes. Esto significa que entre las premisas del argumento se da una contradicción. De esta manera surge la pregunta por una manera en la que las tres premisas sean verdaderas al mismo tiempo.

La manera como Davidson concluye la tesis del MA a partir de tales premisas puede interpretarse como la solución a la contradicción planteada anteriormente: La conjunción de las premisas (1) y (2) establece que debe haber correlaciones entre los eventos físicos y mentales que puedan subsumirse en un sistema determinista cerrado, pero tal conclusión no hace constricciones acerca del lenguaje en el que deben establecerse tales correlaciones. Se tiene por un lado que deben establecerse correlaciones entre predicados que se refieran a lo físico y a lo mental, llámense P_1 y M_1, respectivamente. Se tiene por otro lado que los predicados P_1 y M_1 deben ser instancias de propiedades que puedan ponerse en leyes estrictas, dado que quieren establecerse correlaciones que puedan subsumirse en un sistema cerrado determinista. Como los predicados que pueden ponerse en leyes estrictas no pueden ser mentales, se sigue que los predicados P_1 y M_1 deben ser físicos. Dado que se partió de que M_1 era un predicado que se refería a una propiedad mental y luego se estableció que M_1 era un predicado que se refería a una propiedad física, de ahí se sigue que los predicados mentales son correferenciales con predicados físicos. De aquí se sigue que algunos eventos que tienen una descripción mental o instancian alguna propiedad mental son idénticos a algunos eventos que tienen una descripción física o instancian alguna propiedad física. Esta se considera la tesis del *Monismo Anómalo*. Este monismo es anómalo, puesto que si bien los eventos que instancian una propiedad mental son idénticas a eventos que instancian una propiedad física, cuando esos eventos están descritos en términos mentales (cuando se hace

9 Que las correlaciones entre eventos mentales y físicos son las leyes psicofísicas se ve claramente en la siguiente cita de Davidson [2001a] cuando dice acerca de la inexistencia de leyes psicofísicas que: «*las generalizaciones que combinan predicados psicológicos y físicos no son legaliformes en el sentido fuerte en que pueden serlo las leyes puramente físicas*» Pág. 97.

una descripción psicológica de ellos) esos eventos no son susceptibles de ser subsumidos en un sistema cerrado de leyes deterministas.

Un punto central del anterior argumento que debería tenerse en cuenta tiene que ver con la idea de que los predicados mentales no pueden aparecer en leyes estrictas. Esto está garantizado por una parte, al recordar el carácter de irreductibilidad nomológica que le adscribe Davidson a la Psicología, al atender al hecho de que al hacer atribuciones intencionales a un agente deben tenerse en cuenta los deseos y creencias de ese agente específico. Por tanto, toda atribución intencional tendría inevitablemente un carácter de particularidad que hace que no pueda postularse una ley general que capture los fenómenos cuya especificación implica hacer tales atribuciones. Dado que el estudio de la Psicología implica hacer atribuciones intencionales[10], se sigue que no puede haber leyes generales y estrictas que contengan términos del vocabulario de la Psicología. Además, el que los eventos psicológicos no constituyan un sistema cerrado también muestra que, por definición, los predicados mentales no pueden aparecer en leyes estrictas, puesto que, en la caracterización de Davidson tales leyes se dan al interior de sistemas deterministas cerrados. Los eventos psicológicos no pueden constituir un sistema cerrado, ya que hay fenómenos no-psicológicos que influyen en lo psicológico.

Un último punto expositivo acerca de la doctrina del MA es acerca de la naturaleza de la identidad entre los eventos físicos y mentales. Existen dos maneras posibles de postular la identidad entre eventos físicos y mentales: Identidades entre tokens o entre tipos. Davidson afirma acerca de la doctrina del MA que «*Eso significa que los acontecimientos psicológicos, tomados de uno en uno, se pueden describir en términos físicos, es decir, son acontecimientos físicos*»[11]. Esto parece indicar que el MA es una teoría de la token-identidad entre eventos mentales y eventos físicos. ¿Por qué debe postularse esta token-identidad? El MA establece que todos los eventos son eventos físicos que pueden ser descritos o bien en el nivel de las descripciones neuronales, o bien en el nivel de las descripciones en lenguaje psicológico. Los predicados psicológicos que ocurren en las descripciones psicológicas, dadas las razones aducidas anteriormente, no pueden ocurrir en oraciones legaliformes. De aquí en el MA se deriva que la identidad es de tokens. Una manera de dar sentido a este argumento es decir que las oraciones que expresan tipo-identidades entre predicados de ciencias son enunciados con forma de ley, puesto que los predicados de ciencias son términos de clase natural (ver nota 12) y las leyes establecen correlaciones entre términos de clase natural. De esta manera, sería imposible postular las tipo-identidades entre eventos mentales y físicos, en virtud de que los predicados psicológicos no pueden ocurrir en oraciones legaliformes y tales tipo-identidades serían oraciones legaliformes. Así, se antoja necesario postular la token-identidad entre los eventos mentales y físicos.

Teniendo en cuenta todas las consideraciones anteriores, el MA podría ser caracterizado a través de las siguientes dos tesis. La tesis (1), que podría denominarse como la tesis del *Monismo*, afirma la token-identidad entre eventos psicológicos y eventos físicos. La tesis (2), que podría denominarse como la tesis de la *Anomalía*, afirma que las propiedades psicológicas no son reductibles a propiedades físicas[12]. Esto implica que, como se vio

10 Esto, pues la Psicología en últimas tiene como un punto central de su aplicación el proveer una teoría satisfactoria de la conducta humana y la conducta humana debe explicarse con relación a la intencionalidad de los seres humanos, al menos en la concepción de Davidson.

11 Davidson [2001a]. Pág. 71.

12 La noción de 'propiedades reducidas a otras' puede entenderse claramente en el contexto del proyecto de la unidad de la ciencia, en donde reducir una propiedad a otra significó postular leyes puente [$\forall x$ (Fx $\leftrightarrow$ Mx)] que conectaran los predicados que se refieren a tales propiedades. De esta manera, postular que los predicados que se

anteriormente, los eventos descritos con términos psicológicos (que se refieren a propiedades psicológicas) no pueden subsumirse en un sistema determinista de leyes, pues los predicados psicológicos no pueden aparecer en leyes estrictas. Una manera plausible de interpretar el MA es entender la tesis (1) como una tesis de carácter ontológico y la tesis (2) como una tesis de carácter semántico (en tanto es acerca de *cómo se describen* los eventos y no acerca del *status ontológico* de tales eventos). La tesis (1) afirma una identidad en el plano ontológico. Tal tesis podría entenderse como afirmando que en el mundo sólo existe lo físico, y que lo mental *es* algo también físico[13]. Tal identidad se da entre *eventos*. Por su parte, la tesis (2) afirma una irreductibilidad en el plano semántico. Una interpretación plausible de tal tesis sería la que la considera afirmando que si bien lo psicológico es idéntico a lo físico, las descripciones psicológicas no son idénticas a las físicas, puesto que no puede hacerse una reducción de la una a la otra. Es una irreductibilidad entre *propiedades* que juegan un rol en *descripciones* de eventos. Esto permite plantear una distinción entre los planos semántico y ontológico con relación a lo mental. En el plano ontológico, el MA plantearía que no existe propiamente *lo mental*, es decir, estados mentales, sino que lo que existe efectivamente son estados cerebrales[14]. Se había mostrado que para Davidson los estados mentales eran estados relacionados lógicamente con las actitudes proposicionales (verbos psicológicos que señalan una actitud de un sujeto hacia una proposición). Esto daría cuenta de que las descripciones de los estados mentales tienen contenido, pues las actitudes proposicionales implican la presencia de un contenido proposicional, en tanto son relaciones entre sujetos y proposiciones. El que los estados mentales tengan contenido da cuenta de su *intencionalidad*[15], es decir, *son estados intencionales*. Por tanto, en el plano ontológico, el MA implicaría que los estados mentales intencionales no existen de una manera efectiva, puesto que lo que existe efectivamente son los estados cerebrales (tesis Ψ). Por su parte, en el plano semántico, el MA plantearía que si bien no puede admitirse la existencia real y efectiva de los estados mentales intencionales, es posible hacer descripciones de algunos eventos físicos en términos psicológicos, es decir, que utilicen términos referidos a tales estados mentales intencionales, aunque tales descripciones no puedan aparecer en leyes deterministas que se den en sistemas cerrados. Es decir, es en el plano semántico de las descripciones en donde se postulan los estados mentales intencionales. Así, los estados mentales intencionales (específicamente los términos que se referirían a ellos) juegan un rol puramente descriptivo. Lo intencional acerca de la mente aparece en un plano puramente semántico (tesis Φ).

refieren a las propiedades físicas y mentales aparecen en leyes implicaría que tales predicados son predicados de clase natural, es decir, tendrían como referencia clases naturales. Esto, pues las ciencias buscan establecer las clases naturales de un dominio y encontrar las leyes que gobiernan las correlaciones entre los predicados que se refieren a tales clases naturales. Al respecto ver Fodor [1984].

13 Podría pensarse que dado que la identidad es una relación simétrica, también podría afirmarse que todo lo que existe en el mundo es mental. Pero, esto sería una mala interpretación de la situación, pues el conjunto de las cosas físicas es más amplio que el conjunto de las cosas psicológicas. De esta manera, todo lo psicológico puede ser físico, pero no sería posible que todo lo físico sea psicológico.

14 En un sentido más estricto, tal vez debería hablarse de estados sinápticos neuronales y procesos bioquímicos cerebrales. En este sentido se entenderá la expresión 'estados cerebrales'.

15 En un sentido estrictamente general y vago, puede decirse que la *intencionalidad* es una propiedad de los estados mentales de ser acerca de, representar o estar por cosas, propiedades o estados de cosas.

Argumento standard en contra de la inteligencia artificial

Como se explicó anteriormente, el proyecto de la IA busca construir sistemas[16] que simulen-sean una mente. Uno de los argumentos conceptuales más importantes en el debate acerca de la IA es el que busca probar el imperioso fracaso al que se ve avocado este proyecto a partir de mostrar el carácter necesariamente intencional de la mente y la imposibilidad de capturar tal característica en los sistemas que desarrolla la IA. Tal argumento podría reconstruirse, más o menos, de la siguiente manera:

(1) La IA quiere simular-crear la mente.
(2) La mente es intencional (los estados mentales son intencionales) y por tanto debe describirse con términos que expresen su intencionalidad[17].

(C)Todo intento por simular-crear la mente debe capturar el que ella sea intencional (utilizar términos intencionales).

(1)La IA sólo trata con lenguaje extensionales, es decir, en donde sólo se utilizan términos extensionales

---

(C) La IA no puede tener éxito al intentar proveer programas que simulen-creen la mente.

Este argumento en contra de la posibilidad de la IA se centra en el carácter intencional de la mente y la imposibilidad de modelar tal característica en los lenguajes que usa la IA para escribir sus algoritmos. La IA no puede ser exitosa porque debería capturar el carácter intencional de la mente y no puede hacerlo. El presente escrito busca, a partir de la presunción de la verdad de la doctrina del MA, proponer una réplica a este argumento en contra de la posibilidad de la IA. El primer paso en esa dirección será realizar un análisis del argumento a la luz de la doctrina del MA, para aclarar su sentido y aclarar en qué aspecto puede constituir una crítica a la IA. El segundo consistirá en atacar la validez del argumento.

El argumento Standard en contra de la IA depende de dos postulados: (1) el carácter necesariamente intencional de la mente y (2) la imposibilidad de la IA de capturar tal característica. Si se analiza (1) a la luz de la tesis Φ, debe concluirse que el argumento Standard en contra de la IA se da en un plano puramente semántico y por tanto el carácter necesariamente intencional de la mente debe entenderse en ese sentido. Así, el carácter intencional de la mente no estaría dado por la existencia efectiva de estados mentales intencionales, sino por la posibilidad y aun necesidad de postular descripciones psicológicas de ciertos eventos; descripciones que no son reducibles a descripciones físicas. De esta manera, la imposibilidad descrita por el postulado (2) no sería una imposibilidad ontológica de crear estados mentales intencionales (dada la tesis Ψ, que afirma que no hay tales estados) en los sistemas de IA, sino una imposibilidad semántica de describir el funcionamiento de tales sistemas utilizando descripciones psicológicas. Estas aclaraciones

16 Aunque intencionalmente se usa el término 'sistemas' en un sentido general, actualmente la investigación de la IA se centra en sistemas de naturaleza algorítmica. Palabras más, palabras menos, la IA busca crear algoritmos que representen el desarrollo de los procesos mentales.

17 Es decir, utilizando términos intencionales, tales como verbos psicológicos. Es decir, términos que se refieran a estados intencionales de la mente, como las creencias y los deseos.

son importantes, en tanto la premisa dos del argumento Standard podría interpretarse afirmando que la mente debe describirse intencionalmente *porque* ella es intencional, es decir, porque hay estados mentales y ellos son intencionales. La tesis Ψ previene acerca de hacer tales inferencias, puesto que según ella no es posible postular la existencia de estados mentales intencionales. Esto y otras consideraciones relacionadas se tendrán en cuenta con más detalle en la siguiente sección del escrito.

Argumento standard analizado semánticamente

Como se mostró anteriormente, la aceptación del MA implica que toda apelación al carácter intencional de la mente debe entenderse en un plano puramente semántico, específicamente desde un punto de vista descriptivo. Además, la aceptación del MA permite también que el tratamiento semántico de lo intencional con respecto a la mente se haga con absoluta independencia de consideraciones de tipo ontológico acerca de los estados mentales intencionales, pues uno de los principios del MA es que desde la perspectiva ontológica no puede postularse la existencia de tales estados. Así, el argumento Standard en contra de la IA, dado que se fundamenta en el carácter intencional de la mente, debe entenderse desde un punto de vista estrictamente semántico.

Si el argumento Standard necesita apelar a un realismo acerca de los estados mentales, tal argumento podría rechazarse de entrada, puesto que se está presuponiendo la verdad de la doctrina del MA y en tal doctrina la ontología se reduce a lo físico. De esta manera, apelando al principio de caridad, debería hacerse una interpretación puramente semántica del argumento Standard.

Recapitulando, se había dicho que el argumento Standard establece el carácter necesariamente intencional de lo mental y la imposibilidad de la IA de capturar ese componente. Además, se había mostrado la necesidad de dar una interpretación semántica del argumento Standard. Permaneciendo en este espíritu argumentativo, tal argumento debería reformularse de la siguiente manera:

(1) Hay ciertos eventos que deben describirse mediante descripciones psicológicas. Tales eventos son los eventos que descritos psicológicamente podrían interpretarse como el funcionamiento de la mente.

(2) Dado que la IA quisiera simular-crear el funcionamiento de la mente, entonces en los sistemas de IA deberían suceder tales eventos que se describen con descripciones psicológicas y además tales eventos deberían describirse psicológicamente.

(3) Los sistemas de IA no pueden describirse de una manera psicológica.

(C) La IA no puede ser exitosa al intentar simular-crear el funcionamiento de una mente[18].

La premisa (1) se enmarca en el contexto del MA. Tales premisas pueden entenderse en el sentido de que sólo existen cosas físicas, si bien algunas de ellas son susceptibles de una adecuada descripción psicológica. La premisa (2) establece lo que debería hacer la IA para poder simular-crear una mente dentro del contexto del MA. La premisa *3* establece que la

18 La reconstrucción de este argumento es una interpretación de las consideraciones que hace acerca de la Inteligencia Artificial el profesor Garret Thompson [Thomson,1993].

IA no puede capturar el carácter necesariamente intencional de lo mental, en un sentido semántico, es decir, que en los sistemas de IA no ocurren eventos de los que pueda darse una descripción psicológica adecuada.

Al parecer, en esta versión del argumento no queda recogido un aspecto importante del argumento Standard, que se había postulado como el carácter necesariamente intencional de lo mental. El carácter necesariamente intencional de lo mental desde un punto de vista semántico podría entenderse estableciendo que existen ciertos eventos que deben describirse mediante descripciones psicológicas. ¿Por qué es necesario tal aspecto para el argumento Standard? Si no se postulara el carácter necesariamente intencional de lo mental, dada la token-identidad entre eventos-procesos mentales y eventos-procesos físicos, todos los eventos-procesos del mundo podrían entenderse desde su carácter puramente físico y ser descritos mediante descripciones puramente físicas y por tanto el éxito de la IA estaría garantizado de entrada, ya que no hay un argumento de principio que muestre que los procesos y estados puramente cerebrales, dados en términos de neuronas, que forman circuitos y que se comportan eléctricamente, no puedan darse de una manera idéntica en procesos y estados de un hardware, dados en términos de cadenas de silicio que forman circuitos y que se comportan también eléctricamente. Así, el argumento Standard podría postular una objeción a la IA sólo si puede postular la necesidad un nivel descriptivo de los eventos diferente del nivel puramente físico. Esto muestra que el argumento Standard, aun en la versión semántica, necesita postular el carácter necesariamente intencional de lo mental. Y debe postularlo como necesario y no sólo como posible, pues si fuera sólo posible, cabría la posibilidad de quedarse sólo con el nivel descriptivo físico.

Resumiendo, el argumento Standard depende de postular un carácter necesariamente intencional de lo mental. Además, el argumento no puede entenderse desde una perspectiva ontológica en la que el carácter necesariamente intencional de lo mental se haga depender de la existencia de estados mentales intencionales. El argumento Standard sólo puede ser exitoso si se entiende desde una perspectiva semántica, en la que se postule la necesidad de un nivel descriptivo psicológico y se muestre que los sistemas de IA no pueden ser descritos en tal nivel.

Réplica a la versión semántica del argumento standard

El último paso en el argumento del presente escrito estará dado por una propuesta de réplica a la versión semántica del argumento Standard. El ataque al argumento Standard estará enfocado en falsear la premisa (3) (Los sistemas de IA no pueden describirse de una manera psicológica). Para falsear tal premisa se intentará mostrar un argumento que va más o menos así: La tesis del carácter necesariamente intencional de lo mental es central para el argumento Standard. Se buscará dar sentido a tal tesis, analizando cuáles 'propiedades' de la mente harían necesario apelar a tal carácter. En este sentido, se intentará mostrar que la necesidad de apelar a un carácter intencional de lo mental obedece a motivaciones puramente explicativas. Luego, se intentará mostrar que el funcionamiento de los sistemas de IA podría cumplir con tales características, es decir, es posible explicar su comportamiento en términos psicológicos y por tanto describir su comportamiento con descripciones psicológicas. Esto mostraría que la premisa *3* es falsa, en tanto mostraría que los sistemas de IA podrían describirse psicológicamente.

¿Por qué es necesario apelar a un carácter intencional de lo mental? Desde una perspectiva de tercera persona, la conducta es el único testimonio que se tiene de la mente. Esto a menudo se entiende significando que la conducta es un criterio para atribuir estados

mentales. Jugando de observador externo, sólo se puede saber si un sistema tiene o no estados mentales atendiendo a su conducta, pues sólo su comportamiento es público. Esto hace posible que la pregunta por la necesidad del carácter intencional de lo mental se formule en términos de la pregunta por la necesidad del carácter intencional de la conducta, más exactamente la pregunta por la necesidad del carácter intencional de la *descripción* de la conducta. Si puede determinarse que la conducta de un sistema es intencional, esto sería un criterio para afirmar que tal sistema tiene estados mentales intencionales, puesto que la noción de conducta intencional es la de un comportamiento causado por intenciones de un sujeto. Esto presupondría algo aceptado en las teorías contemporáneas de la mente acerca de que existen relaciones 'causales' en algún sentido entre la mente y la conducta. No hay que perder de vista que en el contexto de la discusión que se está teniendo en cuenta, estas consideraciones deben entenderse desde un punto de vista semántico. Desde un punto de vista semántico, esto significaría que ciertos eventos públicos de un sistema serían un criterio para determinar que tales eventos deben ser descritos de una manera psicológica.

Ahora bien, si atendemos al esquema (a):

(1) La persona P intencionadamente hizo A.
(2) 'A' es equivalente a 'B'.

(C) La persona P intencionadamente hizo B.

Éste puede entenderse como un esquema en el que las variables pueden tomar el valor de alguna conducta intencional de los seres humanos, aunque tal esquema no permite realizar inferencias válidas. Tomemos como ejemplo el siguiente caso: 'La abuela de Pepito está muy enferma. Postrada en su cama, depende para vivir del suministro de una mezcla de oxigeno y nitrógeno que se le proporciona desde unos tanques ubicado en su habitación. Para controlar la mezcla hay un controlador de válvula ubicado en la cocina, para que la empleada, su única compañía, pueda controlar la mezcla mientras hace los quehaceres diarios. El controlador es un interruptor con tres botones: Un botón azul que controla el tanque de oxigeno, un botón verde que controla el tanque de nitrógeno y un botón rojo que apaga el sistema de alimentación de la mezcla desde la válvula hasta la mascarilla que siempre tiene puesta la abuela. Pepito recibe unas vacaciones inesperadas dado que su universidad fue cerrada por un paro y decide ir de improvisto a visitar a su agonizante abuela. Cuando llega a la casa, la empleada muy contenta lo saluda y sale a la tienda a comprarle su postre favorito, pidiéndole antes el favor de que le apague el horno para que no se le quemen las galletas. Pepito desprevenidamente ingresa en la cocina, ve el interruptor, cree que es el mando del horno y decide oprimir el botón rojo para apagarlo. Pepito apesadumbrado asiste al funeral de su abuela tres días después.'

La conducta de Pepito podría ponerse en el esquema explicativo (a), así:

(1) Pepito intencionadamente oprimió el botón rojo del interruptor de la cocina.
(2) 'Oprimir el botón rojo del interruptor de la cocina' es equivalente a 'matar a la abuela de Pepito.

(C) Pepito intencionadamente mató a su abuela.

Pero la conducta de Pepito podría también ponerse en otro esquema explicativo (b), similar a (a) pero en el que no se haga referencia a sus intenciones. Esto daría el siguiente esquema explicativo:

(1) Pepito oprimió el botón rojo del interruptor de la cocina.
(2) 'Oprimir el botón rojo del interruptor de la cocina' es equivalente a 'matar a la abuela de Pepito.

(C) Pepito mató a su abuela.

El esquema (a) no es un esquema válido de inferencia, en tanto utiliza el término 'intencionadamente' lo que crea un contexto opaco en el que no es válido hacer la sustitución de idénticos que se realiza en tal esquema. Por el contrario, en el esquema (b) todos los elementos que intervienen hacen que el contexto lingüístico sea transparente y por tanto la sustitución de idénticos es perfectamente válida. Esto implica que en el caso anterior de la conducta de Pepito, no es válido afirmar que él mató (intencionadamente) a su abuela, según el esquema (a), pero sí es válido afirmar que él mató a su abuela, según el esquema (b). Intuitivamente se consideraría que en el caso de Pepito, no sería adecuado decir que él mató a su abuela. Por tanto, lo intuitivo hablaría a favor de un esquema de explicación de la conducta de Pepito en el que no sea posible concluir que mató a su abuela. Es decir, sería necesario apelar a un esquema de explicación en el que se dé un contexto opaco y en el que por tanto no sea posible hacer una sustitución de idénticos. De esta manera, para referirse a la conducta es adecuado utilizar descripciones psicológicas (v.g. que apelen a intenciones de los sujetos).

Esto es muestra de que la apelación a una descripción psicológica de la conducta obedece a motivaciones de tipo heurístico. La apelación a las intenciones de Pepito para explicar su conducta se da en virtud de que es más fácil proceder de esa manera para explicarla. Una explicación de la conducta de Pepito que sólo atienda al aspecto público del comportamiento (como la que se hace en el esquema (b)) no resultaría adecuada, en tanto no captura la fuerte intuición de sentido común que se tiene al respecto. Estas consideraciones nos permitirían concluir que, desde una perspectiva de tercera persona, la motivación para apelar a un carácter intencional de lo mental es puramente heurística. Esto, pues desde la perspectiva de tercera persona sólo se tiene acceso a la conducta en relación a los demás y con relación a la conducta y lo intencional se necesita como un elemento que hace más satisfactorias las explicaciones de los comportamientos. Es decir, se utilizan descripciones psicológicas de la conducta para explicarla de una manera más adecuada.

Pero, alguien podría objetar que las consideraciones anteriores no podrían capturar el por qué debe apelarse a un carácter necesariamente intencional de lo mental desde una perspectiva de primera persona. Primero, porque desde una perspectiva de primera persona el inventario de lo mental parece ser más amplio que desde la perspectiva de tercera persona. Desde la tercera persona sólo puede hablarse de lo mental en tanto podemos relacionarlo con la conducta pública que podemos observar. Pero desde una perspectiva de primera persona tal vez sería posible apelar a otro tipo de características de los estados mentales, que no estén relacionados directamente con la conducta. Es decir, desde una perspectiva semántica, sería lógicamente posible apelar a ciertos eventos que no se relacionan con la conducta y que deben describirse de una manera psicológica. Segundo,

porque la experiencia en primera persona acerca de la propia mente podría dar razones para pensar que una descripción psicológica del funcionamiento de la propia mente obedece no sólo a razones heurísticas, sino que es una descripción que recoge lo que realmente pasa en la propia mente, es decir, que es una descripción realista del funcionamiento de la mente. Esto podría llevar a postular la realidad de los estados mentales.

En cuanto al primer punto, podría concederse que desde la perspectiva de primera persona exista un inventario ampliado de los estados y procesos mentales. Esto, pues lo que se está buscando es la razón que garantizaría que, desde la perspectiva de primera persona, lo mental deba concebirse de una manera necesariamente intencional, en el sentido semántico de que deba describirse necesariamente de una manera psicológica. Y aun más lejos, el punto estaría en analizar si la experiencia del funcionamiento de la propia mente ofrece razones para pensar que posiblemente los estados mentales puedan tener un status ontológico y sea tal status ontológico el que implique que lo mental deba describirse de una manera psicológica. Esto lleva directamente a un análisis del segundo punto. ¿Da la experiencia de primera persona razones para pensar que debe adoptarse un realismo acerca de los estados mentales? Apelando a la propia experiencia de la mente, uno se vería tentado a postular que en su mente hay creencias, hay deseos y todo otro tipo de estados mentales intencionales que darían cuenta del carácter intencional de la mente. Pero, siguiendo una intuición muy básica al respecto, parece posible apelar que este fenómeno se debe más bien a una especie de *falacia mental*. Para un sujeto es heurísticamente adecuado apelar a descripciones psicológicas para explicar la conducta de los demás sujetos, desde una perspectiva de tercera persona. Este mismo sujeto utiliza esa misma clase de descripciones psicológicas para explicarse su propia conducta, como procedimiento heurístico, pero comete la *falacia mental* de pensar que en su caso la apelación a descripciones psicológicas no se debe a un principio heurístico sino a una posible realidad de los estados mentales a los que en tales descripciones se hace referencia. Esto daría cuenta de que aun desde la perspectiva de primera persona, la necesidad de caracterizar la mente como algo intencional obedece a motivaciones puramente heurísticas. Desde la perspectiva semántica del escrito acerca de lo mental, esto significaría que el mantener la necesidad de dar descripciones psicológicas de algunos eventos se justifica en tanto tales descripciones son un elemento heurístico imprescindible.

La postulación de la *falacia mental* es una propuesta de cómo entender la experiencia de la propia mente sin que esto implique postular un realismo acerca de los estados mentales. A favor de la tesis de la falacia mental podría decirse que es *preferible* sobre la postulación del realismo de los estados mentales a partir de la experiencia propia, puesto que aceptarla implica postular un número menor de entidades en el mundo, es decir, es ontológicamente más adecuada. También podría decirse que, dado el contexto del presente escrito, en el que se da por sentada la doctrina del MA, la tesis de la falacia mental es adecuada en tanto rechaza la realidad ontológica de los estados mentales (tesis Ψ). Además, podría intentar articularse un argumento a favor de su verdad: según experimentos[19] realizados por psicólogos, filósofos del desarrollo y científicos cognitivos, en los bebés de menos de un año de edad pueden encontrarse unos primeros desarrollos de la llamada *inteligencia social*[20]. Además de haber aprendido acerca de la rigidez de los objetos, de las interacciones causales, y de la distinción entre objetos animados e inanimados, tales bebés son capaces de asumir que el movimiento de los objetos animados se da a causa de estados internos

19 Experimentos reseñados en Leda Cosmides y John Tooby [2003].

20 En líneas generales, la inteligencia social es la que le permite a un sujeto interpretar la conducta de los miembros de su nicho social.

invisibles. Al parecer, los niños se sirven de analizar la dirección de los ojos y el movimiento para inferir estados internos en la gente, tales como creencias y deseos. Al parecer, tales experimentos podrían interpretarse en el siguiente sentido: Los bebés de menos de un año están desarrollando su inteligencia social. Interpretan el comportamiento de los objetos animados apelando a ciertos estados internos invisibles. Pero, los niños están experimentándose a sí mismos como objetos animados, en el sentido de que causan su propio movimiento. Por lo tanto, parece plausible pensar que interpreten su propio comportamiento apelando a que en ellos mismos se da la presencia de ciertos estados internos invisibles. La apelación a estados internos se aprendió como un procedimiento heurístico para explicar la conducta externa y la propia. Así, es posible apelar que la utilización de tal procedimiento heurístico se preservó a lo largo del desarrollo del niño hasta convertirse en adulto. Al ser mayor y aplicar el procedimiento heurístico a su propia conducta cometió la falacia mental y se auto-engañó creyendo con que en su caso su conducta se explicaba apelando a al existencia efectiva de estados mentales[21].

Las consideraciones anteriores permiten concluir que es plausible pensar que la necesidad de apelar al carácter intencional de la mente se justifica en tanto es un procedimiento heurístico adecuado. La necesidad de mantener descripciones psicológicas de algunos eventos está justificada en tanto es un buen modelo explicativo de tales eventos.

Para completar el argumento es necesario mostrar que es posible describir el comportamiento de los sistemas de IA utilizando descripciones psicológicas. Es decir, es necesario mostrar que el procedimiento heurístico de apelación a lo intencional es aplicable a los sistemas de IA.

En líneas generales, apelar a la necesidad de describir un evento mediante descripciones psicológicas se fundamenta en razones puramente heurísticas. Esto quiere decir que tal descripción se hace en tanto sirve para explicar de una manera más adecuada el comportamiento de un sistema. La utilización o no-utilización de un procedimiento heurístico es una decisión pragmática que toma el explicador. Esto muestra que no hay una imposibilidad lógica de aplicar el procedimiento heurístico en mención para describir el comportamiento de los sistemas de IA. Si un explicador lo juzgara conveniente podría describir el funcionamiento de uno de tales sistemas en términos psicológicos, una vez que la metafísica acerca de lo mental se ha abandonado. No existen estados mentales. Lo mental intencional se da en un nivel puramente descriptivo y con fines netamente explicativos. No hay una imposibilidad de principio para que los sistemas de IA se describan de una manera psicológica. Esto es suficiente para hacer falsa la premisa *3* de la versión semántica del argumento Standard, en tanto en tal premisa se considera que la imposibilidad de que los sistemas de IA se describan en términos psicológicos está dada de suyo, por todo lo que implican los términos 'psicológico', 'mental' e 'intencional'. Pero una vez que el lenguaje acerca de lo mental se ha limpiado de metafísica y se ha aclarado un poco en qué sentido deberían entenderse esos términos, tal imposibilidad de suyo queda eliminada.

Los sistemas de Inteligencia Artificial son sistemas algorítmicos instanciados en mecanismos físicos que se comportan de acuerdo a las instrucciones consignadas en los algoritmos. Los avances que se vienen dando en las diferentes áreas de la Inteligencia

21 Un problema para este argumento sería el caso de un niño autista. Dado que no desarrollaría la inteligencia social, él no habría aprendido el procedimiento heurístico para interpretar la conducta de los integrantes de su nicho social apelando a postular estados internos. Así que surgirían las preguntas ¿explica él su propia conducta apelando a estados internos? y si lo hace ¿cómo lo aprendió? Responder a esto necesitaría un análisis más adecuado del autismo en su relación con el desarrollo cognitivo de un ser humano normal.

Artificial hacen prever la posibilidad de que en un futuro cercano se logren modelar aspectos complejos de la inteligencia humana, tales como la plasticidad, la interacción social o el desarrollo de organizaciones. Con la instanciación de tales modelos algorítmicos en mecanismos físicos tales como los robots, se habría propuesto un sistema de Inteligencia Artificial digno de ser descrito en términos psicológicos, una vez se ha aclarado que esto no significa que se esté afirmando que tal sistema tenga estados mentales, puesto no hay tal cosa como los estados mentales.

La descripción psicológica y el uso común de lenguaje

Aun aceptando que la apelación a descripciones psicológicas del comportamiento de ciertos sistemas obedece a motivaciones heurísticas acerca de tales sistemas y que esto deja abierta la posibilidad de que el comportamiento de sistemas de IA sea descrito en esa vía, podría pensarse en una objeción de principio a la posibilidad de tal descripción. El uso común del lenguaje parece constreñir la posibilidad de describir psicológicamente sistemas mecánicos como los de la IA. En el uso común del lenguaje no parece adecuado decir cosas como que "El robot *creyó* que…" o "El computador pensó *que*…". Esto hablaría en contra de la posibilidad de describir psicológicamente eventos que ocurren en los sistemas de IA.

Tal objeción desde el lenguaje común parece conservar la infección metafísica acerca de lo mental que se ha querido evitar. La manera como usamos las palabras y el significado que asociamos con ellas depende de una manera importante de las relaciones históricas que se han establecido entre tales palabras y las cosas a las que se refieren. La concepción metafísica y aun mística que ha tenido la humanidad acerca de lo mental ha creado paradigmas muy fuertes acerca de qué es tener un mente. La IA trata con la mente y por tanto se enfrenta con los obstáculos que le impone el significado que se le ha dado tradicionalmente a los términos relacionados con ella. Nuestro desconocimiento acerca del funcionamiento del cerebro ha permitido que 'lo mental' se haya envuelto en un velo místico. Lo místico no es bueno ni malo en tanto expresa aspectos muy interesantes de la naturaleza humana. Pero, lo místico sí es inadecuado cuando impone obstáculos infundados al desarrollo de la ciencia, en este caso la ciencia de la mente, de la que hace parte la IA. Nuestras nociones culturales acerca de lo mental como algo místico e inefable son las que fundamentan las críticas acerca de que no es adecuado decir que un robot o un computador piensan. Así, la IA desde una perspectiva filosófica está comprometida con atender a un análisis juicioso del lenguaje relacionado con la mente, para limpiarlo de aspectos metafísicos sinsentido y para prevenir a la IA de confusiones que puedan surgir a partir del significado que se ha asociado a las palabras que utiliza, puesto que tales asociaciones no han sido el resultado de un análisis juicioso. Eso tal vez está bien para un uso diario del lenguaje, pero no en un uso científico de él. La Filosofía y la Ciencia de la mente deben estar más cerca de la ciencia que de la religión.

BIBLIOGRAFÍA:

DAVIDSON, Donald [2001a]. '*Psychology as Philosophy*'. En: *Essays on Actions and Events*, Oxford Clarendon Press. Segunda edición.

[2001b]. 'M*ental Events*'. En: *Essays on Actions and Events*, Oxford: Clarendon Press. Segunda edición.

FODOR, Jerry [1984]. '*Dos clases de reduccionismo*'. Introducción de *El lenguaje del pensamiento*. Alianza-Madrid.

MALPAS, Jeff [2003]. '*Anomalous Monism*'. En: Stanford Encyclopedia of Philosophy. www.plato.stanford.edu

SEARLE, John [1980]. '*Mentes, Cerebros y Programas*'. En: *Filosofía de la Inteligencia Artificial*. Comp. Margaret Boden. Versión española de Fondo de Cultura Económica, México D.F. 1994.

THOMSON, Garret [1993]. '*Una guía simple para la Filosofía de la Mente*'. En: Ideas y Valores. Nº. 90-91, Abril de 1993. Bogotá.

TOOBY, John y COSMIDES, Leda [2003]. *'Evolutionary Psychology: A primer*'. Publicado por el Centro de Psicológia Evolucionista de la Universidad de California. http://www.psych.ucsb.edu/research/cep/primer.html.