UN PROYECTO DE INTELIGENCIA ARTIFICIAL Y SU REFUTACIÓN POR J. SEARLE

A PROPOSAL ON ARTIFICIAL INTELLIGENCE AND ITS REBUTTAL BY J. SEARLE[1]

Andrés del Corral Salazar[2]
Colombia

Palabras clave: problema mente-cuerpo, test de Turing, inteligencia artificial, estados mentales, intencionalidad, argumento de la habitación china.

Keywords: body-mind problem, Turing test, artificial intelligence, mental states, intentionality, Chinese room argument.

RESUMEN

¿Cuál es la relación entre aquello que llamamos mente y nuestro cuerpo?, ¿mente-cuerpo o mente y cuerpo?; ¿una máquina puede pensar?, ¿es el pensamiento un programa de computador? Estos son los dos conjuntos de preguntas que atraviesan el presente texto. Diferentes respuestas se han dado a estos interrogantes y la presentación de algunas nos remontará hasta la época de la filosofía moderna para avanzar desde allí a pasos agigantados hasta el presente. El texto se divide en tres partes o hitos de la filosofía de la mente. En primer lugar expongo la raíz del problema mente-cuerpo en la filosofía cartesiana. De allí paso a presentar el famoso test del padre de la computación y la informática, Alan M Turing, que ha dado lugar a la posibilidad de pensar por primera vez seriamente, experimentalmente, el tema de la inteligencia artificial. Concluyo con la crítica de John Searle a este proyecto.

1.Traducción de Carlos Arturo Muñoz T. Centro de Traducción del Instituto de Idiomas UAM.

2. Profesional en Filosofía y letras de la Universidad de Caldas. Candidato a Magíster en filosofía por la misma universidad. Docente Departamento de Ciencias Humanas de la Universidad Autónoma de Manizales. Integrante del grupo de Investigación Tántalo de filosofía de la Universidad de Caldas. Correo electrónico: andresdelcorral@yahoo.es

Recibido: febrero 15 de 2010
Aprobado: abril 20 de 2010

ABSTRACT

The following are two sets of questions examined along these lines. What is the relationship between the so-called mind and our body? What relationship between mind and body is conceived? Can a machine think? Is the thought a computer program? Different answers have been given to these questions. Some answers will take us back to the era of modern philosophy and lead us directly to the present. This text consists of three parts or milestones in the philosophy of mind. Firstly, I present the root of the mind-body problem in the Cartesian philosophy. Next, I show the famous test given by the father of computing and computer science, Alan M. Turing. This test has led to the possibility of thinking seriously and conducting experiments about the issue of the Artificial Intelligence. Finally, my conclusion is the criticism of John Searle to this proposal.

I- Descartes, el descubrimiento del 'ego' y el dualismo substancialista: el periodo de transición dado en Europa entre la edad media y la moderna, conocido como renacimiento, intentó devolver la confianza en el conocimiento que proporciona la razón o el discurso argumentado. De esta manera, se buscaba superar las contradicciones e inconsistencias que tanto defendían aquellos que pregonaban con la fe. Descartes, por supuesto, fue uno de los abanderados de este proyecto. Sin embargo, podemos encontrar en el mismo Descartes algunas tesis centrales del escolasticismo medieval, aunque fundamentadas de diferente manera. Así, por ejemplo, la división que hace del sujeto en cuerpo y alma, o substancia material y substancia inmaterial, es diferente a la dada por los filósofos medievales. Esa es una razón para considerar a Descartes como padre de la filosofía moderna. En las *Meditaciones metafísicas*, su principal obra filosófica, Descartes propone un método mediante el cual el concepto de realidad se verá revolucionado en tanto que va a servir como fundamento de toda ciencia que opere con leyes matemáticas[3] y permita la cuantificación del mundo, contrario a la idea metafísica-aristotélica de cualificación del mundo que se tenía a disposición en el momento. Así pues, el estudio del mundo debe hacerse por medio de la ciencia, la ciencia debe hacerse sobre la realidad, y la realidad no es otra cosa que las cualidades primarias de las cosas, las cuales permanecen independientemente de la percepción del sujeto y son de carácter extenso y medible. Las cualidades secundarias de las cosas dependen de la percepción del sujeto y no son

3. No olvidemos la anterior consigna de Galileo: " el gran libro del mundo está escrito en lenguaje matemático

medibles, y por tanto, según Descartes, son sólo apariencias y no pueden ser objeto de la ciencia.

Ahora bien, ¿por qué se llega a esta conclusión, en qué descansa dicho razonamiento? Pues bien, Descartes construye su teoría a partir del descubrimiento o conocimiento del 'ego', es decir, del 'yo'. En su búsqueda de bases firmes que sostengan el andamiaje del conocimiento, Descartes suspende sus juicios y emplea la duda metódica–progresiva y, por último, radical. Siendo la claridad y distinción el criterio de verdad empleado por él, cualquier cosa que produzca la más mínima sospecha o duda, será desechada como fuente de conocimiento y todas las cosas que deriven de ella se tendrán por falsas. En consecuencia, Descartes niega la validez de los sentidos y, por medio de un argumento hiperbólico[4], la de la razón. Sin nada en lo que aferrarse, el filósofo se encuentra en medio de una situación donde nada es claro y distinto, en un escenario irreal lleno de espejismos, máscaras y quimeras, donde el método de la duda no conduce sino a laberintos con una única salida: la misma entrada. Así, con todo, ni siquiera el argumento hiperbólico logrará persuadirlo de que no está empleando la duda metódica y, como si después de divagar en el laberinto saliera por donde entró, se encuentra con la primera verdad necesaria de la cual no puede dudar: el pensamiento. Como no puede dudar de que está dudando, y la duda requiere del pensamiento, no puede pensar que no está pensando; además, como el pensamiento es una cualidad que no puede ser por sí misma sino que necesita de algo, una cosa, en donde residir, se llega a la conclusión de que hay 'algo' que existe, una res o substancia, en la que reside el pensamiento. Se dice entonces: "cogito ergo sum", o, "pienso luego (por tanto) existo". En este sentido, la autoconciencia es equivalente al ego, al yo cartesiano, y es desde allí donde debe comenzar la construcción del mundo.

Pero todavía queda una cuestión sin resolver. ¿Qué es esa substancia?, ¿qué es eso que descubrimos que existía? La respuesta de Descartes es bastante sencilla: una cosa que piensa. La propiedad esencial del alma es pues el pensamiento. Entonces, si existiera el mundo externo, las ideas claras y distintas serían las relacionadas con la razón y no con la experiencia, como las derivadas de la matemática y la lógica. La propiedad esencial del cuerpo o de la materia es la extensión. Para terminar, expongamos el razonamiento, con la ayuda de la ley de Leibniz, mediante el cual se afirma la distinción real entre alma y cuerpo, o substancia material e inmaterial:

4. Me refiero aquí al argumento del genio maligno, el cual va a desechar posteriormente con la primera prueba de la existencia de dios. No es necesario exponerlo para los propósitos del texto, al igual que las pruebas mediante las cuales se afirma el engaño de los sentidos.

1. Dos cosas son idénticas si tienen todas sus propiedades en común. (principio de identidad de los indiscernibles de Leibniz).
2. El alma tiene la propiedad de ser indubitable.
3. El cuerpo no tiene dicha propiedad.
4. Por tanto, el alma y el cuerpo son diferentes.

El problema del razonamiento cartesiano es que no puede explicar el hecho de que cualquier persona, por voluntad, pueda ejercer cierto control sobre su cuerpo y mover por ejemplo una extremidad cuando lo desee. En efecto, si cuerpo y alma son diferentes ¿cómo se explica esa relación necesaria entre mente y cuerpo para producir movimiento?

II- Turing y la prueba del pensamiento: en la mitad del siglo XX, A. M. Turing inventó una prueba para determinar si una máquina podría llegar a tener pensamiento. Esta prueba, tal como la planteó Turing, evita el problema cartesiano del dualismo substancialista en tanto que no intenta explicar cómo se relacionan el alma y el cuerpo o, en términos más cartesianos, la substancia material y la substancia inmaterial. El propósito de Turing era simplemente preguntarse si una máquina podría llegar a pensar. El 'test de Turing', como se conoce dicha prueba, es una afirmación respecto a la posibilidad de pensamiento de una máquina. Si la máquina puede participar correctamente en un juego diseñado de tal manera que sus reglas supongan necesariamente la facultad intelectiva de los jugadores, entonces la máquina puede pensar.

En el 'juego de imitación', test de Turing, intervienen tres personas: un hombre, una mujer y un preguntador. El preguntador se sitúa en una habitación aparte prescindiendo de toda inmediación física y, para él, el juego consiste en determinar, por medio de cualquier tipo de pregunta, quién de los otros dos es el hombre y quién la mujer. Los conoce por medio de la referencia X e Y, y al final del juego determina si <X es hombre e Y es mujer> o si <X es mujer e Y es hombre>. El objetivo de los preguntados es lograr que el preguntador efectúe una identificación errónea. Los preguntados pueden y deben engañar al preguntador.

Ahora planteemos la pregunta: ¿qué sucede cuando una máquina sustituye a uno de los preguntados en el juego?, ¿podrá el preguntador identificar el sexo de la otra persona o identificar la máquina? Estas preguntas sustituyen a la original: ¿pueden pensar las máquinas?

Imagen tomada de http://es.wikipedia.org/wiki/Archivo:Prueba_de_Turing.svg

Pero todavía no tenemos claro qué máquina participará en el juego, por tanto hay que definirla. Las máquinas participantes serán computadoras digitales de estado discreto del tipo 'máquina de Turing', es decir, "máquinas ideadas para realizar cualquier tipo de operación propia de un ser humano" que se puedan programar para cada imitación y tengan una capacidad de almacenamiento adecuado, lo cuál permitirá predecir lo que ésta hará sin importar la cantidad de estados que pueda llegar a tener. El problema estriba, en último término según Turing, en programar adecuadamente la máquina. No obstante, es él mismo quien desarrollara un ordenador que procesa exclusivamente señales electrónicas binarias. Dar una instrucción a un procesador supone en realidad enviar series de unos y ceros espaciadas en el tiempo de una forma determinada. La programación, entonces, ya no es un problema.[5] En la década de 1950, tiempo en que Turing dio luz a *Maquinaria computadora e inteligencia*, no existían, desde luego, tal tipo de máquinas con capacidades tan elevadas. Pero, dice Turing: "personalmente creo que, dentro de unos cincuenta años, se podrá perfectamente programar computadoras con una capacidad de almacenamiento aproximada de 10^9 para hacerlas jugar tan bien al juego de imitación que un preguntador no dispondrá de más del 70 por ciento de las posibilidades para efectuar una identificación correcta a los cinco minutos de plantear las preguntas"[6]. A continuación añade "no obstante, creo que, a finales del siglo, el sentido de las palabras y la opinión profesional habrán cambiado tanto que podrá hablarse de máquinas pensantes sin levantar controversia".

5. Los computadores digitales actuales operan aún con el sistema binario, lo que se conoce como 'máquina de Turing'.

6. Turing, A. M. "Maquinaria Computadora e Inteligencia", *Mind*. 1950.

En conclusión, siguiendo a Turing, se supone que si una máquina así definida puede engañar al preguntador en el juego de imitación, entonces es capaz de pensar o, en otras palabras, de pasar un programa tal como se supone sucede cuando un humano piensa.

III– Searle, el naturalismo biológico y la habitación china: retomemos entonces la pregunta con la que se concluyó la primera sección: ¿cómo podemos dar cuenta de las relaciones entre dos géneros de cosas, mente y cuerpo, en apariencia totalmente diferentes?, ¿cómo puede algo mental tener una influencia física?, ¿cómo un mundo físico o material contiene significados? En el siglo XX surgieron varias concepciones anticartesianas que intentaban diluir el problema. El funcionalismo, el materialismo eliminativo o eliminativismo, la teoría de la identidad, el conductismo –teoría que supone Turing y en la que se apoya su test- son concepciones materialistas que tienden a negar o degradar la existencia de los estados mentales. Esto es, niegan que, en realidad, tengamos *intrínsecamente* estados subjetivos, conscientes mentales, y que sean tan reales y tan irreductibles como cualquier otra cosa del universo. Ahora bien, Searle se acerca a dicho problema desde otra perspectiva: el naturalismo biológico. En la medida en que conozcamos más cómo funciona nuestro cerebro tendremos una explicación clara del problema. La tesis que defiende Searle es que los fenómenos mentales son *causados por, y realizados en,* el cerebro y quizá en el sistema nervioso central[7]. Así pues, el problema mente-cuerpo no es realmente un problema puesto que desde esta perspectiva lo físico no es totalmente opuesto a lo mental. Tener mente es un producto biológico como la circulación, la digestión o cualquier otro. El estado mental existe no como un objeto sino como un rasgo del cerebro. Es un conjunto de una propiedad del bio-sistema.

Sin embargo, habrá que redefinir el concepto de causación. En el pasado, D. Hume había sostenido que la ley de la causalidad debía ser, por definición, entre dos entes diferentes o acontecimientos discretos, secuencialmente ordenados en el tiempo. Searle, en cambio, propone una definición de sentido común. La causación se da cuando una cosa hace que suceda otra, sin importar si son o no entes diferentes, etc. Para explicar cómo el cerebro causa la mente hay que diferenciar entre dos niveles: un macronivel o nivel de los estados mentales y un micronivel o nivel neurofisiológico. El micronivel, es decir, las actividades que desarrollan las neuronas, causa estados en el macronivel, es decir, causa

7. Estas tesis aparecen en las conferencias que llevan por título "Mentes, Cerebros y Ciencia". Capítulo I.

estados mentales como el pensamiento, el dolor, la alegría, etc. que a su vez son realizados en el cerebro y quizá en el sistema nervioso central.

Las propiedades del macronivel son sólo del macronivel y no del micronivel, y viceversa. Es un error afirmar, por ejemplo, que una neurona está triste o pensando, o que el dolor está en sinapsis. Searle acepta, por último, que los estados mentales pueden ser reducidos epistemológicamente pero no ontológicamente. Es decir, se pueden *explicar* los estados mentales en términos de un micro o macronivel solamente, pero, no se pueden *reducir* los estados mentales al micronivel.

Por otra parte, el test de Turing ha impulsado la idea de la inteligencia artificial. En efecto, si una máquina puede pensar entonces tiene inteligencia. Pero antes de pasar a analizar la cuestión diferenciemos dos vertientes de dicha idea. Primero está la inteligencia artificial en sentido débil, IA débil, que sostiene que los estados y los procesos mentales pueden ser simulados. Como segunda vertiente está la inteligencia artificial en sentido fuerte, IA fuerte, que mantiene no ya que se puede simular, sino que se pueden duplicar los estados mentales. Realmente, la IA débil no tiene ningún problema y es por simulación, de hecho, que existen los programas de computador. Con la IA fuerte, por el contrario, existen una serie de inconvenientes. A continuación analizaremos la raíz del asunto, iremos al test de Turing y argumentaremos, según Searle, que el test de Turing no es prueba que satisfaga realmente su cometido, a saber, que las máquinas puedan pensar.[8]

Tal vez la mayor contribución de Searle a la filosofía es su teoría de la intencionalidad, que se aplica a su enfoque de mente y lenguaje. La intencionalidad es una característica de algunos estados mentales, tales como percibir, desear, creer o imaginar, etc, de tener la capacidad de referirse a algo o ser sobre algo o ser de algo o de dirigirse a. En este aspecto, la mente es fundamentalmente diferente de cualquier máquina, por sofisticada que ésta sea. El conocido ejemplo de la 'habitación china' explica esta visión. En él, Searle, por analogía al test de Turing, describe a una persona no china, y que no entiende el idioma chino, estando en una habitación usando o moviendo un juego de ideogramas chinos siguiendo una serie de instrucciones para responder a preguntas. Sin embargo, el hecho de mover los ideogramas, incluso cuando se acierta, no prueba necesariamente que el manipulador comprenda

8. Aquí sólo se critica la IA fuerte que se basa en el tipo de pruebas como la de Turing.

algo. Para comprender, una persona (o un sistema) debe usar además conceptos intencionales y esto es tan sólo potestad de la mente. En este sentido, el test de Turing falla porque incurre en un supuesto conductista que le sirve de apoyo a su prueba. El conductismo es la teoría según la cual los estados mentales son conductas efectivas o disposiciones a la conducta. Como, según Turing, el computador digital *actúa* como si pensara y comprendiera, entonces, el computador digital piensa y comprende. Pero esto es un error. No basta actuar 'como si' se pensara para efectivamente pensar. Turing desconoce el carácter intencional del pensamiento, es decir, su capacidad de ser representacional y semántico. Podemos resumir la crítica de Searle en el siguiente argumento:

1). Algunos de nuestros estados mentales son semánticos.
2). Un programa de computador es una estructura enteramente sintáctica.
3). De la sintaxis no se produce la semántica
4). Por tanto, un programa de computador no tiene estados mentales representacionales como el pensamiento.

Como vimos Searle no es ni dualista ni materialista en lo referente al problema mente-cuerpo. Propone una disolución al problema mediante el conocimiento del funcionamiento del cerebro: naturalismo biológico. Además rechaza el test de Turing como prueba que realmente demuestre el pensamiento de alguien o de algo y, por consiguiente, todo intento de IA fuerte que se apoye en el conductismo y funcionalismo, que desconozcan el carácter intencional, representacional y semántico de los estados mentales.

Bibliografía

DESCARTES, René. (2005) *Meditaciones metafísicas.* Madrid: Alianza.

SEARLE, John. (1985) *Mentes, cerebro y ciencia.* Madrid: Cátedra.

SEARLE, John. (2006) *La mente: una breve introducción.* Bogotá: Norma.

TURING, Alan M. (1950) "Maquinaria computadora e inteligencia" EN: ROSS ANDERSON, A. (compilador) (1985) *Controversia sobre mentes y máquinas.* Buenos Aires: Orbis.