

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/262593340>

DESCARTES: LAS INTUICIONES MODALES Y LA INTELIGENCIA ARTIFICIAL CLÁSICA

Article in *Alpha* · July 2011

DOI: 10.4067/S0718-22012011000100014

CITATIONS

0

READS

20

1 author:



Rodrigo González

University of Chile

15 PUBLICATIONS 6 CITATIONS

SEE PROFILE

All content following this page was uploaded by [Rodrigo González](#) on 04 October 2016.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

DESCARTES: LAS INTUICIONES MODALES Y LA INTELIGENCIA ARTIFICIAL CLÁSICA¹

Descartes, the modal intuitions and the Classical AI

*Rodrigo González**

Resumen

Descartes niega que una máquina pueda ser inteligente, pues los mecanismos son predecibles, inflexibles y limitados. Los seguidores de la Inteligencia Artificial clásica (o IA *fuerte*) argumentan lo contrario. Pese a esto, Descartes y la IA proponen que la mente podría no estar adscrita a propiedades físicas, posibilidad explorada por el primero a partir de una intuición modal que separa mente y cuerpo. La IA *fuerte* se acerca a esta tesis cuando reduce la mente a una Máquina de Turing cuya realización física resulta irrelevante. Aunque ambos argumentos presuponen teorías e implicancias diferentes, en este trabajo examino cómo el Funcionalismo resulta compatible con una forma de Dualismo, lo que aleja a la IA clásica del Materialismo que originalmente la inspiró.

Palabras clave: Inteligencia Artificial Fuerte, intuiciones modales, dualismo, funcionalismo, materialismo.

Abstract

Descartes denies that any machine can be intelligent, since machines are predictable, inflexible and limited. Advocates of AI argue the contrary. Even so, both Descartes and classical AI envisage *the possibility* that thought and intelligence need not be instantiated by physical properties. Such a possibility is entertained by Descartes through a modal intuition according to which mind can exist without body, an idea which seems to have been endorsed by classical AI when it reduces mind to a Turing Machine whose physical realization is irrelevant. Even though both arguments presuppose different theories and consequences, Functionalism turns out to be compatible with a form of Dualism, which discards the Materialism that originally inspired classical Artificial Intelligence.

Key words: Strong AI, modal intuitions, Dualism, Functionalism, Materialism.

INTRODUCCIÓN

En el siglo XIX, Charles Babbage (Swade, 2000:28-29) se propuso solucionar el problema de las tablas de cálculo mediante la creación de una máquina capaz de exhibir la inteligencia de un matemático sin la inevitable falibilidad humana. Un antecedente filosófico a considerar por parte de este incipiente proyecto de Inteligencia Artificial es

¹ Artículo formulado en el marco del Proyecto Fondecyt N° 11080020: “Lo imaginario y lo posible: un estudio desde la experiencia concebible”. Agradezco los comentarios recibidos en las 3ras Jornadas de Filosofía Moderna (Chile) de parte de José Tomás Alvarado y Francisco Pereira; de Martina Fürst y Guido Melchior en el 32 ALWS Symposium y de Tim Crane en el curso del Workshop KNEW 2009 (Polonia). Este trabajo también se enriqueció con los comentarios y sugerencias de Philip Robbins.

Descartes, quien niega que cualquier máquina basada en mecanismos pueda pensar, ser racional e inteligente. Puesto de una manera simple, Babbage pensaba que la inteligencia debía estar *en la máquina*, lo que contrasta notoriamente con Descartes, quien aseveraba que esto era imposible y conceptualmente contradictorio, debido a la natural oposición entre lo extenso y lo mental como metafísicamente distintos.

La filosofía contemporánea ha intentado refutar sistemáticamente el *dictum* cartesiano. En el siglo XX, Alan Turing definió qué era computar (1936), e ideó un controvertido test para reemplazar la pregunta sobre si una máquina podía en efecto pensar (1950).² Turing, contrariamente a lo defendido por Babbage, propone que cualquier máquina digital capaz de imitar los *inputs* y *outputs* del cerebro es inteligente y con independencia de los materiales. En particular, afirma que “si alguna máquina particular puede ser descrita como un cerebro, sólo debemos programar nuestra máquina digital para imitar a ésta y también será un cerebro” (1951a:112. Trad., mía).

La primacía de la imitación por sobre los materiales, que es fundamental para entender la teoría funcionalista de los estados mentales (Putnam, 1967) y el principio de realizabilidad múltiple, se acerca a Descartes y a la intuición modal según la cual es concebible que existan estados mentales sin corporalidad alguna. Esto le ha valido a la IA clásica³ haber sido acusada por Searle de adscribir al Dualismo, quien la ataca así

[...] Ciertamente, la IA fuerte sólo tiene sentido dado *el supuesto dualista* de que, en lo que concierne a la mente, el cerebro no importa. En la IA fuerte (y en el Funcionalismo también) *lo que importa son los programas, y estos son independientes de su realización en máquinas*; sin duda, en lo que respecta a la IA, el mismo programa podría realizarse por una máquina electrónica, una substancia mental cartesiana, o el espíritu hegeliano del mundo (1980:86. Cursivas y traducción mías).

En lo que sigue, me propongo mostrar que dicha acusación, reiterada por Searle en varias otras ocasiones (1990-1992-2002), no es infundada, aunque se requieren algunas precisiones, especialmente si se compara la IA *fuerte* y las

² Para una visión panorámica del Test de Turing y por qué resulta tan controvertido como herramienta probatoria de que máquinas programadas piensan, véase González (2007).

³ Este trabajo se circunscribe al examen a la IA clásica (o *fuerte* según Searle, 1980), aunque resulta obvia la importancia que tienen los desarrollos del conexionismo (Rumelhart., *et al.* 1986) y el de Clark (2008). Mientras que la primera teoría plantea un modelo de procesamiento informacional paralelo y no simbólico, inspirado en el funcionamiento del cerebro, la tesis de Clark se opone a la posibilidad de que la mente no esté corporizada y, en este sentido, no sólo es contraria a la intuición modal cartesiana de mentes abstractas. Pese al interés que revisten ambas teorías, un análisis exhaustivo de la relación entre las intuiciones modales cartesianas y *toda* la IA claramente excede el objetivo central de este trabajo, que se acota a establecer cuán materialista es la IA clásica o *fuerte*, en vista de su proyecto original decimonónico. En lo que sigue me refiero a ambas indistintamente.

intuiciones modales del Dualismo Cartesiano. En la primera sección, se caracteriza la visión cartesiana que niega que una máquina pueda tener la inteligencia y racionalidad de un humano. En la segunda, se examina de qué manera los defensores de la IA clásica conciben que una máquina programada sí podría tener mente, con independencia de sus materiales. El nexo entre la intuición modal de que puede existir mente sin cuerpo y la irrelevancia de los materiales de una máquina programada se examina en la tercera sección. Finalmente, concluyo con algunas reflexiones sobre la peculiar y simbiótica relación entre Descartes y la IA clásica.

EL VÍNCULO CARTESIANO ENTRE PENSAMIENTO, LENGUAJE Y RACIONALIDAD

Descartes sostiene que el pensamiento y la racionalidad están íntimamente ligados al lenguaje, aunque éste impida producir conocimiento cierto y seguro, ya que las palabras varían en su significación y llevan a confusión, a diferencia de las ideas claras y distintas (1985a:12;1994:79).

No obstante, la argumentación que emplea para establecer que una máquina es incapaz de pensamiento no sólo proviene de cómo se alcanza conocimiento, sino de sus consideraciones acerca del uso de signos lingüísticos, de la incapacidad de los animales para usarlos, y de las estructuras mecánicas internas que explican los movimientos corporales automáticos. Naturalmente, existe una notoria diferencia entre los signos naturales, entre los que figuran las respuestas naturales comunes a una especie frente a un estímulo determinado, y los signos lingüísticos que son convencionales e involucran consciencia. Es decir, mientras que existen *signos naturales* como respuestas automáticas y genéricas en una especie, y que hacen que un animal gima de dolor al ser golpeado, por ejemplo, los *signos convencionales* lingüísticos no son empleados mecánicamente automáticamente.

Cabe destacar que esta diferencia entre el uso de signos por parte de animales y humanos no es de grado, sino de naturaleza, tal como Descartes enfatiza en la siguiente en su carta a Mersenne, en 1629

Con respecto a las palabras que significan de manera natural, acepto como una explicación válida que cualquier cosa que impacte nuestros sentidos nos lleva a emitir algún sonido; por ejemplo, si somos golpeados, nos hará quejarnos; si alguien nos hace algo que cause placer, nos hará reír; y que los sonidos que uno emite, al quejarnos o reír, son similares en todas las lenguas. Pero cuando veo el cielo o la tierra, no resulto forzado a llamarlas de una manera u otra y creo que esto sería el caso incluso si estuviera en un estado de inocencia adánica (Clarke 2003:163. Trad., mía).

De esta forma, el uso del lenguaje por parte de los humanos muestra de manera inequívoca la existencia de pensamiento, capacidad que es acompañada de la flexibilidad de la razón y de la voluntad libre, que elige las palabras para comunicar ideas. Claramente, tal libertad no existe cuando alguien golpea a un animal y emite un sonido de queja de manera automática e involuntaria. En este caso, los famosos

“espíritus animales” mueven a ciertas disposiciones naturales de los órganos del cuerpo que causan la conducta, y ésta no es mediada por ninguna decisión racional.

Para Descartes, los espíritus “son como el sutilísimo viento, o más bien como una vivísima y purísima llama, la cual asciende de continuo muy abundante desde el corazón al cerebro y se corre por los nervios a los músculos y pone en movimiento todos los miembros” (1994:111). Por este motivo, resulta natural la elucidación de los espíritus animales en términos de las leyes de la mecánica, y la natural oposición entre éstas y el pensamiento.

Sin embargo, una máquina puede programarse para usar signos lingüísticos, aunque de manera limitada. La inhabilidad para articular lenguaje de una manera programada o mecanizada en animales, y que pueda ser objeto de predicciones, indica que éstos y máquinas son no pensantes, cuestión que explica por qué no adquieren conocimiento práctico a través del ejercicio de la razón y de la facultad del entendimiento. Descartes argumenta sobre este punto de la siguiente forma

Pero [una máquina] no podría ordenar las palabras de formas diferentes para responder al significado que se dice en su presencia, como incluso el menos inteligente de los humanos puede hacer [...]

Incluso, aunque tales máquinas podrían hacer algunas cosas tal como nosotros les hacemos, o de mejor forma, inevitablemente fallarían en otras, lo cual revelaría que no están actuando de acuerdo con su entendimiento, sino por la pura disposición de sus órganos (1994:113).

Una máquina cartesiana podría, así, reaccionar a estímulos ambientales de manera rígida y mecánica, tal como también sucede con los animales que reaccionan mecánicamente al estímulo a través de procesos que incluyen pasiones y espíritus animales, y que involucran toda clase de impulsos naturales. La conexión entre éstos y los estímulos es obviamente *causal* y, en consecuencia, no varía. Luego, la reacción mecánica constituiría el primer signo que muestra por qué una máquina sería incapaz de pensamiento.

El lenguaje humano *no* puede explicarse mecánicamente y tampoco ostenta un vínculo causal entre estímulos ambientales y respuestas lingüísticas. Incluso, los humanos tienen la capacidad de reordenar las palabras de una oración para significar lo mismo de variadas maneras frente al mismo estímulo. La intención de comunicar es lo que guía a los hablantes y está ligada a la experiencia de hablar. Esto, a su vez, muestra la consciencia de las intenciones y la flexibilidad para elegir palabras.

Estas disquisiciones cartesianas en que compara a humanos con animales y máquinas, y que enfatizan la importancia del lenguaje, van más allá todavía, pues el uso de palabras es considerada un signo inequívoco de racionalidad y, por tanto dicha racionalidad constituiría el segundo signo que indica la existencia de pensamiento.

La racionalidad no sólo es definida por Descartes en función de la habilidad para aprender a usar palabras y signos convencionales sino, además, por la capacidad de actuar inteligentemente en función de conocimiento, tal como se

señalaba más arriba. Por ello, el filósofo concluye que mientras “la razón es un instrumento universal que puede usarse en toda clase de situaciones, los órganos necesitan cierta disposición particular para cada acción específica” (1994:113).

Si, además, se toma en consideración que el número de las respuestas lingüísticas apropiadas es indefinidamente grande, y que el número de estímulos lingüísticos al que pueden responder los humanos igualmente lo es, se sigue que es imposible para máquinas y animales almacenar y reproducir respuestas inteligentes que imiten la flexibilidad y variabilidad de la razón humana en circunstancias diversas, pues ambos poseen una capacidad de programación limitada. Sólo la razón es capaz de determinar qué acción es adecuada dentro de un contexto específico y, por tanto, actuar de manera inteligente. Así, Descartes asevera que aquello que es susceptible de explicación mecánica tiene recursos finitos y limitados, a diferencia de la razón humana que tiene un gran y variado número de recursos los que, a su vez, se ligan a la creatividad y a la inteligencia. Incluso, el pensamiento implicaría la inmortalidad del alma, tal como hace ver en respuesta a una objeción de Newcastle en 1646 sobre los animales: “[...] si pensasen como nosotros lo hacemos, tendrían un alma inmortal como la nuestra”.⁴ (1991a:36. Trad., mía). Y tal como insiste posteriormente en 1649, cuando afirma que toda la inteligencia y astucia de los animales se explica por la constitución de sus órganos, pero que resulta imposible probar la existencia de pensamiento en ellos (1991b:37). Sin embargo, esto no significa que los animales no sientan, ni que carezcan de consciencia, sino que —como certeramente aclara Cottingham (1998:225-6)— sólo se sigue que los animales son autómatas que no piensan como nosotros, que carecen de lenguaje, y que no tienen autoconsciencia.⁵

En síntesis, para el filósofo francés todo ser que sustente su actuar en mecanismos, será incapaz de emplear lenguaje, pensar como nosotros y ser racional. En efecto, ninguna tesis podría contrastar más con el proyecto decimonónico de la IA, implementado por Turing a partir de las nociones de algoritmo y computador digital en el siglo XX.

EL PROYECTO DE LA IA CLÁSICA: LA INTELIGENCIA ALGORÍTMICA

A pesar de que los fundadores de la IA se oponen férreamente a la idea de que una máquina no puede ser pensante e inteligente, es Alan Turing quien más argumentó respecto a por qué resulta infundado no atribuir inteligencia a una

⁴ Este argumento metafísico de Descartes, que vincula pensamiento e inmortalidad, también coincide con algunas consideraciones funcionalistas sobre la inmaterialidad de nuestra vida mental y la posibilidad de que sea eterna, cuestión que será explorada en la tercera sección y en la conclusión.

⁵ Ahora bien, respecto de si sienten como el ser humano, sólo puede sostenerse hipotéticamente que lo hacen, pero esto es una cuestión empírica de verificación imposible en principio, dado que no podemos imaginar cómo se siente ser otro organismo, menos si es inconsciente, como sería una máquina-animal. Véase, por ejemplo, el acucioso análisis de Nagel (1974:221) y sus consecuencias para el fisicismo reductivo.

máquina programada, puesto que sí es posible mecanizar el pensamiento y la inteligencia mediante procedimientos efectivos o algorítmicos.

El carácter multipropósito y flexible de una máquina programada es lo que sirve de base para afirmar que es inteligente, cuestión íntimamente ligada a la Máquina de Turing (MT de aquí en adelante). Esta es la idealización matemática de un dispositivo mecánico que opera en base a quintuplos (Turing, 1936:231-5) y que se liga al sentido intuitivo de computar (1936:250-3). Computar involucra una cantidad finita de estados discretos, o la ejecución de un algoritmo, cuyos pasos, también finitos, son calculables mediante lápiz y papel y sin la participación de consciencia. El término del procedimiento algorítmico se alcanza cuando se encuentra la solución a un problema. Justamente, un “computador humano” —que emplea lápiz y papel— puede definirse en términos de una MT, o del Cálculo Lambda, el cual establece operaciones de abstracción y generalización realizables por dicha máquina hasta llegar a una solución (Copeland, 1993:230).

Cabe destacar que los algoritmos, fundamentales para entender qué es una MT, eran ya conocidos en la antigüedad. Un ejemplo es el procedimiento planteado por Euclides para encontrar el máximo común denominador de dos números al dividir sucesivamente la segunda cifra divisora por el remanente de la división entre la primera y la segunda cifra, hasta que dicho remanente sea 0 (Penrose, 1999:40-45).

Este procedimiento sistemático es efectivo, esto es, describe una serie finita de pasos y se detiene cuando se alcanza un resultado específico. Sin embargo, da la impresión errónea de que los algoritmos sólo se aplican a problemas matemáticos, cuestión que recuerda lo que Babbage pensaba respecto al carácter eminentemente matemático de sus máquinas. Por supuesto, esto es inexacto ya que en la vida diaria uno constantemente aplica algoritmos para resolver problemas (por ejemplo, para encontrar la combinación correcta de una cerradura en un manojó de llaves, uno realiza una serie de acciones finitas y recursivas).⁶ Pero ¿Cuál es la relación entre un algoritmo y una MT?

La MT fue planteada como una manera de formalizar términos como “procedimiento mecánico” y “máquina”, puesto que dicho dispositivo simplemente ejecuta un cálculo definible de manera finita y recursiva, lo que lleva a la realización de un algoritmo. Por lo mismo, es una máquina idealizada y formal, con una cantidad de estados discretos posibles, los cuales son finitos aunque pueden ser muchísimos.

⁶ La situación en la que la llave no está en el manojó plantea un caso similar a aquellos procedimientos algorítmicos que nunca se detienen, pues no existe solución para el problema planteado. Estos procedimientos describen el problema conocido como *halting problem*. Según Penrose (1992:3-4), que uno entienda y se dé cuenta de lo inefectivo de un procedimiento algorítmico, y que no se detendrá, es evidencia suficiente para concluir que la inteligencia humana no es reducible a los programas computacionales de la IA clásica. Por supuesto, este argumento ha recibido un sinnúmero de réplicas. Turing mismo, por ejemplo, anticipó esto, pero lo descarta en función de que tal como hay cosas que no puede hacer una máquina programada y sí un humano, hay cosas que un humano no puede hacer y sí una máquina digital (1950:51; 1951b:105).

Pese a esta aparente restricción, la MT no posee límite con respecto a los cálculos posibles, ya que tiene un conjunto de instrucciones que se ejecutan con independencia del tamaño de los números a calcular. Las entradas o *inputs* tampoco necesitan ser restringidos y se utiliza una capacidad de almacenaje de información externa, la que es usualmente descrita como papel para realizar los cálculos y producir el *output*, que tampoco tiene tamaño limitado.

Asimismo, no se supone que la máquina internalice los datos externos o los cálculos, sino que trata con cálculos u operaciones inmediatas. El tamaño ilimitado del *input* vs la finitud de los estados o pasos para los cálculos corrobora (aunque, obviamente, no implica) que las Máquinas de Turing son idealizaciones matemáticas y no máquinas que uno pueda construir en casa, punto sobre el cual Penrose insiste afirmando que “es la naturaleza ilimitada del input, del espacio de cálculo, y del output lo que nos indica que estamos considerando una idealización matemática en lugar de algo que puede construirse en la práctica” (1999:47. Trad., y cursivas mías).

La idealización de qué es computar a través de la MT se vincula a su representación mediante una cinta infinita dividida en celdas, con una cabeza lecto-escritora, que se mueve hacia derecha e izquierda a través de la cinta, y que “recuerda” alguno de los símbolos leídos o estados discretos (por contraposición a continuos). En un tiempo t , la cabeza que se encuentra en un estado interno ($q_0, q_1, q_2, \dots, q_n$), lee el símbolo de la celda de la cinta ($b_1, b_2, b_3, \dots, b_n$). En función del estado interno, de lo leído, y del micro código o programa de la cabeza, esta mantiene el símbolo o lo borra e imprime otro. Posteriormente, se detiene o se mueve a otra celda, continuando las computaciones hasta que se detenga la cabeza, lo cual marca el final del procedimiento.

En el caso de una MT típica, esta tiene 0 y 1 como símbolos en las celdas y su conducta está totalmente determinada por la tabla de la máquina o el programa. Por ejemplo, la máquina —al estar en un estado interno y leer el símbolo de una celda— tiene un *input* I_n , el cual consiste en el par <estado interno, símbolo leído>. Luego, en virtud de I_n y del programa trata de determinar una tripleta de *output*, esto es, <escribe símbolo, movimiento (o detención) y nuevo estado interno>. Si la máquina no es capaz de determinar la tripleta, se detiene y también lo hace cuando la computación ha llegado a su fin. Sobre la base de O_n , la máquina escribe un símbolo (1 o 0) en la celda, se mueve a la izquierda o derecha (o se detiene ahí) e ingresa a un nuevo estado. Eventualmente, la máquina vuelve al primer paso, si no ha llegado a detenerse luego de computar la solución.

Toda la conducta de una MT está expresada a través de enunciados condicionales, y del programa con la lista finita de reglas ordenadas en pares <entrada, salida>. Las reglas se expresan mediante fórmulas “Si... entonces...”, que sirven para conformar quintuplos de la forma <estado actual, leer símbolo, escribir símbolo, moverse (o detenerse) y nuevo estado>. Una cuestión importante de tener en cuenta es que los estados de la máquina, las colecciones de símbolos y las acciones, como borrar,

escribir y moverse, son además finitas, discretas y distinguibles (la máquina no puede estar en dos estados a la vez).

La MT y el Lambda Calculus son fuente de inspiración de *la tesis Church-Turing*, la cual define una clase de procedimientos *efectivos* en lógica y matemáticas (Copeland, 2002). Uno puede entender intuitivamente que un procedimiento *M* es efectivo si y solo si: a) *M* se concibe mediante una lista finita de instrucciones, las que son expresadas por una lista finita de símbolos; b) *M* se ejecuta sin errores en una serie de pasos finitos; c) *M* puede ejecutarse por parte de un humano que emplee lápiz y papel para los cálculos; d) La ejecución de *M* no requiere de consciencia para realizar los cálculos. A su vez, si un procedimiento es efectivo y su función es computable, lo es por una MT o, de manera equivalente, a través del Lambda Calculus.

El poder de una MT radica en que puede, en principio, simular cualquier sistema algorítmicamente calculable y que sea expresable en términos del Lambda Calculus. Cualquier máquina que implemente un procedimiento efectivo *M* podrá ser imitada por una MT, y esto es supuestamente crucial para la construcción de máquinas programadas capaces de exhibir la misma inteligencia de un humano. En efecto, el proyecto de la IA clásica contempla que máquinas programadas puedan ejecutar tareas y resolver problemas que imiten los *inputs* y *outputs* del cerebro.

Tradicionalmente, se ha interpretado que la independencia de los materiales de la MT cuenta como un importante antecedente del principio de realizabilidad múltiple, uno de los pilares del Funcionalismo. Puesto de manera simple, dicho principio señala que “ser una máquina X” es funcionar como X de modo que los materiales son irrelevantes, pues qué hace y cómo funciona la máquina es lo único que importa.

Por supuesto, esto no significa que la MT y el Lambda Calculus impliquen el principio de la realizabilidad múltiple. En efecto, ambas tesis son compatibles con que exista una sola realización material de la inteligencia (por ejemplo, el cerebro) y que este sea una MT. No obstante, la insistencia de la IA clásica o *fuerte* de que no es necesario replicar las propiedades físicas del cerebro asume sin más que la inteligencia es el programa del cerebro que puede imitarse, en principio, por cualquier MT y con independencia de sus materiales.

Justamente, los defensores de la denominada IA *fuerte* se inspiran en el Funcionalismo para sostener que la mente puede modelarse computacionalmente, y que una máquina que opere como un cerebro *es* un cerebro artificial. Es decir, la mente humana no sería más que un computador digital apropiadamente programado, tesis que resulta complementaria con lo postulado por la ciencia cognitiva respecto a que el cerebro es una máquina sintáctica operando una máquina semántica (Block, 1990:267). Es decir, los seguidores de la IA *fuerte* y de la ciencia cognitiva sostienen

que existen ciertas estructuras simbólicas en el cerebro que se correlacionan y que explican las actividades del pensamiento y de la inteligencia humana.

Si a esto se suma la irrelevancia del material en que se realiza un programa, una idea propia del Funcionalismo de MT, la cercanía con la intuición modal cartesiana que separa mente y cuerpo parece aún más patente, pues —tal como Searle asevera en diversas ocasiones (1980-1990-1992-2002)— existe un resabio dualista en la IA *fuerte*. El siguiente pasaje esclarece cómo la IA clásica o *fuerte* no sólo sería compatible con el dualismo, sino que contaría como una forma de este

La IA fuerte es una extraña mezcla de Conductismo y Dualismo. Es conductista en su aceptación del Test de Turing, pero en un nivel más profundamente filosófico es dualista, porque rechaza la idea de que la consciencia y la intencionalidad son fenómenos biológicos comunes como la digestión. Según Dennett y Hofstadter (1981:15), debemos pensar en la mente como en “un tipo de cosa abstracta cuya identidad es independiente de una encarnación física” (Searle, 2002:57. Trad., mía).

Según Searle, el dualismo, que divide el mundo en un dominio mental y en un dominio físico, se combinó con otras dos importantes ideas que influyeron la vida intelectual del siglo XX (2002:59). Primero, con el método científico y la necesidad de restringir el saber de la ciencia a lo que puede enunciarse en forma matemática; y segundo, lo indubitable de los dogmas religiosos *vs* la provisionalidad de las hipótesis científicas postuladas por la razón. Esta combinación de factores explicaría la aparición de la IA *fuerte* como teoría computacional de la mente en que ésta, en lugar de ser un fenómeno mundano y ordinario, cuenta como hecho eminentemente formal y abstracto.

Ponderaremos, luego, cuán justa es la acusación de Dualismo a la IA *fuerte* a la luz del examen del Funcionalismo de MT y de la intuición modal cartesiana de que es posible separar mente y cuerpo. Como se examinará en detalle, aunque el Funcionalismo no implica una forma de Dualismo, el principio de realizabilidad múltiple sí resulta compatible con una de sus formas. Pese a esta cercanía, lo que distingue crucialmente a Descartes y la IA clásica es que inteligencia y pensamiento no pueden ser mecanismos para el primero, incluso si tienen la naturaleza sintáctica y formal de una MT.

DESCARTES Y LA IA CLÁSICA O FUERTE: COINCIDENCIAS Y DIFERENCIAS CRUCIALES

En el siglo XX, el Funcionalismo nace como una teoría acerca de la naturaleza de los estados mentales, alternativa a la teoría de la identidad que típicamente reduce estados mentales tipo a estados neuronales tipo, es decir, intenta reducir lo mental a lo físico. De acuerdo con este *Fisicismo* reductivo, cualquier propiedad mental (tener dolor) debería identificarse o reducirse a una propiedad física (la activación de las fibras C), y la existencia del primer tipo de propiedades no podría darse sin la existencia del segundo tipo, ya que existiría una relación de

identidad y dependencia entre ambas. Una serie de objeciones modales (Kripke, 1980:144-6) al chauvinismo neural que se sigue del intento de reducción materialista, precisamente basadas en intuiciones cartesianas, pusieron en entredicho el Materialismo reduccionista tipo-tipo, lo que ayudó a catapultar al Funcionalismo y a su apelación a los estados discretos de una MT como capaces de caracterizar funcionalmente los estados mentales. El acerbo cartesiano del argumento de Kripke en contra de la Teoría de la Identidad puede apreciarse en el siguiente pasaje

Descartes y algunos seguidores argumentaron que una persona o mente es distinta de su cuerpo dado que la mente podía existir sin el cuerpo. Pudo haberse argumentado a favor de la misma conclusión a partir de la premisa de que el cuerpo podría haber existido sin la mente. Ahora bien, la respuesta que considero sencillamente inadmisibile es la que acepta gustosamente la premisa Cartesiana en tanto niega la conclusión Cartesiana. Sea "Descartes" un nombre, o designador rígido de su cuerpo, de determinada persona, y sea "B" un designador rígido de su cuerpo. Entonces, si Descartes fuera efectivamente idéntico a B, la supuesta identidad al ser una entre dos designadores rígidos, sería necesaria y Descartes no podría existir sin B y B no podría existir sin Descartes. El caso análogo no es ninguna manera comparable al supuesto caso análogo del primer director general de Correos y el inventor de los lentes bifocales (1980:150-1).

Dado que sí es posible concebir estados mentales sin identificarlos con estados neurales, puesto que puede darse el dolor sin la activación de las Fibras C o viceversa, esto socavaría las bases de la Teoría de la Identidad y motivaría que los funcionalistas propongan un *Fisicismo* no reductivo. Explicado de una manera no técnica, el Funcionalismo afirma que un estado mental es sólo un estado con causas y efectos típicos (Lewis, 1966), y que se debe distinguir entre el rol causal de un estado mental y el ocupante, o la función que desempeña un estado mental en la economía funcional de un organismo y su realización física. Un rol podría tener varios ocupantes y realizadores. Por ejemplo, el dolor de cabeza sería un conjunto de causas y efectos conductuales que podría realizarse en distintos sustratos materiales, siendo las propiedades físicas fundamentales no reductivas de las propiedades mentales.

A su vez, una variante de esta aproximación sostiene que un estado mental es compatible con una serie de predicados físicos diferentes (Putnam, 1967:162), tesis que da lugar al Funcionalismo de MT que, justamente, profundiza la idea anterior, al sostener que los estados mentales pueden reducirse a los estados de una tabla de MT, o a la disyunción de dichos estados. Originalmente, la tesis del Funcionalismo de MT surgió como una aclaración sobre qué objetos pueden compartir propiedades psicológicas y, luego, a cuáles se les pueden aplicar tales predicados con independencia de sus propiedades físicas.

Esta teoría, que versa sobre la naturaleza de los estados mentales, se apoya en el principio de realizabilidad múltiple, tesis según la cual un estado mental tipo, al que se le aplica un predicado psicológico, puede tener distintas realizaciones en distintas propiedades físicas. Pese a esta diferencia, los organismos que tengan estas últimas compartirán el mismo perfil causal (el rol funcional) de tal estado mental.

Al describir computacionalmente los estados mentales, la MT puede potencialmente ser realizada por una gran variedad de materiales. Por lo mismo, no importa que sus estados no sean especificaciones de estados neurales tipo, tal como defiende el *chauvinismo neural* de la teoría de la identidad de tipos. Por el contrario, la caracterización funcional de los estados mentales mediante un *automaton* finito permitiría obviar el material y los realizadores de tales estados. Y ello ocurre porque un *automaton* finito es definido de manera similar a una MT, “excepto que las transiciones entre los “estados” se permiten con varias probabilidades en vez de determinísticamente” (Putnam, 1967:162).

Teniendo presente estos argumentos en contra del *chauvinismo neural*, la mente en su conjunto podría verse como una MT con estados probabilísticos. Si ésta caracteriza funcionalmente ciertos estados, podría existir una gran variedad de implementaciones de dicha máquina que no tuvieran nada que ver con cerebros y sistemas nerviosos, siempre y cuando se preservara la economía funcional y la caracterización de los estados mentales en términos de causas y efectos. En consecuencia, el Funcionalista de MT propone que la organización funcional mediante perfiles causales, y no la realización material de los estados mentales, es crucial para caracterizarlos.

Sin embargo, esta afinidad entre una MT y la caracterización de los estados mentales mediante estados funcionales daría pie a algo bastante contra intuitivo: la posibilidad de entidades conscientes no-físicas, con perfiles causales compatibles al humano o animal. Desde un punto de vista conceptual, es posible plantear que existen entidades conscientes, pero sin realización física. Por ejemplo, resulta concebible la existencia de un estado mental como el dolor en seres ectoplásmicos,⁷ o en ángeles en el sentido de que estos seres podrían tener un perfil causal similar al de los humanos o los animales y una realización no física, problema metafísico que Heil aborda, así

Si los estados de la mente son estados funcionales, y si los estados funcionales se “realizan” en creaturas conscientes a través de estados con un perfil causal apropiado, esto deja abierta la posibilidad de que seres inmateriales puedan estar conscientes, pensar, sentir dolor. Ellos podrían poseer tales estados de la mente si pudiesen estar en estados con los perfiles causales adecuados. Una sustancia ectoplásmica, por ejemplo, o

⁷ Ectoplasma es la sustancia espiritual (o etérea) que le daría presencia a espíritus y fantasmas. Es, por lo mismo, un tipo de realización, aunque parezca controvertido establecer si es *física* o de algún tipo de energía.

un ángel con la organización interna ectoplásmica o angelical podría pensar o sentir dolor (Heil, 2004:148).

Pero ¿cómo ángeles y seres ectoplásmicos podrían tener una implementación no física en consideración de que interactúan con objetos físicos? Tal posibilidad, aunque no parece implicar contradicción desde un punto de vista conceptual —pues la sustancia ectoplásmica supuestamente cuenta como un tipo de energía— resulta chocante desde el punto de vista de la física popular y del sentido común. Por ejemplo, tal como Dennett critica —en su alusión al Dualismo y a *Casper el fantasmita amistoso*— argumentar en base a este tipo de seres lleva a una paradoja ante la cual los niños se asombran: “¿Cómo es posible que Casper traspase murallas y sea capaz de agarrar una toalla que se cae?” (1991:35. Trad., mía). En consecuencia, si bien el Funcionalismo y el principio de realizabilidad múltiple no implican al Dualismo, la compatibilidad entre éstos y la idea de que el material no importa genera un importante problema metafísico, toda vez que el Funcionalismo y otras teorías, en principio fisicistas, son planteadas como alternativas materialistas contra Descartes.

Más aún, decir que los programas son independientes de sus realizaciones físicas es una manera de sostener que el nivel sintáctico, los patrones formales, es *separable* del nivel físico, y que lo mental *no* está adscrito a las propiedades físicas de ningún material. La realizabilidad en múltiples materiales supone que las propiedades de ellos no son relevantes para la implementación de un programa e, incluso, que la realizabilidad podría concretarse a través de propiedades no físicas, tal como sugiere el ejemplo de las sustancias ectoplásmicas y ángeles. Esta posibilidad conceptual sobre bases *a priori* es crucial para apreciar cómo el Funcionalismo se acerca al Dualismo al proponer dos niveles metafísicos de lo real, a saber, el de los estados funcionales-abstractos y el nivel de realización.⁸

⁸ Incluso, uno no requiere de seres tan polémicos para mostrar por qué la separación de dos niveles metafísicos involucra que uno podría existir sin el segundo, como es el caso de programas y realizaciones. Un proyecto de inversión podría contar como un programa con una serie de pasos finitos regulados sintácticamente y con un objetivo central, a saber, obtener la mayor plusvalía posible. Las etapas del proyecto podrían justamente verse como los pasos de dicho programa, necesarios para la consecución del objetivo central y por lo mismo, podría considerarse un procedimiento efectivo mundano. Ahora bien ¿Es necesario que el proyecto de inversión se materialice de alguna manera, que se compren bienes y que mediante ellos se obtenga la mayor plusvalía posible? En estricto rigor, no es necesario que el proyecto mismo se materialice en ningún bien concreto, ya que podría invertirse todo el capital en acciones y en otros instrumentos financieros que no tienen una realización física y que, en este sentido, serían mentalmente dependientes. Incluso, el capital final obtenido mediante la ejecución del proyecto podría figurar como una cifra abstracta, sin que se materializase de ninguna forma hasta que alguien fuera a un banco e hiciera el retiro de los fondos. Por supuesto, el hecho de que un proyecto de inversión pueda contar como un programa, entendido como una serie de pasos finitos recursivos tendientes a la resolución de un problema específico, no significa que deba realizarse en computadores electrónicos específicos. Este procedimiento efectivo y mundano podría sólo contar como una serie de ideas conectadas sintácticamente en las narraciones biográficas de algún especulador.

En consecuencia, existe una importante implicancia al afirmar que el nivel funcional no es reducible al nivel material concreto, y que el primero podría realizarse de múltiples maneras en el segundo o que, incluso, podría no realizarse en un objeto físico. En efecto, afirmar que las propiedades físicas no son relevantes para lograr la implementación de los estados funcionales implícitamente hace “flotar” los primeros sobre la realidad concreta del mundo físico, cuestión algo paradójica si, nuevamente, se tiene en cuenta que el Funcionalismo pretende erigirse como una clase de Físicismo auténticamente materialista que refute a Descartes y a la idea de que lo mental puede escindirse de lo físico.

Con respecto a la posible separación de niveles abstractos y concretos por parte de funcionalistas, conviene recordar que el filósofo francés propone que es posible distinguir clara y distintamente que puede existir una mente sin realización física concreta, y que, entonces, desde el punto de vista de lo concebible, la propiedades mentales no están adscritas a las propiedades físicas, ni lo mental sería dependiente en modo alguno de lo físico. Putnam mismo reconoce esta compatibilidad entre el Funcionalismo y el Dualismo, aunque sin tildarlo de cartesiano

Sin embargo, los estados funcionales de los sistemas son algo distintos [a los estados físico-químicos del cerebro]. En particular, la hipótesis de los estados funcionales, ¡No es incompatible con el Dualismo! Aunque resulta patente que la hipótesis es mecanicista en cuanto a su inspiración, es un hecho marcadamente notorio que un sistema que consiste en un cuerpo y un “alma”, si existe esto, podría perfectamente ser un automaton probabilístico (1973:76. Trad., mía).

Por otra parte, en dos pasajes, Descartes argumenta que resulta concebible separar cuerpo y mente y que esta última prima por sobre lo corpóreo de los que somos directamente conscientes mediante ideas claras y distintas. Enfatiza que el pensamiento no está adscrito al cuerpo, ni a lo mecánico por cuanto

[...] El *pensar* debe identificarse no sólo con entender, desear e imaginar, sino también con la consciencia de lo que sentimos. Debido a que si digo “estoy viendo, o estoy caminando, por lo tanto pienso”, y tomo esto como aplicable al ver y al caminar como actividades, entonces, la conclusión no es absolutamente cierta. Esto sucede porque, como a menudo pasa durante el sueño, es posible que piense que estoy viendo o moviéndome, aunque mis ojos estén cerrados o no me mueva en absoluto; tales pensamientos podrían ser posibles si no tuviese cuerpo alguno (1985b:195. Énfasis en el original. Trad., mía).

En el segundo pasaje de *Las Meditaciones* elabora mucho más la intuición modal, argumentando que conduce a una distinción *real* entre cuerpo y mente

Primero, porque sé que todas las cosas que concibo clara y distintamente pueden ser producidas por Dios tal como las entiendo. Por lo tanto, del hecho de que pueda entender clara y distintamente una cosa separada de otra es suficiente para darme certeza de que *las dos cosas son distintas*, pues son susceptibles de ser separadas, al menos por Dios. [...] Es verdad que podría

tener (o que podría anticipar tener) un cuerpo que está muy cercanamente unido a mí. No obstante, tengo una idea clara y distinta de mí, en cuanto a que soy simplemente una cosa no extensa pensante; y, por otra parte, tengo una distinta idea del cuerpo en cuanto esta es una cosa extensa no pensante. Y, en consecuencia, *es cierto que soy muy distinto a mi cuerpo, y que puedo existir sin este* (1985a:54. Trad., y cursivas mías).

La intuición modal cartesiana tiene, por tanto, un estrecho vínculo con cómo uno puede concebir dos cosas separadas y, luego, que una ni está adscrita a, ni es dependiente de la otra.

En vista de la defensa de la imposibilidad de reducir los estados mentales a propiedades físicas, Descartes y la IA clásica o *fuerte* se acercan notablemente, aunque con supuestos teóricos muy diferentes, pues el pensamiento y la racionalidad serían máximamente objetivos, aunque ninguno de éstos puede ser mecánico o automático para este filósofo. Por otra parte, conviene también aclarar que nunca postuló la existencia de una multiplicidad de realizaciones posibles del pensamiento, ni menos lo hizo teniendo en consideración la máxima funcionalista de que tales realizaciones son mecánicas y equivalentes en función de sus *inputs/outputs* y estados internos.

Justamente, es importante recalcar otra diferencia fundamental entre Descartes y la IA *fuerte*. En efecto, mientras que el primero afirma tener certeza de la distinción mente-cuerpo, en tanto tiene una idea clara y distinta de aquella, la segunda propone la hipótesis de que los estados mentales son los estados discretos de un *automaton* finito y esto no es incompatible con el Dualismo, toda vez que las similitudes en la conducta de un organismo llevan a concebir similitudes en la organización funcional sin que sea relevante especificar detalles físicos (Putnam, 1967:164-5).

En consecuencia, mientras Descartes propone una intuición modal de separación de mente y cuerpo, que se basa en la claridad y distinción de tal diferencia escrutada por la mente, se infiere de la IA *fuerte* que es concebible que existan seres con mente que no tengan realización material, como las entidades ectoplásmicas y angelicales, los cuales *podrían* tener la misma economía funcional humana y, por lo tanto, los mismos dolores y otros estados mentales. Esto último, al menos, despeja la duda de si el Funcionalismo cuenta como una genuina forma de Materialismo, como a veces se lo caracteriza⁹ y, ciertamente, corrobora el distanciamiento entre Babbage y la IA *fuerte*.

CONCLUSIÓN: LA PECULIAR SIMBIOSIS ENTRE DESCARTES Y LA IA CLÁSICA

Es claro, entonces, que Descartes y la IA *fuerte* proponen cosas diferentes e inspiradas por suposiciones teóricas distintas, aunque el principio de realizabilidad

⁹ Incluso, la cercanía entre el *fisicismo* no reductivo del Funcionalismo de Máquina de Turing y el Dualismo Cartesiano parece mayor si se considera que, al menos, resulta discutible que Descartes sea un genuino dualista de substancias como Clarke (2003) sostiene, dado el rol redundante y no explicativo del concepto de substancia y la identificación de éstas mediante propiedades.

múltiple del Funcionalismo resulte compatible con el Dualismo. Pese a que sólo Descartes posee una teoría desde la cual la intuición modal que separa cuerpo y mente resulta plausible, y ésta nunca puede contar como un mecanismo ni concreto ni abstracto, los argumentos de la IA clásica hacen referencia explícita a Descartes e intentan refutar de manera sistemática por qué una máquina no podría pensar.¹⁰

Aparte de la preeminencia que otorgan al lenguaje, tanto la IA clásica como Descartes, existe una gran coincidencia entre lo propuesto por ambos en relación con la irrelevancia de lo material en el caso de lo mental. De ambas posiciones se infiere que es innecesario que la inteligencia esté realizada en propiedades físicas, aunque la segunda considera que el pensamiento es mecanizable formal y computacionalmente. Si esto es así, podría darse el caso de que nuestra mente “flotase” en algún ambiente virtual, pudiésemos cambiar de *hardware* en caso de ser necesario, y descargar nuevamente los estados mentales en una implementación diferente, o en muchas. Por supuesto, en el ámbito contemporáneo, tales ideas —que se sintetizan en que cualquier máquina que funcione como un cerebro será un cerebro con los mismos estados mentales— han incentivado no sólo disquisiciones filosóficas sobre el problema metafísico del Funcionalismo, sino hasta trabajos en ciencia ficción.¹¹

En síntesis, la IA clásica o fuerte ha establecido una peculiar relación con Descartes, en tanto intenta refutar su *dictum* acerca de la imposibilidad de que alguna máquina sea pensante, pero al mismo tiempo no descarta de plano al Dualismo, cuyo exponente más notable es el filósofo francés. La IA no rechaza que el desempeño lingüístico y conductual sean determinantes para juzgar la existencia de inteligencia (Moor, 1976), ni tampoco que puedan haber mentes inmortales. Todo lo anterior le ha

¹⁰ En efecto, una buena cantidad de investigadores se propusieron mostrar que una máquina podía tener la flexibilidad de la razón humana y su competencia lingüística. Ejemplos sobran. Entre algunos que vale la pena destacar están programas como ELIZA de Weizenbaum (1984), o PARRY de Colby (1975), los cuales intentan mostrar que una máquina programada sí puede conversar con humanos. Si bien estos programas son en realidad ejemplos de cómo se podría pasar el Test de Turing, no es claro que involucren las habilidades cognitivas necesarias para resolver problemas que requieran inteligencia genuina. Como el propio Weizenbaum destaca, su programa muestra tres cosas: a) la ingenuidad de pensar que un programa podría tratar con los problemas emocionales y mentales humanos, lo que involucra compartir experiencias; b) el candor de cómo algunas personas lograban enganchar emocionalmente con una máquina programada, y de cómo podían antropomorfizarla; c) la inocencia de pensar que un programa como ELIZA podría representar una solución genérica al problema del entendimiento lingüístico (1984:5-8).

¹¹ La conexión entre la supuesta inmaterialidad del pensamiento y la inmortalidad del alma es explorada por Rudy Rucker (1982), quien en su notable novela *Software* caracteriza a Cobb Anderson, una especie de hippie futurista que alcanza la inmortalidad al almacenar toda su vida mental fuera de su viejo y deteriorado cerebro. Aquí también existe una importante coincidencia entre lo planteado por Descartes y la IA clásica, pues suponen que la irrelevancia de la realización de lo mental en lo material cuenta como un factor crucial para alcanzar la inmortalidad, ya que sólo lo material decae y está afecto a la corrupción. Así, mientras Descartes afirma que el pensar es indivisible, incorruptible y no mecánico, y eso hace inmortal al alma humana, algunos entusiastas seguidores de la IA clásica especulan sobre esta posibilidad en base al carácter sintáctico y formal de una MT.

valido críticas a la IA clásica de escindir el pensamiento y la inteligencia de un modo dualista de la corporalidad, del cerebro y de la materia, lo cual significa que Searle está en lo cierto, y que el Funcionalismo no apoya de manera consistente al monismo materialista que originalmente inspiró a la IA en el siglo XIX, con Babbage como su gran defensor. La sombra de Descartes y su intuición modal continúa proyectándose, entonces, sobre la IA clásica o *fuerte*.

*Universidad de Chile**
Facultad de Filosofía y Humanidades
Departamento de Filosofía y Centro de Estudios Cognitivos
Ignacio Carrera Pinto 1025, Ñuñoa. Santiago (Chile)
rodgonfer@gmail.com

BIBLIOGRAFÍA

- Block, Ned. "The Computer Model of the Mind", en D. Osherson y E. Smith (Eds.), *in invitation to Cognitive Science: Thinking* (Vol. 3). Cambridge, Mass.: MIT Press, 1990.
- Clarke, Desmond. *Descartes's Theory of Mind*. Oxford: Clarendon Press, 2003.
- Clark, Andy. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: OUP, 2008.
- Colby, Keneth. *Artificial paranoia*. New York: Pergamon Press, 1975.
- Copeland, Jack. *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell, 1993.
- "The Church-Turing Thesis", en <http://plato.stanford.edu/entries/church-turing/>, 2002. Consultado: Enero 2011.
- Cottingham, John. "Descartes treatment of animals", in *Descartes*. Oxford: Oxford University Press, 1998:225-233.
- Dennett, Daniel. *Consciousness Explained*. London: Penguin Books, 1991.
- *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Basic Books, 1981.
- Descartes, René. *Discurso del Método*. Madrid: Alianza, 1994.
- "Letter to More, 5 February 1649", in D. Rosenthal. (Ed.). *The Nature of Mind*. New York: Oxford University Press, 1991:36-7.
- "Letter to (the Marquess of Newcastle) 23 November 1646", en D. Rosenthal (Ed.). *The Nature of Mind*. New York: Oxford University Press, (1991):35-6.
- "Meditations on First Philosophy", in J. Cottingham, R. Stoothoff y D. Murdoch (Eds.). *The Philosophical Writings of Descartes. Vol. II*. New York: Cambridge University Press, 1985:1-62.
- "Principles of Philosophy", en J. Cottingham, R. Stoothoff y D. Murdoch (Eds.), in *The Philosophical Writings of Descartes Vol. I*. New York: Cambridge University Press, 1985:177-266.
- González, Rodrigo. "El Test de Turing: Dos Mitos, un Dogma", en *Revista de Filosofía*. Universidad de Chile, N° 63 (2007):37-53.
- Kripke, Saul. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, 1980.

- Lewis, David. "An argument for the identity theory", in *Journal of Philosophy* N° 63:17-25. Reimpreso en J. Heil (Ed.). *Philosophy of Mind: A Guide and Anthology*. Oxford: Oxford University Press, 1966:150-157.
- Moor, James. "An Analysis of the Turing test", in *Philosophical Studies* N° 30. 249-57. Reimpreso en S. Shieber (Ed.). *The Turing test: Verbal Behaviour as the Hallmark of Intelligence*. Cambridge, Mass.: MIT Press, 1976:297-306.
- Nagel, Thomas. "What is it like to be a bat?", in *Philosophical Review* N° 83. 435-50. Reimpreso en D. Chalmers (Ed.). *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press, 2002:219-226.
- Penrose, Roger. *The Emperor's New Mind*. Oxford: Oxford University Press, 1999.
- "Setting the scene: The claim and the issues", en D. Broadbent (Ed.). *The Simulation of Human Intelligence*. Oxford: Blackwell, 1992:1-32.
- Putnam, Hilary. "The nature of mental states", originalmente publicado como "Psychological Predicates", in W. Capitan y D. Merrill (Eds.). *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press. Reimpreso en D. Chalmers (Ed.). *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, 1973:73-79.
- "Psychological predicates", in W. Capitan y D. Merrill (Eds.). *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press. Reimpreso en J. Heil ed., *Philosophy of Mind: A Guide and Anthology*. Oxford: Oxford University Press, 1967:160-167.
- Rucker, Rudy. *Software*. New York: Harper Collins, 1982.
- Searle, John. "Twenty-one years in the Chinese Room", in J. Preston y M. Bishop (Eds.). *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford: Oxford University Press, 2002:51-69.
- *The Construction of Social Reality*. London: Penguin, 1995.
- *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press, 1992.
- "Is the brain's mind a computer program?", in *Scientific American*, January. (1990):20-25.
- "Minds, brains and programs", in *Behavioral and Brain Sciences* 3:417-24. Reimpreso en M. Boden (Ed.). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1980:67-88.
- Rumelhart, D.; Hinton, G. *et al.* "Learning internal representations by error propagation", in D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1. Cambridge, Mass.: MIT Press, 1986:318-62.
- Swade, Doron. *The Difference Engine: Charles Babbage and the Quest to Build the First Computer*. London: Penguin, 2000.
- Turing, Alan. "Can Digital Computers Think?". Típo de una entrevista radial en el tercer programa de BBC, del 15 de mayo de 1951. Número de referencia de los Archivos Turing: B.5. Reimpreso en S. Shieber (Ed.). *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*. Cambridge, Mass.: MIT Press, 1951:111-116.
- "Intelligent machinery, an heretical theory". Manuscrito inédito de una conferencia en la "51 Society" en Manchester, Inglaterra. Número de referencia de los Archivos Turing: B.4. Reimpreso en S. Shieber (Ed.). *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*. Cambridge, Mass.: MIT Press, 1951:105-109.

Rodrigo González

----- “Computing intelligence and machinery”, in *Mind* LIX, N° 2236 (Oct. 1950):433-60. Reimpreso en M. Boden (Ed.) *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1950:40-66.

----- “On computable numbers, with an application to the *Entscheidungsproblem*”, in *Proceedings of the London Mathematical Society*, series 2. Vol. 42:231-65. (Con correcciones en Vol. 43:544-6, 1936.

Weizenbaum, Joseph. *Computer Power and Human Reason*. London: Penguin, 1984.