

# Supuestos en la discusión acerca del pensamiento mecánico: la relevancia de la máquina universal

Pio Garcia- Javier Blanco

Universidad Nacional de Córdoba

## Introducción

Las tesis mecanicistas encontraron históricamente un límite en la consideración de la naturaleza del pensamiento. La respuesta negativa a la pregunta acerca de si las máquinas pueden pensar, parece fundarse en una comprensión tanto de lo que es pensamiento como de lo que las máquinas pueden hacer. Se supone que si bien puede haber dudas acerca de lo que es el pensamiento, lo que parece estar claro es lo que una máquina puede generar o hacer. Habitualmente esta pregunta se ha desarrollado en el contexto de la filosofía de la mente o del lenguaje. Ahora bien, ¿qué puede aportar el hecho de reubicar esta discusión en el contexto de la computación?

En la primera mitad del siglo pasado se comenzó a constituir un ámbito de indagación, investigación y problematización en torno a la noción de computación. Aquí confluían perspectivas teóricas, ingenieriles y filosóficas. En particular se encontró la tradición de construcción de máquinas con aquella que buscaba especificar la idea de cálculo efectivo en la matemática. La noción de procedimiento efectivo o cálculo mecanizable fue un logro crucial de principios de siglo XX. En cierta forma, este logro reavivó la discusión acerca de si las máquinas pueden pensar. Si bien este problema se puede rastrear en la tradición de la filosofía moderna (Leibniz y Descartes), es sin dudas el artículo ya clásico de Turing de mediados del siglo pasado el marcó la discusión contemporánea.

En el presente trabajo no vamos a tomar el problema general acerca de si las máquinas pueden pensar sino más bien vamos a analizar algunos de los supuestos particulares de esta discusión. En especial tomaremos en consideración una de las propuestas de Turing, la máquina universal, en tanto estimamos que la relevancia de este tipo particular de máquina no ha sido suficientemente destacada en la literatura. Estimamos que las implicaciones de la noción de máquina universal – y de la computación en general- puede permitir reubicar la indagación acerca de los supuestos filosóficos centrales en la aquella tradición que ha discutido el problema de si las máquinas pueden pensar. En particular analizaremos el supuesto de que siempre podemos conocer el comportamiento de una máquina. Esta consideración a priori acerca de lo que una máquina puede hacer aparece de diferentes maneras en Descartes e incluso se puede encontrar, al menos a partir de una cierta lectura, en la obra wittgensteniana.

## Turing y el pensamiento mecánico

En 1950 Alan Turing publica un controvertido artículo en el cual se preguntaba si las máquinas podían pensar. Al parecer algunas de las propuestas de dicho artículo, como el juego de imitación, estaban diseñadas como una respuesta a un objetor tradicional del pensamiento mecánico: Descartes. La prueba de corte conductista descrita por Turing se puede asociar directamente con la parte V del Discurso del método. Además de este test de imitación Turing sugería diversos argumentos para intentar socavar los prejuicios que descartaban la posibilidad de máquinas pensantes. Pero hay una propuesta, que Turing no la presenta como argumento sino como un contexto nuevo, que estimamos que se destaca sobre el resto. En primer lugar Turing no hablaba de cualquier máquina sino de las computadoras. Pero además destacaba la importancia de un tipo particular de computadora. Y, como veremos, este aspecto no es trivial.

Regresando a la prueba de imitación, hay una importante tradición que la ha interpretado en términos conductistas, probablemente porque el artículo de Turing se publicó por la misma época que El concepto de lo mental de Ryle o el Ciencia y conducta humana de B. F. Skinner. De cualquier manera la interpretación conductista del test de imitación parece natural y directa, aunque no parece que una perspectiva fuerte, como

el conductismo lógico, sea la más apropiada. Esto es, suponer que el juego de imitación da las condiciones necesarias y suficientes para decir que una máquina puede pensar.

Habría al menos tres formas alternativas de plantear el alcance de la interpretación conductista. Una respuesta es defender que Turing estaba comprometido con una condición de suficiencia, aunque no de necesidad, en su test de imitación (v.gr. Ned Block). De esta manera poseer un determinado comportamiento – indistinguible en un sentido relevante de los seres humanos- sería suficiente para señalar inteligencia.

Además de la condición de suficiencia, se ha argumentado que Turing defendería una condición de necesidad (Abramson 2011). Dicho aspecto estaría asociado con la predictibilidad de la conducta y se ha llamado la condición de limitación epistémica (Abramson 2011 p. 546).

Una forma distinta de interpretar este problema sería el siguiente. Entre las respuestas a las objeciones, Turing señala que podría haber máquinas que tengan un comportamiento que pueda ser visto como inteligente pero que no se parezca al de los seres humanos. Son, entonces, dos cuestiones diferentes: por un lado el que se puedan llevar adelante tareas no triviales de resolución de problemas y, en segundo lugar, el que dicha tarea se asemeje al modo humano de resolver estos problemas. Esta distinción será retomada en los años 70 en las disputas asociadas a la inteligencia artificial.

Regresando a las interpretaciones anteriores la condición de limitación epistémica tiene un interés especial para nosotros porque puede remitir de manera directa al comportamiento de la máquina universal. Pero antes de abordar esta cuestión realicemos un breve paréntesis para indagar acerca de un posible origen del test de imitación: la propuesta cartesiana.

## **Descartes: argumentos ontológicos y conductistas**

En lo que se refiere a la oposición entre mecanicismo y pensamiento, la tradición cartesiana suele ser una de las primeras en ser citada. Hay un primer argumento, de tipo principista, para defender esta oposición y que refiere al dualismo ontológico. El que haya una sustancia pensante con propiedades diametralmente opuestas a la materia explicaría por qué, en principio, las máquinas no pueden pensar. Pero, como es bien sabido, Descartes estaba preocupado por la forma en la cual los autómatas imitaban tanto el aspecto como el comportamiento de los seres humanos. Por tanto propone algunos test comportamentales para poder “probar” cuándo aquello que se parece a un ser humano es meramente una máquina. El más conocido es la prueba del lenguaje. Un animal- una máquina- puede imitar el lenguaje humano, pero son meramente sonidos que no deberíamos confundir con palabras genuinas. Estas últimas suponen un significado que no tiene la mera imitación del sonido. Nuevamente aquí vemos una reedición de la versión principista del argumento cartesiano – el dualismo – y una versión conductual. Si sabemos quién es el que profiere los sonidos, podríamos saber si hay un significado por detrás. Si este no es el caso, entonces lo que hoy se suele llamar la productividad del lenguaje sería aquello que nos permite realizar la distinción. La imitación se quedaría siempre corta con respecto a la variabilidad que supone la productividad del lenguaje.

Al parecer Turing leyó acerca de este test del lenguaje por medio de un artículo de Jefferson y en este sentido se puede entender el test de imitación como un desafío empírico a la propuesta cartesiana. Abramson analiza en un artículo reciente el eventual alcance de la relación entre Turing y Descartes. En el presente trabajo queremos destacar un aspecto, vinculado con la máquina universal, que no aparece en Abramson 2011. Este aspecto se puede plantear como un argumento. Recordemos que en lo que estamos interesados aquí es en el supuesto de que conocemos, a priori, lo que una máquina puede hacer.

Turing (1950) refiere a la objeción de Lady Lovelace. Ada Lovelace fue una matemática del siglo XIX, colaboradora de Charles Babbage y la primera programadora de una computadora. En un comentario muy citado, Lovelace dice que una máquina sólo puede hacer aquello para lo que fue programada. Esta objeción apunta directamente a lo que llamábamos más arriba la condición de limitación epistémica. Si se estima que un ser humano inteligente se distingue de una máquina por ser impredecible, por la posibilidad de generar algo nuevo con respecto a las condiciones iniciales, entonces parece que el comentario de Lady Lovelace estaría cerrando esta alternativa.

Pero Descartes ofrece algo más que un test de lenguaje. Sugiere asimismo una caracterización de lo que una máquina puede hacer – y lo que no-. Los animales – máquinas- pueden realizar diversas tareas mucho mejor que los seres humanos. Pueden ver mejor o pueden demostrar más fuerza que nosotros, pero actúan siempre en relación con un mecanismo. El comportamiento de una máquina siempre se explica por la relación entre las partes -finitas- de un mecanismo – como los órganos del cuerpo humano o de los animales.

La razón por otra parte es un “instrumento universal” y por tanto puede generar un comportamiento “inteligente”. Es importante notar que aquí hay al menos dos cuestiones. Por un lado está la universalidad de la causa que genera un comportamiento como el humano, pero también está la idea de que debemos tener alguna prueba “empírica” de dicho comportamiento. Y este último aspecto es el que se podría vincular con la objeción de Lady Lovelace. De esta manera, el primer aspecto del argumento se relaciona con la respuesta “en principio” de Descartes – el alma es la causa del comportamiento inteligente-. Y el segundo aspecto se relaciona con una condición más débil- la predictibilidad del comportamiento-. Para Descartes no hay ninguna dificultad en distinguir estos dos aspectos porque él está seguro de la respuesta a la segunda cuestión: una máquina no puede hacer nada que no podamos saber previamente indagando en el tipo de comportamiento que se sigue de su “programa”- la estructura del órgano o del mecanismo-. Y es este segundo supuesto el que queremos discutir. Si ya no somos dualistas -o al menos vemos a dicha posición difícil de defender- no tenemos ya un argumento principista sino sólo uno que descansa en nuestro conocimiento de lo que una máquina puede hacer. Y Turing, con su propuesta de la máquina universal ataca precisamente este último argumento.

### **Otra respuesta a priori a lo que una máquina puede hacer: normatividad**

Uno de los supuestos que la computación puede ayudar a identificar y discutir es el de la impredecibilidad de - algunas- máquinas. Vimos con Descartes que el supuesto de la no predictibilidad – o la condición de limitación epistémica- puede ser visto como uno de los núcleos del argumento de la indefinición del comportamiento a partir de una causa no limitada por lo particular – la materia-.

Complementaremos ahora nuestro análisis con otro contexto más complejo que ha sido sugerido para responder a priori a la cuestión de lo que una máquina puede hacer. Como vimos, a partir de la caracterización de procedimiento mecanizable o efectivo parece haberse configurado una concepción intuitiva y característica de máquina y de comportamiento mecánico. Se presupone entonces qué características y límites deberían tener los comportamientos generados por dicha máquina. Este parece ser el supuesto de filósofos como Stuart Shanker. En un artículo de fines de los 80 Shanker (Shanker 1987) sugiere que en el *paper* de 1936 Turing propone en realidad dos tesis: una de carácter matemático - la que se refiere a lo que es un procedimiento efectivo - y una tesis epistemológica - la que se refiere a si es posible afirmar que la extensión de lo computable es tal que las máquinas puedan pensar-. A esta segunda propuesta, supuestamente implícita en el citado artículo de Turing, la denomina Shanker la “tesis mecanicista”. Evidentemente Shanker vincula el Turing de 1936 con el de 1950. Como ya hemos dicho, no nos interesa aquí involucrarnos en la discusión de si las máquinas pueden pensar - uno de los temas centrales del artículo de Shanker- sino más bien con algunos de los supuestos de la argumentación. Si bien el objetivo del artículo de Shanker es contraponer las posiciones de Wittgenstein con las de Turing, en varios párrafos se muestra que Shanker pretende utilizar la argumentación wittgensteiniana para discutir en contra de algunas interpretaciones del pensamiento mecánico.

Parte de la argumentación del artículo de Shanker es como sigue. La única referencia explícita de Wittgenstein a las máquinas de Turing parece ser la de los *Remarks on the Philosophy of Psychology*. Aquí se habla de las máquinas de Turing como “humanos que calculan”. Si bien el sentido de la referencia, más bien irónico, no es del todo claro, en otros pasajes citados por Shanker de la obra wittgensteiniana se quiere diferenciar entre un proceso meramente empírico de “contar” y un proceso como el de calcular. En este último caso habría un aspecto normativo que no estaría presente en el primero (Shanker 1987, 619-620). La mencionada cita wittgensteiniana sobre Turing haría referencia entonces a esta confusión entre aspectos meramente mecánicos - computación- y aspectos normativos- típicamente humanos-.

Lo que queremos destacar especialmente es que la pregunta acerca de si las máquinas pueden pensar parece responderse antes de cualquier indagación acerca de los comportamientos que genere una computadora. La cuestión de si las máquinas pueden pensar es como la pregunta acerca del color de un número. No es una pregunta empírica. Parece entonces que, como se supone que se sabe de antemano lo que significa mecánico y por lo tanto lo que un dispositivo mecánico puede hacer, esto es qué tipos de comportamientos puede generar, entonces se puede responder a priori la pregunta acerca si dicha máquina puede tener un comportamiento que se denomine “pensamiento”.

Una forma de cuestionar esta perspectiva y la reseñada en la sección anterior dedicada a Descartes es haciendo referencia a las implicaciones de la máquina universal. En la propuesta de Turing de 1936 no sólo se proponía lo que después se iba a llamar máquina de Turing sino también la máquina universal.

## **Turing y la máquina universal**

Hay un acuerdo generalizado de que la noción de computación ha sido caracterizada de manera sustantiva por Turing en su artículo acerca de los números computables de 1936 (Turing 1936). El significado de lo que implica calcular de manera efectiva se entendería a partir de ahora como la ejecución paso a paso del funcionamiento de máquinas abstractas. Este acuerdo parece ir más allá de encontrar un formalismo para caracterizar una noción. Así parece entenderlo Gödel cuando dice que con Turing encontramos por primera vez una noción “absoluta” de computación.

Mucho se ha dicho sobre el alcance de la noción de “máquina” - y de “máquina de Turing”- para entender a la de “computación”. Menos explorado es el eventual aporte de la idea de máquina universal (MU) que Turing presenta en el mismo artículo de 1936. Por supuesto que ha sido ampliamente reconocida la importancia de las MU como una nota distintiva de la propuesta de Turing (Davis 2000). Lo que queremos problematizar es el aporte de las MU a la discusión acerca de si podemos saber a priori lo que una máquina puede hacer. Este aporte parece ir más allá de una mera sofisticación de las máquinas de Turing o de una extensión de su campo de aplicación. Las MU muestran algunos aspectos interesantes de lo que las MT pueden hacer.

Una máquina de Turing está definida a partir de un conjunto finito de estados, un alfabeto finito de símbolos de entrada (se suele usar un alfabeto binario) y una función de transición, la cual suele ser presentada como una tabla de transición. Los datos de entrada vienen en una cinta infinita (pero con una cantidad finita de símbolos no nulos), sobre la cual habrá una posición diferenciada que indicará el símbolo corriente. Cada una de las filas de la tabla indica para cada estado y símbolo de entrada dados cómo se cambia el símbolo corriente de la cinta, si se mueve para la izquierda o la derecha la posición corriente, y cuál será el nuevo estado.

El "computador" presentado por Turing para realizar las computaciones prescriptas por sus máquinas abstractas es una persona equipada con lápiz y papel que toma una tabla de transición como codificación de un comportamiento cuyos datos de entrada están la cinta, y va aplicando mecánicamente los pasos prescriptos en esa tabla. Como decíamos, cada paso indica posibles modificaciones a una posición de la cinta, un posible cambio de posición y cuál es el siguiente estado. Si se llega a un estado y un símbolo de entrada para el cual no hay ninguna regla en la tabla, el programa termina.

La MU es una máquina particular (dada por su propia tabla de transición) cuyo comportamiento consiste en recorrer la cinta leyendo el código de una máquina dada, y luego comportarse como dicha máquina tomando como entrada el resto de la cinta. En este sentido, la máquina universal puede imitar el comportamiento de cualquier otra máquina, al menos en lo que respecta al comportamiento de entrada/salida. Esta posibilidad depende de que se pueda codificar una MT y que luego esta codificación sea parte de los “datos” de entrada de otra MT particular que ahora funciona como una MU.

Una propiedad es que no podemos predecir el comportamiento de cualquier MT (aunque podamos predecir el comportamiento de algunas). Esto se sigue de la solución -negativa- al problema de la detención. Esto es poder predecir si una MT cualquiera dará un resultado - positivo o negativo- o no se detendrá nunca. Esta es una manera muy fuerte de establecer que no podemos saber a priori lo que hará una máquina (en este caso una computadora). Dado que la MU puede imitar a cualquier máquina y que no puede predecirse el

comportamiento de algunas de dichas máquinas (es decir, no hay manera mecánica de determinar conociendo la máquina y los datos para esa máquina, si va a producir un resultado o no), no puede predecirse en particular el resultado de evaluar una MU para un conjunto de datos iniciales (los cuales en particular pueden estar codificando alguna otra máquina). Es de algún modo curiosa la situación: disponemos de una máquina particular, la MU, para la cual podemos demostrar que es impredecible, que no hay ningún método para determinar a priori su comportamiento. Por supuesto que una manera de saber cómo se va a comportar es llevar adelante el proceso mecánico que ella prescribe, pero no hay ninguna garantía de que ese proceso arroje algún resultado observable.

## **Consideraciones finales**

En el presente trabajo hemos analizado algunos supuestos de la discusión acerca de si las máquinas pueden pensar. Responder a esta pregunta presupone que podemos dar cuenta de lo que significa pensar y que sabemos qué puede hacer una máquina. Vimos muy brevemente cómo esta cuestión aparecía implícita en la propuesta cartesiana y de qué manera se reconfiguraba en la discusión filosófica posterior que destacó la normatividad como aspecto central. Sin embargo, a pesar de las diferencias identificamos en ambas tradiciones un supuesto que tiene que ver con saber con anticipación qué puede hacer una máquina. En este sentido estimamos que el mecanicismo del siglo XVII no es el mismo que el del siglo XX. Las propiedades y el comportamiento resultante de esta máquina particular llamada computadora posiblemente exija una reevaluación del mecanicismo resultante. Nuestra intención no es defender la tesis de que las máquinas pueden pensar sino reexaminar a la luz de la computación el supuesto de que podemos tener un conocimiento a priori del comportamiento de una máquina.

## **Bibliografía**

Turing, A.M. (1936). "On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*. 2 (1937) 42: 230–265.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.

Davis, Martin (2000) *The universal computer. The road from Leibniz to Turing*, Norton.

Shanker, (1987) 'Wittgenstein versus Turing on the Nature of Church's Thesis', *Notre Dame Journal of Formal Logic*, vol. 28, no. 4.

Descartes, R.: (2003) *Discurso del método; estudio preliminar, traducción y notas de Bello Reguera, E.*; ed. Tecnos, Madrid.

Webb, J (1980). *Mechanism, Mentalism and Metamathematics*. Kluwer,.