

## CAPÍTULO V

### GÉNERO, IMITACIÓN E INTELIGENCIA: UNA REVISIÓN CRÍTICA DEL ENFOQUE FUNCIONALISTA DE ALAN TURING

*Rodrigo A. González Fernández*

#### Introducción

Desde que el juego de la imitación vio la luz (Turing, 1950), renació una importante discusión en filosofía de la mente: ¿Es posible en principio la inteligencia de máquina? Tal como analizo aquí, la simplificación de la versión estándar de dicho juego ha tenido un impacto negativo sobre la apreciación de la teoría de la mente que inspira a Turing, lo que a su vez ha eclipsado su proyecto funcionalista del niño-robot. Este trabajo, dividido en cuatro secciones, revisa el Test de Turing, mostrando las dificultades de la versión estándar, las virtudes de la versión de género y el problema al que lleva el proyecto funcionalista del niño-robot.

Justamente, en la primera sección del ensayo se discute el objetivo central del juego de la imitación, en función de la disputa conceptual de la inteligencia de máquina heredada a partir de Descartes, quien descarta de plano la posibilidad de la inteligencia de máquina. En la segunda sección se critica cómo la versión estándar del juego omite del todo el género. En relación con este punto, argumento que, entendido en sentido minimalista como el sexo de los participantes, es crucial para comprender qué tiene en mente Turing con imitar, y también para ponderar correctamente su proyecto: la programación, el aprendizaje y la educación como medios para crear inteligencia de máquina. En la tercera sección se ofrecen



razones sobre por qué Turing se basa en la imitación, relacionándola con cómo la inteligencia de mujeres y hombres puede aprenderse por parte de máquinas programadas, de un modo análogo al que lo harían niños-robots. Finalmente, en la última sección, se critica cómo la imitación de inteligencia de la forma hacerse-pasar-por, fundamento tanto del test como del proyecto del niño-robot, es insuficiente para la acción inteligente en la ontología social; en efecto, como argumento, esta no solo requiere conducta inteligente sino además involucrarse-conscientemente-en.

### El Test de Turing: un método empírico para poner punto final a un debate conceptual cartesiano

Es claro que la propuesta de Turing se inserta de lleno en la filosofía de la mente. Hay consenso respecto de que su test pretende poner punto final al problema filosófico de si una máquina puede pensar, o puesto de otra manera, si la inteligencia de máquinas es posible en principio. Esta es justamente una pregunta de inspiración cartesiana que ya había provocado una larga y controvertida discusión, la cual ciertamente no fue resuelta antes y, a mi juicio, sigue provocando polémica por la naturaleza del problema examinado. En efecto, Descartes en diversos pasajes sostiene que es imposible en principio que una máquina tenga estados mentales como los de un ser inteligente, con racionalidad. En relación con el nexo entre lenguaje e inteligencia, Descartes (1637) indica que: “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” (p. 113). Es decir, una máquina podría usar palabras, pero de modo mecánico y automático, sin ser sensible y flexible al contexto y ambiente.

Justamente, para el filósofo francés, animales y máquinas carecen de inteligencia humana, porque solo pueden reaccionar mecánicamente frente a los estímulos ambientales. A diferencia de ambos, el ser humano es capaz de ordenar las palabras en formas

diferentes para significar exactamente lo mismo. Esto se debe a que el lenguaje, puesto de una manera simple y breve, implicaría racionalidad según Descartes, y esta, a su vez, sería signo inequívoco de flexibilidad e inteligencia. En consecuencia, el lenguaje contaría como una prueba empírica confiable de la existencia de racionalidad y, por tanto, de inteligencia.

No obstante, también sostiene que la inteligencia también requiere flexibilidad en la conducta. Con respecto a la relación entre conducta e inteligencia, argumenta lo siguiente:

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 (Descartes, 1637, p. 113).

Luego, la ausencia de lenguaje y la conducta mecánica en animales representa evidencia empírica de que ellos y las máquinas no son inteligentes a la manera humana. Por supuesto, la calificación de no inteligentes de animales y máquinas está asociada con la distinción metafísica entre dos substancias propuesta por Descartes. De hecho, los animales son máquinas, en la medida que su corporalidad, su *res extensa*, está constituida por complejos mecanismos que reaccionan de manera puramente causal al ambiente. La *res cogitans*, por el contrario, es la razón flexible e inteligente, capaz de producir conducta apropiada para la resolución de problemas complejos, cuestión que no sucede por una mera relación causal efecto, sino por la participación del entendimiento. Ciertamente, la relación causa-efecto se da entre cuerpo y ambiente, e indica por qué el primero reacciona de modo mecánico, automático y causal de manera totalmente predecible, en función de la constitución biológica, a diferencia de lo que ocurre con el entendimiento.

A la postura que niega mente e inteligencia a animales y máquinas se le ha llamado conductismo cartesiano (Clark, 1982). De



cuando con este enfoque, la inteligencia es producto de un tipo específico de racionalidad ajena a lo biológico, que además de no poder ser imitada, involucra que los animales carecen de estados mentales intencionales. Tal interpretación de Descartes sin duda ha provocado polémica, por ser no solo contraintuitiva, sino además poco consistente con lo que puede observarse en la propia conducta animal.

Pese a la inspiración cartesiana de un filósofo posterior como de La Mettrie, es sin duda un notable adversario del conductismo cartesiano. Con un estilo provocador, este materialista declara que:

Ser una máquina y sentir, y ser capaz de distinguir lo correcto de lo incorrecto, como el azul del amarillo —en una palabra, nacer con inteligencia y un instinto certero de moralidad, y ser solo un animal— son así cosas que no son más contradictorias que ser un simio o un loro, y ser capaz de darse placer a uno mismo (de La Mettrie, 1747, p. 54).

Hubo muchos otros a quienes también impactó el dualismo cartesiano que divide el mundo entre *res cogitans* y *res extensa*, y más aún, el tratamiento de los animales como mecanismos totalmente carentes de inteligencia. En general, la simpatía por el materialismo, asociado al avance de las ciencias, hizo que la filosofía cartesiana fuera sometida progresivamente a crítica. Por ejemplo, un famoso materialista decimonónico, Charles Babbage, quien es considerado uno de los padres de la inteligencia artificial, planteó en diversas ocasiones que no existe imposibilidad en principio de que una máquina emule la inteligencia matemática. Intentó construir dos máquinas para mostrar que las máquinas inteligentes no son imposibles en principio. La máquina de las diferencias, pensada para reemplazar los computadores humanos, y la máquina analítica, diseñada para realizar tareas no solo matemáticas.

A pesar de estos proyectos, la postura de Babbage es cauta cuando se trata de determinar si la máquina analítica en efecto tiene estados mentales, o solo imita la inteligencia matemática. Tóme-

se en consideración la cautela de estas declaraciones en el siguiente pasaje de Swade (2000) en donde alude a Babbage:

Babbage habla de “enseñarle a la máquina a prever”. En otras ocasiones afirma que la máquina “sabe”. El tenía la suficiente perspicacia para saber si este uso era adecuado, y evidentemente sentía que antropomorfizar mecanismos requería de cierta justificación o excusa: La *analogía* entre estos actos y los procesos mentales me forzó al uso *figurativo* de tales términos. El uso de estos fue ponderado como *económico y expresivo*, y prefiero seguir usándolos que sustituirlos por largos circunloquios (En Swade, 2000, pp. 103-4, énfasis míos).

En vista del mentalismo cartesiano y de la cautela de Babbage, la estrategia adoptada por Turing en 1950 no consistió en proponer un nuevo argumento basado en la discusión de términos y conceptos como “piensa” y “máquina”. Por el contrario, diseñó un test, un método empírico preciso, que fuera capaz de reemplazar la pregunta: ¿Pueden pensar las máquinas? Pero, ¿por qué Turing quiere abandonar la arena conceptual, en consideración de que su discusión entra de lleno en la filosofía de la mente y, en particular, en el problema de las otras mentes?

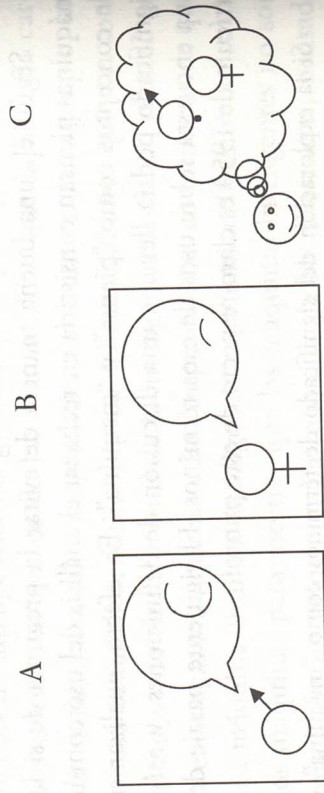
Según él, una buena manera de evitar la pregunta de si las máquinas piensan consistiría en rechazar el análisis del uso común de conceptos como “piensa” y “máquina”. En efecto, analizar su significado podría llevar a una discusión de definiciones, y así a una encuesta sobre usos de esos términos. El siguiente pasaje del artículo de 1950 es claro respecto de este punto:

Si la exploración del significado de términos como “máquina” y “piensa” se debe efectuar a partir del análisis de cómo estos se usan regularmente, es difícil evitar la conclusión de que el significado y la respuesta a la pregunta: ¿Pueden pensar las máquinas?, debe encontrarse a través de una investigación estadística similar a una encuesta Gallup (Turing, 1950, p. 40).



Para evitar la discusión conceptual y una eventual encuesta propone un método empírico que reemplace la polémica pregunta. El test se basa en un juego, a saber, el denominado juego de la imitación. Consiste en que hay un hombre en una habitación (A), una mujer en una segunda habitación (B) y un juez cuyo sexo no tiene importancia (C). En rondas sucesivas, el juez formula preguntas sencillas mediante un teletipo, con el objeto de identificar el sexo de los participantes. Las preguntas son simples y orientadas a ese fin, por ejemplo: ¿Tiene usted el pelo largo? ¿Usa maquillaje en algunas ocasiones? Mientras que la mujer responde de manera sincera a las preguntas del juez, el hombre se debe hacer pasar por una mujer. El objetivo del hombre, entonces, es que los jueces se equivoquen y no logren identificarlo correctamente. Así, el juego de la imitación, al menos en su primera etapa, es que un hombre induzca al error en su identificación, y que sea confundido por una mujer. Esta, en cambio, responde de manera sincera. La primera etapa puede esquematizarse así:

**Figura 1**  
**Primera etapa del juego de la imitación**



Pero, ¿qué tiene que ver el problema del género con la imitación de la inteligencia? Más aún, ¿por qué Turing pasa a la segunda etapa? Mucho se ha elucubrado sobre este cambio, por ejemplo,

viendo un proyecto político emancipador de género (Genova, 1994; Lassègue, 1996). A diferencia de estas interpretaciones, estimo que un indicio para entender el cambio a la segunda etapa son las tres preguntas que la motivan:

¿Qué sucederá cuando una máquina juegue el rol de A en este juego? ¿Decidirá el juez incorrectamente tan a menudo como cuando se juegue así a como se lo hará cuando en el juego participen un hombre y una mujer? Estas preguntas reemplazan nuestra pregunta original “¿Puede pensar una máquina?” (Turing, 1950, p. 41).

En vista de la forma en que Turing reemplaza la pregunta cartesiana sobre la inteligencia de máquina, al centrarse en la imitación, la siguiente sección examina cómo el juego propuesto por él representa una visión vanguardista con respecto a la inteligencia: es imitable e independiente de la biología, o de cualquier otra realización material.

### El rol del género en el juego de la imitación: ponderando la justa medida

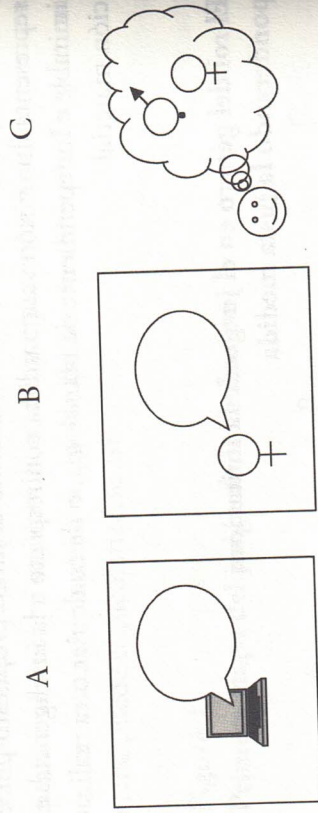
A diferencia de Descartes, Turing estima que existe evidencia para aseverar que una máquina tiene estados mentales y es inteligente. Por ello es un defensor de la llamada inteligencia de máquina, concepto basal de la inteligencia artificial según la cual el pensamiento mecanizado sí es posible en principio. No obstante, y tal como se analizará con mayor profundidad en la tercera sección, no es necesario que una máquina tenga propiedades biológicas específicas, o que tenga cerebro para que sea inteligente. Por el contrario, la inteligencia sería una propiedad mental ajena a las propiedades físicas de un sistema cognitivo y así, a lo biológico.

Para comprender el enfoque funcionalista y anti-biológico de Turing, y especialmente su propuesta vanguardista sobre la inte-



ligencia de máquina, es necesario examinar con más detención el juego de la imitación y discutir algunas posturas que señalan que lo central de este es precisamente la cuestión del género. En relación con este punto, cabe destacar que el juego no apunta al problema de la identificación de especies, sino que originalmente se centra en el género, esto es, en el sexo de los participantes, una cuestión que no es ni casual ni trivial. Lo anterior queda de manifiesto en la segunda etapa que Turing describe, donde se puede inferir que un computador digital podría reemplazar lo que hace el hombre en A, tal como este esquema muestra:

**Figura 2**  
**Segunda etapa del juego de la imitación**



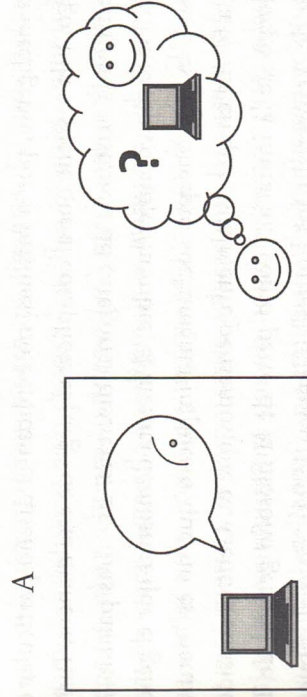
Naur (1986), Hayes y Ford (1995) y Traiger (2003) denominan a esta etapa la "versión de género" del test. Tal como ellos, considero que el género no es una pista falsa [*red herring*] y que no puede ignorarse del todo, tal como algunos aconsejan con la versión estándar (Davidson, 1990). Cabe destacar que los argumentos en pro y en contra de simplificar el test son tan discutidos como la eficacia de este.

Por una parte, los defensores de la versión estándar argumentan que hay cuestiones crípticas en el rol que juega el género en el test, y que Turing describe de modo algo ambiguo que en la etapa 2 la

máquina programada debe jugar "tal como [A]" [*like this*], de la Figura 1 (Shieber, 2004, p. 101). Asimismo, insisten en que en 1952 el propio Turing describe una versión simplificada de su test, dejando de lado la cuestión del género. Tal versión propone el reemplazo de la pregunta sobre si las máquinas piensan así: "Quisiera proponer un tipo particular de test que uno podría aplicar a una máquina [...]. La idea del test es que la máquina trate y pretenda ser un hombre" (Turing, 1952 en Copeland, 2003, p. 6). Con base en esta simplificación, se ha tomado la versión estándar como la esencia del test, ignorándose la cuestión del género.

La versión estándar supuestamente cuenta con la ventaja de centrarse solo en la cuestión de la imitación por parte de una máquina, tal como se ilustra a continuación:

**Figura 3**  
**Versión estándar del juego de la imitación**



Pese a la simplicidad de la versión estándar, parece central en el juego de la imitación cómo un hombre puede imitar la conducta lingüística inteligente de una mujer, y no solo como una máquina puede hacerse pasar por un humano al imitar su conducta lingüística (o como una máquina puede hacerse pasar por cualquier especie). De hecho, según la predicción de Turing, para el año 2000 el computador digital, con más velocidad y capacidad de almace-



namiento, debería haber engañado al menos al 30% de los jueces (Turing, 1950, p. 49).

Es importante enfatizar que la simplificación de la versión estándar ha dado lugar a toda clase de interpretaciones incorrectas sobre la filosofía y el proyecto de largo plazo de Turing. Aunque parece más “simple y neutral” que las etapas 1 y 2, la versión estándar soslaya del todo el género, entendido de manera minimalista: la imitación del sexo de los participantes. Por supuesto, la alusión al género, entendido de manera no minimalista, también ha generado interpretaciones polémicas.

Por ejemplo, Judith Genova (1994) sostiene que podría existir una abierta declaración de Turing a favor de seres inteligentes con nuevas oportunidades de género y sexuales, y esto podría estar inspirado por el pensamiento de Charles Darwin, quien también esperaba la evolución de nuevas formas dentro de un contexto naturalista. Una síntesis de la hipótesis de Genova es la siguiente:

Genova cree que las nociones sexistas acerca de la mujer siendo menos inteligente, por sí mismas, no explican el diseño particular del juego. Ella sostiene que al complicar el juego de esta manera, Turing cuestiona la existencia de categorías discretas. En otras palabras, al referirse al tópico mujer/hombre, él intenta demostrar que el género mismo es un concepto socialmente impuesto que no es “natural” de la forma en que usualmente pensamos que es. Genova considera el juego de la imitación como parte de la filosofía general de Turing de “transgredir los límites” (Genova, 1994). Bajo el supuesto de que Turing admirara tales transformaciones que no encajan con las categorías discretas, Genova sugiere que Turing podría estar discriminando a la mujer como un ser pensante de categoría inferior por ser incapaz de engañar a un hombre (Saygin, Cicekli y Akman, 2003, p. 59).

Otro ensayo que sostiene otra hipótesis muy discutible, y que se apoya en cuestiones psicológicas y biográficas de Alan Turing<sup>1</sup>, es el de Lassègue (1996). Por ejemplo, resulta curioso que la fecha del juicio a que Turing fue sometido por conducta inmoral, a causa de su relación homosexual, haya prácticamente coincidido con la entrevista que dio a *BBC* en 1952, en enero de ese año, y que en ese entonces haya habido preocupación pública por el descubrimiento de dobles agentes homosexuales trabajando para la KGB. ¿Habrá motivado esto que Turing obviara la cuestión del género? ¿Serán solo coincidencias?

Un análisis riguroso del juego de la imitación aconseja no adoptar la versión estándar, que omite el género del todo. Pero tampoco parece aconsejable exagerar tal elemento al extremo de ver un proyecto político. Por el contrario, el género es relevante para caracterizar por qué la imitación es suficiente para recabar evidencia acerca de la inteligencia de máquina. Cabe recordar que el objetivo de Turing es proponer un método empírico para detener la discusión filosófica de índole cartesiana respecto de la posibilidad de que máquinas piensen y sean inteligentes. La apreciación de este objetivo se refuerza cuando se identifica el *quid* central de dicho juego, a saber, el engaño [*deception*] mediante imitación, criterio decisivo para juzgar la inteligencia al comparar el desempeño de una mujer con hombre y máquina (Saygin *et al.*, 2003, p. 26). Por este motivo, es importante aclarar que hacerse pasar por una mujer (y eventualmente por un hombre) es una propiedad que, al igual que la inteligencia, no requeriría de implementación material específica, como la biológica. Luego, el punto central de Turing es que la imitación de conducta lingüística inteligente en función de un estándar con el cual comparar sería suficiente para determinar,

<sup>1</sup> Supuestamente, la vida de Turing sería relevante con respecto a las teorías que propone. Por ejemplo, se ha interpretado que su condición homosexual tuvo que ver con sus desarrollos teóricos. Para una síntesis de su vida, véase Hodges (1983).



inductivamente<sup>2</sup>, si una máquina programada es inteligente, pese a no tener sexo, ni poseer propiedad biológica alguna. En consecuencia, Turing cree que la evidencia observable aportada por el juego de la imitación es suficiente para exorcizar el mentalismo cartesiano y, así, para detener la disputa conceptual al que este da lugar.

El propósito de la siguiente sección es justamente examinar sucintamente qué filosofía habría motivado a Turing a introducir la cuestión del género y su relación con la imitación y la inteligencia, elementos centrales de su concepción anti-biológica de la mente.

### La filosofía funcionalista y anti-biológica de Turing: el nexo entre imitación, género e inteligencia

Dos antecedentes serán centrales para sustentar mi posición respecto del juego de la imitación planteado por Turing, quien critica que lo biológico determina crucialmente la inteligencia. Mostraré que los trabajos de Turing de 1936 y de 1950 son los pilares fundamentales del funcionalismo, una teoría de la mente y la inteligencia, y vincularé dicho funcionalismo con un argumento que presenta Turing respecto del aprendizaje y la educación: cómo podría programarse un computador digital a aprender del mismo modo en que un niño se transforma en una mente adulta inteligente, es de-

cir, ocupando mecanismos simples, que guiados por la educación, ayuden a crear aquella (Turing, 1950, p. 62).

En particular, argumentaré que el aprendizaje mediante programación que tiene en mente Turing representa una perspectiva tan vanguardista como polémica respecto de la inteligencia, el funcionalismo, una filosofía anti-biológica.

### El funcionalismo de máquina de Turing

Además de ser uno de los padres de la inteligencia artificial, Turing es un precursor de un enfoque filosófico denominado funcionalismo. La tesis central de esta postura filosófica es que la inteligencia es una función no biológica, puesto que lo que interesa respecto de los estados mentales son las relaciones de *input/output* y los estados internos de una máquina. Todo esto puede ser descrito mediante el funcionamiento de una máquina de Turing, el fundamento conceptual para entender qué es una máquina programada o computador digital.

Una máquina programada se caracteriza por tener carácter multipropósito, es decir, por tener la capacidad de resolver problemas de diferente índole. Una máquina de Turing, a su vez, es una idealización matemática de un dispositivo mecánico-abstracto y la base para entender qué es una máquina programada. Junto con operar sobre la base de quintuplos<sup>3</sup> descriptibles por una tabla (ver Turing, 1936, pp. 231-235), la máquina de Turing puede vincularse con el sentido intuitivo que posee computar (ver Turing, 1936, pp. 250-253). Esta actividad consiste en la ejecución de un algoritmo, cuyos pasos son finitos y calculables mediante lápiz y papel, sin la participación de la conciencia. Para ejecutar un algoritmo la máquina opera de modo mecánico, ya que la tabla dicta el proceder.

<sup>2</sup> El elemento inductivo del Test de Turing puede apreciarse cuando se sopesa la predicción que él hace: "Creo que en cerca de cincuenta años será posible programar a un computador, con una capacidad de almacenaje de 10<sup>9</sup>, para hacerlo participar en el juego de la imitación tan bien que un juez promedio no llegará a tener más de un 70% de probabilidades de hacer la identificación correcta luego de preguntarle por cinco minutos" (Turing, 1950, p. 49). Moor (1976 y 1987) y Copeland (2003) han interpretado el Test de Turing en un sentido inductivo, debido a esta predicción. Según ellos, lo que Turing tiene en mente no es definir la inteligencia, ni tampoco proponer condiciones necesarias o suficientes para esta. Por el contrario, el hecho de que participen diferentes jueces en varias rondas, y que en cincuenta años estos no tengan más de 70% de probabilidades de descubrir a la máquina programada, reafirma que es un método empírico para evaluar la inteligencia. En vista de la evidencia inductiva, tanto Moor como Copeland privilegian solo la versión estándar del juego de la imitación. Las razones de esto están expuestas en Copeland 2003, pp. 6-8.

<sup>3</sup> Más abajo me refiero a qué son los quintuplos, cuando se describe de qué manera opera una máquina de Turing en función de una tabla, o un programa.



Tal como sucede con un algoritmo, el término de la ejecución de este ocurre cuando la máquina de Turing logra resolver un problema mediante procedimiento efectivo. Un "computador humano" que emplea lápiz y papel, también puede definirse en términos de una máquina de Turing, o del cálculo Lambda. Este establece operaciones de abstracción y generalización que son realizables por una máquina de Turing, hasta que se alcanza una solución mediante computaciones recursivas (ver Copeland, 1993, p. 230).

Tanto la máquina de Turing como los algoritmos permiten caracterizar qué es computar. En realidad, esta máquina es una idealización de qué es computar, o de la ejecución de un algoritmo mediante la implementación de un programa. Por este motivo, usualmente se representa a la máquina de Turing mediante una cinta infinita y una cabeza lectora-escritora (o *scanner*) que se mueve hacia la derecha y la izquierda, cambiando de estado interno. Los estados en que se encuentra la máquina de Turing están totalmente condicionados por los símbolos leídos, o estados discretos, y los mencionados estados internos. Por ejemplo, la cabeza se encuentra en un tiempo  $t$  en un estado interno específico  $(q_0, q_1, q_2, \dots, q_n)$ , luego un símbolo de la celda es leído  $(b_1, b_2, b_3, \dots, b_n)$ . Dado el estado interno, lo que se lee en la cinta y el micro código de la cabeza, esta mantiene el símbolo, o bien lo borra, imprimiendo otro. Luego, la cabeza para o se mueve a otra celda de la cinta, continuando las computaciones hasta que se detiene. Si esto sucede, se ha llegado al final del procedimiento y, por tanto, a la resolución de un problema.

Tradicionalmente, se representan los símbolos de una máquina de Turing mediante 0s y 1s. Cabe destacar que su conducta está absolutamente condicionada por el programa, descrito mediante la mencionada tabla. Si la máquina está en un estado interno y lee una celda de la cinta, esta dupla implica un input  $I_n$ : <estado interno, símbolo leído>. El programa, entonces, dicta una tripleta de *output* en función del programa: <escribe símbolo, movimiento (o detención) y nuevo estado interno>. Este  $O_n$  hace que la máquina escriba un 1 o 0 en una celda, que se mueva izquierda o derecha, ingrese a un nuevo estado, o que se detenga.

Las dupletas de *input* y tripletas de *output* pueden describirse en forma de quintuplos y enunciados condicionales, todo lo cual ayuda a que se conforme la lista finita de reglas ordenadas de pares <*input*, *output*>. Los quintuplos tienen la forma <estado actual, leer símbolo, escribir símbolo, moverse (o detenerse) y nuevo estado>. Es pertinente enfatizar que los símbolos de una máquina de Turing pueden representar objetos que no necesariamente sean matemáticos, y esto permite que exista computación de problemas de diverso orden.

Una cuestión crucial respecto de la máquina de Turing es que puede imitar el comportamiento de cualquier otra máquina, sin importar los materiales con que están confeccionadas. Lo único que se requiere es que ambas funcionen de la misma manera, es decir, que implementen el mismo programa. Así, una máquina de Turing puede imitar la conducta de un cerebro, pero no requiere de la "papilla neural" para tales efectos. Y tal cuestión representa una de las tesis centrales del funcionalismo de máquina de Turing: para tener la inteligencia de una mujer, no es necesario tener sus propiedades físicas o biológicas. Basta con que el funcionamiento de su máquina-cerebro y del computador digital sean similares, tesis de Turing que luego sistematiza y defiende Purnam (1973).

El siguiente extracto de una entrevista radial a Turing en 1951 es iluminadora respecto de la relación entre imitación de inteligencia y funcionalismo:

Para lograr que nuestro computador imite a una máquina solo es necesario programarlo para que calcule lo que la máquina en cuestión haría bajo ciertas circunstancias [...]. Ahora bien, si una máquina en particular puede describirse como un cerebro, tenemos que solamente *programar* nuestro computador digital para imitarlo y *también será un cerebro*. Si se acepta que los cerebros reales, descubiertos en animales, y en especial en el hombre, son una clase de máquina, se seguirá entonces que nuestro computador digital, debidamente programado, *se comportará como un cerebro*. Este argumento presupone una idea que puede ser razonablemente cuestionada [...] que esta máquina debiera ser de una naturaleza cuya conducta sea en



principio predecible mediante cálculo [...]. Nuestro problema es, entonces, cómo programar una máquina para *imitar al cerebro*, o si lo pudiésemos expresar de una manera más breve y menos rigurosa, *para que piense* (Citado en Copeland, 2003, p. 11, énfasis míos).

Aquí se aclara aún más qué tiene Turing en mente en la primera etapa del juego de la imitación, la cual ha originado tanta confusión. Para que un computador o un hombre imiten la conducta inteligente de una mujer, no es necesario que se repliquen ciertas propiedades físicas o biológicas, a saber, aquellas por las cuales se tiene mente y conducta lingüística femenina. Para ponerlo de manera cruda y clara: el punto de Turing es que tal como no es necesario replicar las propiedades físicas de la papilla neural del cerebro para tener la mente y las respuestas inteligentes de este, la primera etapa del juego muestra que tampoco es necesario tener el aparato endocrino y reproductor de una mujer para poseer su mente e inteligencia<sup>4</sup>. Basta que el computador o el hombre imiten la inteligencia de la mujer para tener esta, tal como la mujer podría eventualmente imitar la inteligencia de un hombre; similarmente, no se requeriría poseer las propiedades físicas o biológicas asociadas al sexo masculino<sup>5</sup>. La pregunta que cabe hacer, entonces, es si

tal como la inteligencia, la conducta lingüística orientada por los roles de género puede aprenderse, de un modo similar a lo que Turing plantea con respecto a la creación de la mente adulta. Esto es, mediante la puesta en práctica de mecanismos simples que operen sobre la experiencia y la educación como inputs, que implementen aprendizaje basado en estímulo-respuesta, pero centrado en definiciones y proposiciones expresadas por un lenguaje simbólico, con el objeto de dejar de lado aspectos más bien emocionales (Turing, 1950, p. 63). Si esto fuera factible, la filosofía de Turing y su teoría de la inteligencia serían los pilares de una concepción tan vanguardista como controvertida de la mente, que toca aspectos como el aprendizaje y la educación, todo lo cual haría desestimar que género e inteligencia están puramente determinados por la biología.

### Turing, el aprendizaje y la educación: del género al proyecto de crear inteligencia de máquina

Tal como se examinó en la subsección anterior, el juego de la imitación se apoya en una concepción funcionalista de la mente. Puede seguirse de esta propuesta que aprender y ser educado se complementan al crear cierto tipo de inteligencia, la cual no sería un mero producto biológico. Por el contrario, al no importar el material de una máquina que puede imitar la conducta, la inteligencia no está constreñida por lo biológico y lo material. Por esto, su filosofía está lejos de posiciones materialistas contemporáneas, que identifican lo mental con lo neural y que incluso supone que si las mismas propiedades físicas de un cerebro se replican, se tendrá la inteligencia de este. Tómese en consideración, por ejemplo, un experimento mental desarrollado por Block (1990), a propósito de si la inteligencia requiere intencionalidad, la propiedad de algunos estados mentales, como creencias y deseos, de representar y ser acerca de:

<sup>4</sup> Por extraña que parezca esta cuestión, ciertamente se da en el ámbito empírico con el fenómeno de la transexualidad, esto es, de mentes que no corresponden con el cuerpo. Como puede apreciarse, curiosamente hay cierta cercanía entre el dualismo de Descartes y funcionalismo de Turing (González, 2011), pues cuerpo y mente ni son idénticos ni se condicionan mutuamente desde un punto de vista causal. Al contrario, son separables. Debo la sugerencia sobre la transexualidad a Alfredo Minervini.

<sup>5</sup> En relación con esto, difiero de Genova (1994). El funcionalismo anti-biológico de Turing hace pensar que no hay prejuicio contra la mujer, como incapaz de imitar la inteligencia de un hombre. En efecto, dado que imitar la conducta lingüística no requiere de ninguna propiedad biológica, ¿por qué la inteligencia masculina no podría ser imitada por una mujer también? Esto, además, podría ser una pista de por qué Turing abandonó la versión de género en 1952. En efecto, si se comprende el funcionalismo anti-biológico en pos de la inteligencia de máquina, resulta algo engorroso introducir la cuestión del sexo de los participantes. Mi tesis minimalista acerca del género es compatible con tal posición: el sexo y la conducta lingüística inteligente parecen un buen punto de partida para ejemplificar por qué lo biológico es irrelevante. Copeland (2003, pp. 6-7) también defiende una hipótesis sobre el abandono de Turing del tópico del género.



Imagínese un evento con una probabilidad bajísima, pero no cero: dado cierto movimiento aleatorio de moléculas, las partículas de un pantano se juntan y por azar resultan en un duplicado molécula por molécula igual a usted. La criatura del pantano tendrá todas las capacidades (conductuales) que usted tiene, y ellas producirán los mismos procesos psicológicos que ocurren en usted. Luego, podría ser (controvertidamente) inteligente. Pero hay razones para negar que la criatura tiene la intencionalidad que usted tiene, y ciertamente para negar que tiene estados intencionales del todo. La criatura del pantano dice, como usted: "Gorvachov influenció a Thatcher cuando viajó a Inglaterra". Pero, a diferencia de usted, la criatura nunca ha visto a Gorvachov o a Thatcher (o a alguien) en televisión, o ha leído acerca de ellos en los periódicos [...]. La criatura del pantano está simplemente *repietiendo palabras*. Si hubiera tenido sus moléculas ligeramente distintas estaría diciendo: "Los sobres Sr. Tattoo Eisenhower están sobre Nepruno" (Block, 1991, p. 263, énfasis mío).

Nótese que aunque Block y Turing coinciden en que hay (controvertidamente) inteligencia cuando hay imitación de esta, las suposiciones materialistas del experimento mental del primero difieren de la filosofía del segundo cuando se trata de qué causa la inteligencia. Mientras que en el contexto del pantano, si hay réplica de propiedades físicas del cerebro, entonces hay inteligencia, según Turing, basta la imitación del comportamiento de la máquina-cerebro para que haya estados mentales e inteligencia. No se requiere, por tanto, del sustrato biológico, y esto hace que para Turing ser inteligente sea algo flexible, alcanzable mediante programación de mecanismos simples que doten a las máquinas de la capacidad de aprender y educarse.

Un pasaje iluminador con respecto a cómo la inteligencia no es un producto físico, y por tanto puede ser programada, aprendida y enseñada es uno en que Turing se refiere al desarrollo futuro de la inteligencia de máquina. Justamente, esta debe tomar en consideración de qué forma aprende un niño, y es educado, para alcanzar la mente adulta. El pasaje en cuestión es el siguiente:

En el proceso de tratar de imitar a una mente humana estamos condicionados a pensar bastante acerca de los procesos de que fue sujeto para llegar a su estado actual. Podemos notar tres componentes:

1. el estado inicial de la mente en el nacimiento;
  2. la educación que ha recibido;
  3. otra experiencia, no describable como educación, que ha tenido.
- En vez de intentar producir un programa para simular la mente humana, por qué no tratamos más bien de simular uno que sea como la de un niño. Si se le sometiera a la educación apropiada, uno obtendría una mente adulta. Tal vez el cerebro del niño es algo así como un libro de notas que uno compra en una tienda de artículos de oficina. *Pequeños mecanismos y un montón de páginas en blanco.* (Turing, 1990, pp. 61-62, énfasis mío).

Tales mecanismos simples son, precisamente, programables para que trabajen el input de la experiencia, y se logre crear mente adulta inteligente mediante aprendizaje y educación. Nótese que este filósofo no se aboca a dar detalles de cómo debería aprender una máquina a tener la inteligencia de un niño, sino que propone que los mecanismos simples programables son suficientes para activar la cadena aprendizaje-educación-mente adulta. Pero, las preguntas que caben son las siguientes: ¿Puede la inteligencia del niño-robot ser suficiente para la acción inteligente en la ontología social? Más aún, ¿puede el niño-robot de Turing ser inteligente en vez de parecer serlo en dicho contexto?

### El problema de la inteligencia de máquina del niño-robot en la ontología social

Estimo que la filosofía de Searle (2010) constituye un sistema. Su concepción de la ontología social explica todos los hechos brutos, de la ciencia, más los institucionales. Aunque es un hecho bruto tener cerebro, este da lugar a todo hecho institucional, declarado



mediante lenguaje. Turing, tal como se examina en la sección precedente, está convencido de que la inteligencia es una función, algo programable en máquina de Turing, con independencia de lo biológico. Esto es imposible en la ontología social, como defenderé en el resto de esta sección, pese a la argumentación basada en el niño-robot.

En aras de la claridad y de la economía, es necesario tener presentes los siguientes elementos de la ontología social serleana: cerebro, conciencia, intencionalidad individual y colectiva, realidad física, hechos brutos, hechos institucionales, declaraciones, lenguaje, funciones y marcadores de estatus, reglas constitutivas y poderes deónticos. Brevemente se engarzan estos conceptos de la siguiente forma.

El cerebro, máquina biológica, tiene capacidad de sentir, gracias a la conciencia, y de representar, a causa de la intencionalidad. Tal máquina está inserta en un mundo físico, con el cual se conecta sintiendo. Pero, dada la complejidad de la inteligencia, el cerebro representa objetos y les da significado, los que aparecen en el ámbito de la ontología social: elecciones, dinero, matrimonios, premios, condenas, etc. Estos objetos descansan sobre propiedades físicas como el papel, las urnas, los anillos de matrimonio, los podios, las cárceles, etc. El hombre, como animal social, se asocia, coopera y participa junto con otros hombres: nace la intencionalidad colectiva; esta origina las atribuciones de funciones de estatus: X cuenta como Y en C. Hago que esta piedra X cuente como mojón frontera Y en el contexto de límites entre dos países C. Ciertas reglas se instauran; a saber, se emigra al traspasar la frontera del país propio para habitar territorio extranjero. A partir de esto se originan poderes deónticos: hay una caseta con policías que regulan la entrada y salida por el paso fronterizo. Visten uniformes, sus marcadores de estatus, asociados al reconocimiento de su autoridad. ¿Cómo? Solo puedo salir de la frontera presentando mi pasaporte válido. Pero, aquí viene la parte crucial: estoy consciente de las reglas que regulan la emigración. Para salir de Chile debo poseer un pasaporte.

No deambulo de un país a otro. Si emigro, conscientemente presento mi pasaporte y atravieso la frontera a, por ejemplo, Perú. Eso es lo esencial de la acción inteligente emigrar, de involucrarse y hacerse parte de la misma. En vez de eso, los computadores digitales solo siguen reglas formales al modo de un algoritmo que se implementa. Pero la pregunta que cabe es la siguiente: ¿Es la conciencia irrelevante en la ontología social? Por ejemplo, ¿emigraría estando inconsciente, en un estado de catalepsia? Por supuesto que no. Y este punto es central aquí: tampoco lo haría si me transformase en un robot o un zombi. ¿Por qué? El robot sigue las reglas de un algoritmo para implementarlo. Claramente, no hay un sentir que acompañe la ejecución del algoritmo: prueba de esto es que no hay nada que es ser-como-una-máquina-programada. Por otra parte, los zombis por definición también carecen de estados mentales conscientes cualitativos. Por lo tanto, ni zombis ni robots ni hombres con catalepsia sienten y actúan de acuerdo con las reglas constitutivas de "emigrar". Ese sentir implica estar consciente e involucrarse en las reglas mediante las cuales se participa en acciones de la ontología social.

Y, ¿qué hay entonces del niño-robot y su aprendizaje para convertirse en mente adulta? ¿Sería parte de dicha ontología? Si lo intentase sería una especie de zombi que, al tratar de actuar de acuerdo con reglas constitutivas, produciría acciones mecánicas, imitando la conducta de un hombre. Aún así, no sería un habitante genuino de la ontología social y aunque estaría en esta no participaría genuinamente en esta. Tal como hay robots que imitan, haciendo como si juegan fútbol, los niños-robots harían como si participasen en la ontología social con acciones inteligentes. Si caben dudas, considérese una orquesta en que se participa colectiva y conscientemente, ¿lo harían los niños-robots de Turing? De la misma forma, ¿podrían tales niños hacer promesas y comprometerse a tocar en su graduación escolar?

La teoría funcionalista de Turing, tan vanguardista como polémica, lleva a un menudo problema en el contexto como la on-



tología social. No es posible la acción inteligente en un contexto complejo sin un producto biológico como la conciencia, porque de otra forma se actúa como un individuo cataléptico, un zombi o un robot. Es importante finalizar esta sección advirtiendo que no somos zombis ni robots: el hecho de que estemos conscientes de las reglas que constituyen acciones inteligentes en la ontología social hace que seamos los seres más inteligentes del planeta Tierra. También los más peligrosos, por el carácter multipropósito de la inteligencia. Y esta es mucho más multipropósito en la ontología social, en donde se requiere de la conciencia para llevar a cabo la acción inteligente, o flexible, en términos cartesianos. ¿Cuán inteligente? Lo suficiente para navegar exitosamente en el complejo mundo social de funciones de estatus, instituciones y poderes deónticos. Una navegación que requiere más que la pura imitación, o el hacerse-conductualmente-pasar-por, que defiende Turing. Al contrario, requiere de genuina participación o involucrarse-concientemente-en, condición necesaria de la acción inteligente en la ontología social.

## Referencias

- Block, N. (1990). "The Computer Model of the Mind". En E.E. Smith y D.N. Osherson (Eds.) *Thinking: An Invitation to Cognitive Science* (pp. 247-89). Cambridge, Mass.: MIT Press.
- Clark, S. (1982). *The Nature of the Beast*. Oxford: OUP.
- Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.
- (2003). "The Turing Test". En J. H. Moor (Ed.) *The Turing Test: The Elusive Standard of Artificial Intelligence*, pp. 1-21. Dordrecht: Kluwer.
- Davidson, D. (1990). "Turing's test". En *Modelling the Mind*, pp. 1-11. Oxford: Clarendon Press.
- De La Mettrie, J. O. (1747). *Machine Man*. Reimpreso en S. Shieber (Ed.) *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, pp. 45-57. Cambridge, Mass.: MIT Press.
- Descartes, R. (1637). *Discurso del Método*. Reimpresión 1984, Madrid: Alianza Editorial.
- Genova, J. (1994). "Turing's sexual guessing game". *Social Epistemology*, 4, Vol. 8, pp. 313-26.
- González, R. (2011). "Descartes, las intuiciones modales y la inteligencia artificial clásica". *Revista Alpha*, 32, pp. 181-198.
- Hayes, P. y Ford, K. (1995). "Turing Test as considered harmful". *Proceedings of the International Joint Conference on Artificial Intelligence IJCAI 95*, Vol. 1, p. 972.
- Hodges, A. (1983). *Alan Turing: the Enigma*. London: Burnett.
- Lassègue, J. (1996). "What kind of Turing Test did Turing have in mind?" *Tekhnema*, 3, pp. 37-58.
- Moor, J. (1973). "Turing Test". En S. C. Shapiro (Ed.) *Encyclopedia of Artificial Intelligence* Vol. 2, pp. 1126-30. New York: John Wiley.
- Moor, J. (2004). "An Analysis of the Turing Test". En S. Shieber (Ed.) *The Turing Test: Verbal Behavior as the Hallmark of Intelligence* (pp. 297-306). Cambridge, Mass.: MIT Press.
- Naur, P. (1986). "Thinking and Turing's Test". *BIT*, 3, pp. 175-187.
- Putnam, H. (1973). "The nature of mental states". Reimpreso en D. Chalmers (Ed.) *Philosophy of Mind: Classical and Contemporary Readings*, pp. 73-79. Oxford: OUP.
- Saygin, A., Cicekli, I. y Akman, A. (2003). "Turing Test: 50 years later". En J. H. Moor (Ed.) *The Turing Test: The Elusive Standard of Artificial Intelligence*, pp. 26-78. London: Kluwer Academic Publishers.
- Shieber, S. (2004). "The Ephemera". En *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, pp. 97-104. Cambridge, Mass.: MIT Press.
- Swade, D. (2000). *The Difference Engine: Charles Babbage and the Quest to build the First Computer*. London: Penguin.
- Traiger, S. (2003). "Making the right identification in the test". En J. H. Moor *The Turing Test: The Elusive Standard of Artificial Intelligence*, pp. 99-110. Dordrecht: Kluwer.
- Turing, A. (1936). "On computable numbers, with an application to the Entscheidungsproblem". En *Proceedings of the London Mathematical Society*, 2, pp. 230-265.



*Society*, series 2, Vol. 42, pp. 231-265 (con correcciones en Vol. 43, pp. 544-546).

— (1950). "Computing intelligence and machinery". Reimpreso en M. Boden (Ed.) *The Philosophy of Artificial Intelligence*, pp. 40-66. Oxford: OUP.

## CAPÍTULO VI

LAS INTENCIONES COMO COMPROMISOS: UN ANÁLISIS DE LAS  
POSTURAS DE LA DOBLE PERSPECTIVA SOBRE LA AGENCIA HUMANA

*María Ayelén Sánchez*

### **La teoría estándar de la acción: una concepción causalista de la intención**

Entre la infinidad de eventos que ocurren en el mundo, algunos involucran movimientos corporales (como cantar una canción, abrir una ventana o estornudar). Algunos de los movimientos corporales son considerados cosas que le ocurren a la persona (temblar o ruborizarse, por ejemplo) pero otros son entendidos como cosas que la persona hace ( patear una pelota o servirse un café). Solo a estos últimos los denominamos acciones. El intento por comprender y explicar esta última subclase de eventos motivó el surgimiento de una rama de la filosofía que se ocupa específicamente de reflexionar sobre las acciones. Así nació la filosofía de la acción en la segunda mitad del siglo XX con el objeto de teorizar sobre las acciones intencionales humanas, es decir, aquellos comportamientos motivados por creencias, deseos e intenciones y explicables en estos términos. Esta rama de la filosofía se interesa por las intenciones en tanto que la relación entre la intención y la acción le permite delimitar el campo de las acciones intencionales (Dancy y Sandis, 2015, p. 10).

En la filosofía de la acción se pueden identificar un conjunto de tesis que conforman lo que se denomina "teoría estándar de la acción" (*standard theory of action*), cuyas bases remiten a la obra de Davidson (1963/1980). Sus seguidores han continuado con el espíritu general de la propuesta davidsoniana, ofreciendo