

EL TEST DE TURING: DOS MITOS, UN DOGMA¹

Rodrigo González
Katholieke Universiteit Leuven.
rodgonfer@gmail.com

Resumen

Este artículo analiza el Test de Turing, uno de los métodos más famosos y controvertidos para evaluar la existencia de vida mental en la Filosofía de la Mente, revelando dos mitos filosóficos comúnmente aceptados y criticando su dogma. En primer lugar, se muestra por qué Turing nunca propuso una definición de inteligencia. En segundo lugar, se refuta que el Test de Turing involucre condiciones necesarias o suficientes para la inteligencia. En tercer lugar, teniendo presente el objetivo y el tipo de evidencia que recopila, se considera si el Test de Turing cuenta como un experimento científico a la luz de la concepción de Fodor. Finalmente, se argumenta que Turing simpatiza con una forma de Conductismo, confundiendo la simulación —un proceso epistémico que, gobernado por la verosimilitud, es eficaz cuando alguien es causado a creer que el computador es inteligente— con la duplicación de la inteligencia en cuanto propiedad, lo que ocurre a nivel ontológico. Tal confusión implica un dogma y explica por qué, a pesar de haber sido propuesto como una solución final a la problemática de si las máquinas programadas piensan, el Test de Turing ha tenido precisamente el efecto contrario en más de cinco décadas, estimulando el debate filosófico torno a la naturaleza de lo mental.

PALABRAS CLAVE: test, inteligencia, simulación, duplicación, verosimilitud.

Abstract

Debunking two commonly held myths and fleshing out its dogma, this article deals with the Turing Test, one of the most famous and controversial methods to assess the existence of mental life in the Philosophy of Mind. Firstly, I show why Turing never gave a definition of intelligence. Secondly, I dispute claims that the Turing Test provides a necessary or sufficient condition of intelligence. Thirdly, in view of its aim and the sort of evidence it offers, I consider whether or not Turing's test can be regarded as a scientific experiment in light of Fodor's theory. Finally, I argue that Turing is committed to a form of behaviourism and, further, confuses simulation —an epistemic process which, being governed by verisimilitude, is successful when someone is caused to believe that the computer is intelligent— with the duplication of intelligence qua property, which takes place at an ontological level. This confusion involves a dogma and explains why, despite being devised

¹ Agradezco al revisor anónimo de la *Revista de Filosofía* por sus valiosos comentarios y sugerencias, y muy especialmente a Carolina Páez por los esquemas.

as the final solution to the dilemma of whether or not programmed machines think, the Turing Test has precisely had the opposite effect for longer than five decades, stimulating the philosophical discussion on the nature of mind.

KEYWORDS: *test, intelligence, simulation, duplication, verisimilitude.*

No simulation by itself ever constitutes duplication

John Searle

INTRODUCCIÓN

RI La historia de la Inteligencia Artificial no ha estado ajena a las frecuentes disputas y embates propios de una disciplina emergente. Es posible afirmar que lo que hoy se conoce como Inteligencia Artificial, entendido como un paradigma de investigación, comenzó a consolidarse solo a partir de la segunda mitad del siglo XX, puesto que con anterioridad solo se denominaba ‘Inteligencia de Máquina’ [*machine intelligence*] (Copeland 2003). Ciertamente, muchos de los desacuerdos de ese entonces, algunos de los cuales incluso perseveran hasta hoy, se originaron a partir de la falta de consenso acerca de qué significaba que una máquina programada pensara. Términos como ‘inteligencia’, ‘máquina’, ‘pensamiento’, etc., generaban polémica, a pesar de que existía una intuición compartida de a qué podrían referirse. Sin duda, lo anterior indica la existencia de un problema de fondo: cómo determinar si una máquina piensa y actúa de manera inteligente, uno de los tópicos más debatidos en la Filosofía Contemporánea. Curiosamente, dicha complicación ya había sido vislumbrada en el siglo XIX por Charles Babbage, en el albor de la Inteligencia de Máquina. Un claro signo de esto son las excusas que el matemático británico ofrece por el uso de lenguaje mental para describir la potencial conducta de una de sus creaciones. Swade comenta las complicaciones teóricas enfrentadas por Babbage de la siguiente forma:

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 (Swade 2000, pp. 103-104, énfasis mío).

Estas sensatas reflexiones surgieron en el seno del diseño de la máquina analítica, que muchos estiman fue el primer proyecto de máquina programada o computador. No obstante, la historia ha juzgado que fue Augusta Ada, la condesa de *Lovelace* e hija de Lord Byron, quien descubrió su verdadero potencial, ya que este artefacto, que podía ser programado mediante tarjetas podía servir para *diversos usos*, lo que

obviamente incluía los complicados cálculos matemáticos que motivaron la parcial construcción de su predecesora, la máquina diferencial². Todo ello ocurrió en el siglo XIX, cuando lo que hoy se conoce como Inteligencia Artificial estaba en su alborada y en una etapa francamente precientífica.

Pero, ¿cómo se explica el grado creciente de consenso y unificación después de la primera mitad del siglo XX? Existen una serie de acontecimientos que muestran avances entre las investigaciones conducidas después de la década del cuarenta. Diversos computadores son construidos casi simultáneamente en varios países, entre los cuales destacan el Z-3 de Konrad Zuse en Alemania, el Colossus en Inglaterra y el ENIAC en EE.UU. (Copeland 1993). Por otra parte, otros síntomas del consenso alcanzado son el planteamiento por parte de Newell y Simon de un programa computacional, el Lógico Teórico, capaz de probar treinta y ocho de los cincuenta y dos teoremas de *Principia Mathematica*, y la conferencia *Dartmouth Research Project on Artificial Intelligence*, organizada por John McCarthy, cuya importancia es vital, pues cambió el nombre de la disciplina de Inteligencia de Máquina a Inteligencia Artificial.

Indiscutiblemente, Alan Turing fue uno de los mayores contribuyentes al consenso teórico post 1950, a tal punto, que muchos lo consideran el padre de la Inteligencia Artificial. Este matemático-filósofo aportó dos trabajos seminales, la máquina de Turing y el Test de Turing (Turing 1936 y 1950, respectivamente), que pueden considerarse hitos fundacionales de la Ingeniería Informática. En efecto, el primero define el carácter algorítmico de ‘computación’, cuya realización material es completamente irrelevante, mientras que el segundo precisa de manera sistemática un método para acopiar evidencia de vida mental inteligente en computadores programados.

Este artículo analiza el Test de Turing, mostrando cómo ha sido mal interpretado por numerosos filósofos, quienes han instaurado dos mitos en torno a su naturaleza. Asimismo, investiga cómo el método planteado por Turing ha ayudado a consolidar un dogma filosófico en relación con la naturaleza de la mente. El primer mito es que el test ofrece una definición de inteligencia, cuestión que claramente no se condice con la actitud de Turing de *evitar* discusiones, tanto del significado de ‘máquina’ e ‘inteligencia’, como de su uso común. El segundo mito afirma que el test ofrece condiciones necesarias y suficientes para la existencia de inteligencia, en consideración a que solo plantea una forma de obtener evidencia de que una máquina programada *es* inteligente. Y aquí justamente radica el dogma del Test: éste confunde lo epistémico, esto es, el éxito del proceso causal en que observadores son convencidos de que existe mentalidad, con lo ontológico, es decir, la existencia de vida mental inteligente en una máquina programada. Naturalmente, tal confusión explica por qué el Test de Turing representa una controvertida respuesta al problema de las otras mentes, y por qué dicho test ha enfrentado tantas refutaciones en casi sesenta años.

² El hecho de que Babbage no se diese cuenta del carácter multipropósito de la máquina analítica sugiere, pero no asegura, que la tradición está en lo cierto.

El artículo está dividido en cuatro secciones. La primera discute cómo varios filósofos contemporáneos sostienen que el test define qué es la inteligencia, a pesar de la máxima de Turing de eludir cualquier clase de discusión conceptual en este sentido. En la segunda sección se explica por qué el test no se centra en el planteamiento de condiciones necesarias y suficientes para la existencia de inteligencia, sino que acentúa la postulación de un método que, mediante la obtención de evidencia estadística, indique que un computador posee mente e inteligencia. La tercera sección examina en detalle qué es, finalmente, el Test de Turing, y si dada su naturaleza satisface las condiciones de los experimentos científicos, tal como los concibe Fodor. Finalmente, la cuarta sección elucida cómo el Test de Turing ha apoyado un dogma al que se aferran numerosos defensores de la Inteligencia Artificial tradicional, a saber, que la conducta observada y obtenida a partir de simulación es suficiente para concluir la duplicación de mentalidad e inteligencia.

1. Derribando el mito N°1: por qué el Test de Turing no define 'inteligencia'

Mucho se ha discutido acerca de cómo el Test de Turing ofrece una definición de inteligencia. Tal vez una buena manera de comenzar a analizar este problema es tener como antecedente el objetivo de dicho test, que Turing manifiesta abiertamente. Según el matemático-filósofo, una buena forma de apartarse de la espinuda pregunta sobre si los computadores piensan es a través de rechazar el análisis del uso común de términos como 'piensa' y 'máquina', ya que dicho examen podría llevar a una encuesta sobre definiciones y usos. Turing manifiesta explícitamente la inconveniencia de debatir acerca del uso común de 'máquina' y 'piensa' de la siguiente manera:

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).

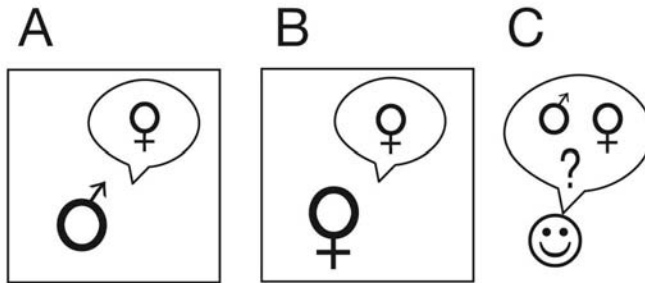
La desazón ante tal posibilidad motiva a Turing a plantear un test, entendido precisamente como un reemplazo de la mencionada pregunta. Turing, entonces, formula un método, el denominado *Juego de la Imitación* [*Imitation Game*], que consiste en que ciertas personas juegan a descubrir la identidad de uno de ellos mediante un breve interrogatorio. Es importante destacar que las bases de tal juego fueron ideadas antes del Test de Turing, cuando el matemático-filósofo propone un experimento en el cual hay tres jugadores de ajedrez y uno de ellos debe adivinar con quién está jugando (Turing 1948, p. 23).

Turing planteó diversas versiones del *Juego de la Imitación*. En la original, tres personas participan en el juego: un hombre, una mujer y una persona cuyo sexo no es relevante. Tanto A (hombre) como B (mujer), quienes se encuentran en dos piezas distintas, responden a las preguntas formuladas por C ('juez' de sexo no relevante),

quien debe identificar el sexo de los otros participantes desde el exterior. Mientras que la mujer responde a las preguntas de manera veraz, el hombre debe hacerse pasar por una mujer, de modo que C crea que A también es de sexo femenino.

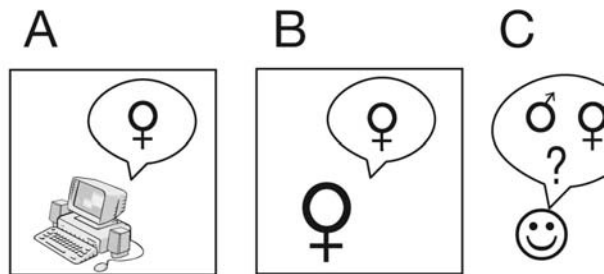
Para evitar que se descubra de manera obvia quién es quién, las preguntas y respuestas son escritas e impresas a través de tele-impresoras. La figura 1 esquematiza la versión original del *Juego de la Imitación*:

Fig. 1



Si A es lo suficientemente convincente, C concluirá que hay dos mujeres en vez de una, lo cual revela el peculiar papel que A desempeña, a saber, debe causar que C se equivoque en la identificación. A juicio de Turing, lo anterior indica cómo el juego puede reemplazar la pregunta acerca de si las máquinas piensan. Si un computador reemplaza a A, y ello no involucra ninguna diferencia respecto de cómo C identifica erróneamente a los participantes, entonces: portamiento inteligente, puesto que puede

La segunda versión del juego consis-
plaza a A, haciendo creer a C que es un
verazmente, tal como ilustra la figura 2:



Si el computador responde en forma eficiente a las preguntas que se le formulan, el resultado más probable es que C se equivoque y concluya que hay dos mujeres en vez de una. Una cuestión que adicionalmente ha sido debatida y que ha generado variadas hipótesis es por qué Turing introdujo el peliagudo tema del género en el *Juego de la Imitación*. Una de estas hipótesis es que él podría haber considerado que

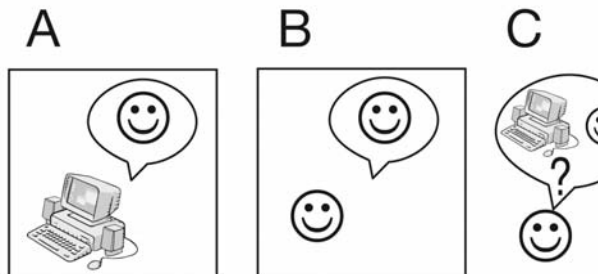
pasar por mujer era relativamente más fácil que pasar por hombre. Sin embargo, tal conjetura solo lleva a especulaciones infundadas e irrelevantes con respecto a la naturaleza del test. Ignorando las posibles ambigüedades y disquisiciones acerca de las razones que tuvo el matemático-filósofo para centrarse en el género de los participantes, la tradición ha propuesto la siguiente versión como la interpretación estándar del Test de Turing (Fig. 3):

Fig. 3

Esta versión revela de manera indiscutible qué está detrás del *Juego de la Imitación*. El hecho de que la finalidad del juego sea inducir a C a efectuar identificaciones erróneas sugiere que el Test de Turing intenta verificar la existencia de inteligencia en base a la imitación y la simulación eficaz. Este objetivo del test, enfatizado por Saygin et al. (2003, p. 26), es crucial, tal como se examinó en otra oportunidad (González y Vergauwen 2005), pues ha influenciado de manera importante al concepto de simulación en Inteligencia Artificial.

A pesar de que Turing manifiesta su pregunta acerca de si las máquinas piensan, el test propone una definición de inteligencia. El biógrafo Hodges, afirman que Turing ofrece una definición particular, Block (1991, p. 249) sostiene que la definición de inteligencia en términos no mentales no puede plantear una definición conductista. Por otro lado, los críticos del Test de Turing, asegura de manera tajante que éste fue originalmente propuesto como una definición operacional simple de inteligencia. Finalmente, Hodges (1992, p. 415) también apoya este mito cuando afirma que el matemático-filósofo introdujo una definición operacional de ‘piensa’ o ‘inteligente’ a través del *Juego de la Imitación*.

Cabe preguntarse, entonces, si Turing quiso postular una definición operacional de inteligencia, de modo de restringirla a un ámbito no mental. A pesar de que existe una obvia conexión entre la conducta lingüística y la evaluación de la inteligencia según Turing, ello no necesariamente significa que haya pensado su método como una manera de constreñir el significado de ‘inteligencia’, pues el acento está puesto en la evidencia obtenida. Tanto Moor (1976 y 1987) como Copeland (2003) apoyan esta clarificación. Es más, el último transcribe de manera íntegra una entrevista radial a Alan Turing difundida por BBC en 1952, que refuta *el mito de la definición* (Copeland 2003,



p. 6). En una parte de la entrevista, Turing se refiere a este punto de la siguiente manera:

No quiero dar una definición de qué es pensar, pero si tuviese que darla, probablemente sería incapaz de expresar nada más acerca de ésta que decir que fue un tipo de zumbido mental [*buzzing*] en mi cabeza. Pero no veo que tengamos que estar de acuerdo en una definición en modo alguno. Lo relevante es tratar de distinguir entre las propiedades de un cerebro o las de un hombre, que queremos discutir, y aquellas que no queremos. Para ponernos en un caso extremo, no estamos interesados en el hecho de que el cerebro tenga la consistencia de la papilla. No queremos decir ‘esta máquina es muy compleja, luego no es un cerebro y no puede pensar’. Me gustaría sugerir una clase particular de test que uno pudiese aplicarle a una máquina. Ud. podría querer llamar a este un test para ver si la máquina piensa, pero sería mucho mejor no formularlo así y caer en la petición de principio, diciendo que las máquinas que pasen el test serán, por decirlo de algún modo, máquinas grado A. La idea del test es que la máquina tiene que simular ser un humano a través de responder a las preguntas que se le hacen, y sólo pasará este test si la simulación es suficientemente convincente.

(Turing 1952, énfasis mío).

Este largo pasaje zanja definitivamente por qué el mito N° 1, esto es, que el Test de Turing define qué es la inteligencia, es falso.

2. *Derribando el mito N°2: las mentadas condiciones necesarias y suficientes*

Otro mito al que adhieren algunos filósofos es que el test plantea condiciones necesarias o suficientes para la inteligencia, lo que de alguna forma de nuevo se conecta con entender su método como una definición. Block es justamente uno de los filósofos que cree en este mito. En particular, además de aseverar que Turing propuso una definición conductista de inteligencia, sostiene en una nota lo siguiente:

Turing reaccionó a lo último [un experto puede averiguar que las máquinas resuelven problemas que la gente no puede resolver] negando que pasar el test involucra una condición necesaria para la inteligencia, y debilitando su posición a: pasar el Test de Turing es una condición suficiente para la inteligencia (Block 1990, nota a pie en pp. 249-250).

Sin embargo, Turing nunca manifiesta que su test involucra condiciones necesarias o suficientes para la inteligencia. Por el contrario, y tal como se ha enfatizado anteriormente, él considera que pasar su test produce evidencia de que el computador es inteligente.

En este sentido, Turing mismo argumenta que lo central es que una máquina inteligente sea capaz de suplantar a un humano en el *Juego de la Imitación*. Una posible objeción es que el juego es demasiado estricto con la máquina, porque si un

hombre tuviese que reemplazarla, sin duda tendría un desempeño muy pobre en matemáticas. Por otra parte, ¿que pasaría si ello no fuese posible porque la máquina hace algo cuya naturaleza es totalmente distinta de aquello que hace un hombre cuando piensa y actúa con inteligencia? Turing considera ambos problemas, y en especial la última objeción, pero los desestima³ de la siguiente forma:

¿Podría acaso la máquina hacer algo que debiese ser descrito como pensando pero que es muy diferente a lo que hace un ser humano? Esta objeción es crucial, pero podemos decir que sí, no obstante, puede construirse una máquina para jugar el *Juego de la Imitación* de manera satisfactoria, no tenemos que preocuparnos de ella (Turing 1950, p. 42).

Lo relevante, entonces, es que los humanos sean convencidos mediante el método propuesto por el *Juego de la Imitación*. Turing, en este sentido, enfatiza que los jueces (C), en tanto evaluadores, deben ser presa de la simulación eficiente y, por lo mismo, no pueden ser expertos en Ingeniería Informática o en alguna ciencia relacionada con el estudio de la cognición. Por el contrario, Turing recalca que los jueces deben ser gente común o legos, es decir, incapaces de descubrir a la máquina mediante cualquier tipo de tecnicismos o argucias⁴. Por lo tanto, los interrogadores deben ser adecuados para la simulación, aunque ello tampoco implica que deban ser candorosos. En consecuencia, se requieren interrogadores *promedio*, que no sean ni expertos ni demasiado crédulos.

Todo lo anterior sugiere que el acento del Test de Turing está en el convencimiento de los jueces y que esto involucra, en vez del planteamiento explícito de condiciones necesarias o suficientes para la inteligencia, la obtención de evidencia para justificar que un computador es inteligente. En relación con este problema, un aspecto interesante del test y que también ha generado mucho debate es la centralidad que tienen los interrogadores humanos. Por ejemplo, ¿podrían ser dichos interrogadores reemplazados por un programa, tal como la máquina programada reemplaza a A o B? Todo parece indicar que si no existe *en principio* un algoritmo que reemplace a los

³ Ciertamente es justo decir que Turing desestima la diferencia entre lo que explica la conducta de la máquina y la del humano de una manera extraordinariamente liviana. De hecho, el experimento mental de *La Pieza China* de Searle justamente explota esta diferencia (Searle 1980 y 1990), replicando las mismísimas condiciones del Test de Turing para refutar la tesis de que la mente no es más que un operador algorítmico.

⁴ Los expertos saben qué tipo de cosas preguntar a un computador. Por ejemplo, existen oraciones ambiguas como: i) 'Los detectives detuvieron a los sospechosos porque cumplieron con su deber'. ii) 'Los detectives detuvieron a los sospechosos porque habían estado en la escena del crimen'. Tanto en la primera como en la segunda oración es más o menos obvio percatarse a qué se refiere el segundo verbo. Sin embargo, para un computador no es posible clarificar qué interpretaciones caben para el significado de cada oración, por tanto es altamente probable que se equivoque al contestar preguntas como: ¿por qué es improbable, aunque no imposible, que *habían estado* refiera a los sospechosos en la segunda oración? Hay muchos otros casos como, por ejemplo, entender chistes basados en diferencias culturales, los cuales un computador no entiende, pese a que simula hacerlo.

jueces y a su *criterio*, entonces la inteligencia humana resulta irreducible desde un punto de vista computacional, ya que el humano parece ser un estándar irremplazable. De acuerdo al test planteado por Turing, un programa puede pasar por un humano y convencer a interrogadores, pero no puede reemplazar a los interrogadores y al criterio que estos poseen para evaluar qué es inteligente. A pesar de este sesgo antropocéntrico, Turing propone su método como una manera de acopiar evidencia estadística o inductiva de que la máquina programada piensa.

3. *El Test de Turing como método empírico-estadístico: ¿un experimento científico?*

Dado que los mitos 1 y 2 del Test de Turing han sido derribados, es posible ahora ahondar en la exploración de su naturaleza. Debido a que Turing destaca el convencimiento de interrogadores y, por tanto, la evidencia en tanto generación causal de creencia, cabe preguntarse si dicho test es un experimento científico, tal como Fodor entiende este.

Su posición, examinada críticamente antes (González 1997 y 2002), puede sintetizarse de la siguiente manera. Fodor (1990 y 1994) entiende que los experimentos son instancias en que alguien es causado a creer que p si y solo si p es verdadero —y así, si una proposición corresponde con un hecho—, a diferencia de cómo los concibe la tradición, es decir, como lugares para la observación y corroboración (o refutación) de teorías. En este sentido, realizar un experimento es formular una pregunta a la naturaleza, pero más importante aún, es predisponerse mentalmente a ser causado a creer que p . Por ejemplo, uno predispone su mente a creer que p cuando pregunta la hora a alguien que tiene reloj, o cuando ve el informe meteorológico. Sin embargo, tal como en los experimentos, hay condiciones *ceteris paribus* que deben satisfacerse, ya que nadie sería causado a creer la hora o el informe meteorológico en un pabellón psiquiátrico de mitómanos o en el año 732, en medio de la Batalla de Poitiers. Los experimentos tienen que ser respaldados por enunciados contrafácticos y satisfacer otras condiciones, detalles que no vale la pena examinar aquí. Más relevante es que, dada la concepción fodoriana de los experimentos, existe un estrecho vínculo entre la verdad de una creencia y la posibilidad de que ésta sea producida causalmente, ya que un experimento no causa la creencia que p a menos que p sea el caso. Es por ello que los científicos usan los experimentos como artefactos cognitivos capaces de convencer al resto de la comunidad científica, o al *club*, de la solidez de una hipótesis. ¿Se da esta situación en el Test de Turing, esto es, puede contar como un experimento científico donde alguien es causado a creer que p si y solo si p es verdadero?

El Test de Turing no cumple esta condición, a pesar del énfasis que otorga al convencimiento de jueces y al método que, dependiendo de un proceso causal de generación de creencias, tiene como objetivo la obtención de evidencia inductiva para corroborar la hipótesis de que el computador piensa. No obstante, a diferencia de un experimento científico cualquiera y a que depende fundamentalmente de la simulación,

el Test de Turing induce una creencia falsa, a saber, que el computador es una persona. Aun así, Turing considera que la simulación exitosa y verosímil es aquello que apoya la hipótesis de que la máquina es inteligente y posee vida mental.

La dependencia en la imitación, la simulación y la producción causal de una creencia falsa en los interrogadores es aquello que no permite clasificar al Test de Turing como un experimento científico adecuado, pese a su acento en la obtención de evidencia para la hipótesis de que el computador piensa. Cabe destacar que Turing, incluso, formuló una interesante predicción:

Creo que en alrededor de cincuenta años será posible programar un computador, con una capacidad de memoria de 10^9 , para que participe en el *Juego de la Imitación* tan eficientemente que un interrogador lego no tendrá más de un 70% de probabilidad de hacer la identificación correcta después de cinco minutos de interrogatorio. Creo que la pregunta original, ¿pueden las máquinas pensar?, es demasiado absurda para seguir analizándola. No obstante, pienso que a finales de este siglo las ideas de la gente ilustrada y el uso de las palabras habrán cambiado de manera tal que uno será capaz de hablar de *máquinas que piensan* sin incurrir en contradicción alguna

(Turing 1950, p. 49, énfasis mío).

Lo anterior claramente permite concluir que el Test de Turing describe un método para la obtención de evidencia inductiva en apoyo a la hipótesis que afirma que computadores y máquinas programadas poseen vida mental. Dicho método, empero, no cumple con un importante requisito que parecen satisfacer la mayoría de los experimentos científicos, esto es, la generación causal de creencias verdaderas⁵. Por el contrario, el Test de Turing se basa en convencer a observadores mediante la simulación. Ello hace que el método propuesto por el matemático-filósofo haya respondido a una buena cantidad de objeciones iniciales, previas a 1950, a la idea de que las máquinas programadas piensan.

Sus defensores, por otra parte, han intentado salvar objeciones posteriores. Moor, por ejemplo, sostiene que el Test de Turing representa un objetivo a largo plazo para la Inteligencia Artificial (Moor 1987), cuestión con la que Copeland (2003, pp. 8-9) coincide cuando afirma que los cincuenta años a los que se refirió Turing no fueron *exactamente* cincuenta años, argumento que es esgrimido teniendo en consideración la competencia *Loebner Turing* organizada en 2000, cuyos resultados también son controvertidos.

⁵ Un aspecto que no abordo aquí es cómo resulta al menos cuestionable que un experimento genere eficazmente creencias si y solo si son verdaderas, una idea que es central en la concepción fodoriana. En la historia de la ciencia existe un sinnúmero de casos en que experimentos han generado causalmente creencias consideradas verdaderas en un momento, pero que han resultado refutadas con posterioridad (González 1997 y 2002). Por ejemplo, es sabido que los experimentos de Robert Boyle influyeron de manera importante la teoría de Becker y Stahl acerca del flogisto. Sin embargo, lo central, para propósitos de la presente discusión, es que el Test de Turing, mediante la simulación, induce una creencia no necesariamente verdadera.

En todo caso, y a pesar de que la predicción de Turing es relativamente modesta, pues solo tres de diez interrogadores deben ser convencidos por la máquina, resulta francamente dudoso que dicho convencimiento cuente como evidencia incuestionable de la inteligencia de ésta. Por otra parte, lo que resulta filosóficamente interesante es que el Test de Turing ha tenido, paradójicamente, el resultado inverso al previsto por su inventor: en vez de terminar con las discusiones conceptuales acerca de qué son la mente y la inteligencia, ha sumido a los filósofos en un debate acerca de la naturaleza de ambas en casi seis décadas y aún no se aprecian signos de consenso.

4. El dogma del Test de Turing: epistemología versus ontología

Una cuestión sospechosa en el Test de Turing es por qué la simulación efectiva o verosímil de conducta lingüística resulta no solo adecuada para convencer a los jueces interrogadores, sino además capaz de aportar evidencia de que las máquinas programadas piensan y son inteligentes. Todo parece indicar que Turing funde dos cosas en una cuando afirma que ser convencido por el computador permite justificar que tal máquina *es*, como cuestión de hecho, inteligente. Dos aspectos básicos y fundamentales acerca del problema de las otras mentes se mezclan aquí. Por una parte, una cosa es *saber* acerca de otras mentes, lo que es una cuestión epistemológica. Por otra parte, una cosa muy diferente es la existencia de otras mentes y, más importante todavía, de propiedades como la inteligencia, cuya existencia no parece depender de su corroboración por parte de observadores externos o de una actitud interpretativa que avale la existencia de estados mentales (contrariamente a lo que cree Dennett 1987).

En efecto, Turing confunde lo puramente epistemológico con la existencia de una propiedad, es decir, mezcla la generación causal de la creencia que *p*—originada a partir de las respuestas de la máquina, *similares* a las de una persona— con la existencia de inteligencia, cuestión que claramente no se sigue. Porque el matrimonio entre lo epistémico y lo ontológico nunca ha sido feliz. Siempre se tiende a reducir un ámbito al otro, y esto es precisamente lo que Turing hace cuando propone un método para corroborar la existencia de una propiedad, en este caso, la inteligencia, en función del convencimiento de observadores de que un computador es inteligente.

A pesar de que Block evidentemente yerra el tiro cuando sostiene que el Test de Turing propone una definición y que postula una condición suficiente para la inteligencia, acierta cuando asegura que existe un claro componente conductista en el método propuesto por Turing⁶, cuestión que ha sido criticada por diversos filósofos,

⁶ El Conductismo Metodológico es una escuela de Psicología, según la cual esta ciencia, de índole experimental, debe centrarse en el estudio y la predicción de la conducta. El Conductismo Lógico, en cambio, es una escuela filosófica que postula que los estados mentales deben reducirse a disposiciones conductuales.

tales como el mismo Block, Searle (2004) y este autor (González y Vergauwen 2005). Sin duda lo que Turing tiene en mente con esta maniobra es reducir la pregunta acerca de si las máquinas piensan, a un método de verificación que corrobore disposiciones conductuales, una solución comúnmente adoptada como respuesta al problema de las otras mentes, aunque extraordinariamente simplista y ciertamente poco satisfactoria.

La conducta, que es públicamente observable y que por su carácter manifiesto muchas veces resulta útil para zanjar discusiones conceptuales de carácter metafísico, no es igualmente eficaz en el caso de la mente, toda vez que ésta posee un punto de vista de primera persona que difícilmente se refleja de manera íntegra en las disposiciones conductuales.

Toda clase de experimentos mentales han sido propuestos para desbaratar a la conducta como signo inequívoco de inteligencia y mentalidad, los que van desde comunidades de super-espartanos o super-estoicos (Putnam 1965, pp. 102-103), quienes no manifiestan conducta asociada a dolor bajo ninguna circunstancia, hasta *La Pieza China* de Searle, en donde él manipula símbolos de acuerdo a reglas, lo que induce a creer que es un hablante nativo de chino, aunque solo es capaz de hablar inglés⁷.

Si bien parece evidente que el Test de Turing coquetea con el Conductismo Lógico, muchos de sus detractores soslayan que el problema filosófico principal no solo se origina a partir de la insuficiencia de la conducta lingüística para corroborar la existencia de mentalidad, sino que la simulación, en tanto fuente de evidencia estadística y elemento central del test, no permite aseverar de manera categórica que una máquina programada piensa y es inteligente. Este es precisamente el dogma del Test de Turing y aquello que lo ha convertido en uno de los problemas más controvertidos de la Filosofía de la Mente.

Pero, ¿es posible postular la existencia de estados mentales en un computador capaz de suplantar a un humano y por tanto apto para simular exitosamente vida mental? Si alguien se manifestase a favor de tal posibilidad, sin duda confundiría la simulación de la inteligencia o la generación de un proceso epistémico eficaz por el cual los interrogadores son convencidos de que el computador es inteligente, con la duplicación de la inteligencia. Y esto es precisamente lo que confunde Turing con su test y con el sesgo conductista que posee. Un ejemplo de esta confusión y de su

⁷ Searle concibe el experimento mental de *La Pieza China* teniendo en cuenta no solo el Conductismo sino también al Funcionalismo de Máquina de Turing, según el cual los estados mentales se explican por y se reducen al funcionamiento de un programa y, por tanto, no requieren de realización material. Para propósito de la presente discusión, que se centra en el dogma del Test de Turing, no es necesario ahondar en dichos detalles. Solo cabe mencionar que el *Gedankenexperiment* de Searle ha generado tantos detractores como partidarios, provocando un debate filosófico que se ha extendido por más de veinticinco años. Resulta, entonces, tan polémico como el test planteado por Turing (para ponderar algunas de las posiciones en torno a *La Pieza China*, véase Preston & Bishop 2002).

simpatía por el Conductismo Lógico puede encontrarse en el siguiente extracto de otra entrevista radial a Turing para BBC en 1951:

Para lograr que nuestro computador imite a una máquina sólo 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* (En Copeland 2003, p. 11, énfasis mío).

Frente al argumento de Turing, cabe señalar que la simulación o imitación eficaz no implica duplicación de propiedades, y es absolutamente cuestionable su afirmación de que algo que imita a un cerebro resulte ser un cerebro. Si las simulaciones fuesen la cosa real, uno saldría disparado con las simulaciones meteorológicas de huracanes, y terminaría aporreado o, peor aun, tendría un fin trágico cada vez que colisiona en un simulador de vuelo. Justamente, uno de los aspectos más interesantes de la simulación es la peculiar relación que existe entre ésta, lo imaginario, lo concebible y lo real.

El éxito de los procesos de simulación está en función de su verosimilitud, cuestión que a su vez depende de que la imitación de propiedades se aleje mínimamente de ciertas propiedades del mundo real. Tal proceso está, en efecto, comandado por el principio de alejamiento mínimo [*Principle of Minimal Departure*], que Ryan postula en el contexto de la ficción para elucidar cómo ésta resulta suficientemente creíble (Ryan 1991, pp. 50-51). La simulación también depende de que la imitación se apege a las propiedades del mundo real, pues solo de esta manera resultará exitosa y, en consecuencia, convincente. No obstante, el éxito de una simulación no implica que se genere la cosa real. Ejemplos de esto sobran en la Inteligencia Artificial, tales como PARRY, un programa escrito por el psiquiatra Kenneth Colby en la Universidad de Stanford en 1972 que simula a un paciente paranoico (Heiser, Colby et ál. 1980), o su contrapartida ELIZA (Weizenbaum 1984), que simula a una terapeuta *rogeriana* interrogando pacientes.

Desde el punto de vista del sentido común, es más o menos evidente que todas estas simulaciones no son la cosa misma y que lo simulado no tiene el mismo estatus que lo real. Como bien dice Searle, aunque en la vida real nadie está dispuesto a aceptar algo tan absurdo, cuando se trata de la mente, muchos asumen livianamente que la inteligencia es una propiedad abstracta y que por ello la simulación es suficiente para su replicación. ¿Dónde se originó esta concepción de la inteligencia como algo abstracto, como una propiedad que puede generarse mediante cálculo y, así, cuya simulación basta para su replicación? La Historia de la Filosofía está llena de héroes y

villanos, todos erigidos por distintas escuelas, épocas, y teorías dominantes. Es posible afirmar que Turing, pese a que ayudó a consolidar el nacimiento de la Inteligencia Artificial y el paradigma dominante de la Ingeniería Informática, legó una visión de la inteligencia y de la mente profundamente antibiológica. Tal prejuicio puede apreciarse en el siguiente pasaje del Test de Turing, en donde se refiere a las ventajas del *Juego de la Imitación* y a cómo *no* se debe producir inteligencia artificialmente:

El nuevo problema tiene *la ventaja de distinguir claramente entre las propiedades físicas y las propiedades intelectuales del ser humano*. Ningún ingeniero o químico afirma ser capaz de producir un material que sea indistinguible de la piel humana. Es posible que en un tiempo futuro se logre esto, pero incluso si esta invención fuese posible, deberíamos pensar que tendría muy poco sentido hacer ‘una máquina pensante’ más humana poniéndole piel natural (Turing 1950, p. 41, énfasis mío).

No es de extrañar entonces que Searle, con la locuacidad que lo caracteriza, acuse a Turing de haber legado a la tradición un método conductista de verificación de vida mental y de haber simpatizado con una forma de dualismo (Searle 2004, pp. 48-49). Dicha concepción de la mente separa de manera radical los factores que causan la inteligencia, de índole biológica, de la inteligencia misma. Aunque no ahondaré en los detalles de este problema y en cómo Searle ataca otras teorías acerca de la naturaleza de lo mental, tales como el Funcionalismo, basta decir que la moraleja final es que el Test de Turing asume injustificadamente que es posible la producción artificial de inteligencia mediante pura simulación. Porque, como se ha examinado en este artículo, la simulación eficaz de la inteligencia, que ocurre a nivel epistemológico, no necesariamente implica su replicación. Y ese dogma, que desestima la sabia prudencia de Babbage de usar lenguaje mental para evaluar la conducta de un computador solo por economía, nos legó Turing con su polémico test.

CONCLUSIÓN

El Test de Turing es uno de los criterios de vida mental más debatidos y polémicos de la Filosofía de la Mente. Tal como se examinó en la primera y segunda secciones, diversos filósofos han apoyado mitos con respecto a su naturaleza, tales como que dicho método incluye una definición de inteligencia y que propone condiciones necesarias y suficientes para la existencia de ésta. Sin embargo, pese a las ambigüedades que se le han atribuido a Turing, este recalca de qué manera su test reemplaza la pregunta acerca de si las máquinas programadas o computadores pueden pensar, teniendo especialmente en cuenta las complicaciones que presenta analizar el uso y significado común de términos como ‘piensa’ y ‘máquina’. De esta manera, el Test de Turing propone un método para recabar evidencia que indique la existencia de mentalidad en computadores, y el método escogido es el denominado *Juego de la Imitación*, que consiste en que un computador induce a interrogadores a creer que es una persona. Según Turing, si la máquina logra convencer a los jueces humanos, resulta

justificado creer que es inteligente y pensante, pues es capaz de suplantar a humanos mediante comportamiento lingüístico.

El énfasis de Turing en la evidencia obtenida mediante su test y las precisiones que hace de éste como procedimiento para causar creencias, llevaron a analizar, en la tercera sección, si este test podía contar como un experimento científico tal como Fodor lo concibe. No obstante, una importante diferencia hizo desestimar que fuese un experimento adecuado, ya que el test ideado por Turing intenta convencer a los interrogadores mediante la imitación y la simulación de que el computador es una persona.

En la cuarta y última sección se analizó críticamente cómo la conducta y la simulación propuestas en el Test de Turing ayudan a consolidar el siguiente dogma: el proceso exitoso de simulación es suficiente para la duplicación de la inteligencia, puesto que ésta es una propiedad abstracta que puede ser replicada con independencia de los factores causales que la producen. Además de refutar este argumento y revelar cómo Turing apoya una forma de Conductismo Lógico, se hizo hincapié en cómo los procesos exitosos de simulación son gobernados por la verosimilitud, esto es, al igual que en la ficción, una simulación resulta eficaz y creíble si y solo si se aleja mínimamente de ciertas propiedades en el mundo real. Y esto, a su vez, sugiere por qué una simulación no implica replicación, puesto que la verosimilitud o la credibilidad de una simulación eficaz no requieren de la duplicación de propiedades. Así, el convencimiento de jueces o interrogadores es un proceso estrictamente epistémico y, por tanto, no conlleva ni mucho menos requiere de duplicación.

Contrariamente a lo que concibió el matemático-filósofo, no es posible reducir la cuestión epistemológica, involucrada en determinar si existe mentalidad, con lo que produce mentalidad (por ejemplo, los factores causales observador-independientes). Esta confusión es comúnmente apoyada por los simpatizantes del Test de Turing. Dennett, por ejemplo, afirma lo siguiente:

La imitación perfecta de Chateau Plonque es exactamente tan bueno como el vino real, aunque es una falsificación, y lo mismo sucede con los Cézanne adulterados, si realmente son indistinguibles... (Dennett 1994, p. 156).

Estos comentarios, que provocan dudas acerca del real conocimiento enológico y artístico de Dennett, sintetizan el dogma del Test de Turing. Con todo, este artículo ha expuesto dicho dogma y ha analizado cómo el divorcio entre lo epistemológico y lo ontológico impide que dicho test sea una respuesta satisfactoria al problema de las otras mentes, cuestión que a su vez explica por qué Turing, pese a su máxima, no pone punto final al debate filosófico en torno a la naturaleza de lo mental.

Referencias bibliográficas

Block, N. (1990), "The computer model of the mind", en D.N. Osherson & E.E. Smith (eds.), *Thinking: An Invitation to Cognitive Science*. Vol. 3. Cambridge, Mass.: MIT Press, pp. 247-89.

- Copeland, B.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*. Dordrecht: Kluwer Academic Publishers, pp. 1-21.
- Dennett, D.C. (1987), *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- _____ (1994), “The practical requirements for making a robot”, *Philosophical transactions of the Royal Society A* **349**: 133-46. Reprinted in D.C. Dennett *Brainchildren*. Cambridge, Mass.: MIT Press.
- French, R. (2000), “The Turing Test: The First Fifty Years”, *Trends in Cognitive Science* **4**: 115-22.
- Fodor, J.A. (1990), “The dogma that didn’t bark (A fragment of a naturalized epistemology)”, en H. Kornblith (ed.), *Naturalizing Epistemology*. Cambridge, Mass.: MIT Press, pp. 191-216.
- _____ (1994), *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, Mass.: MIT Press.
- González, R. (1997), *La Modularidad de la Mente, Observación y Teorías*. Publicaciones Especiales N° 67 (Ensayo). Serie Documentos del Grupo Cognición y Praxis. Santiago, Departamento de Filosofía, Universidad de Chile.
- _____ (2002), “Escepticismo y Epistemología Naturalizada”, *Revista de Filosofía Universidad de Chile*, **58**: 53-69.
- González, R. y Vergauwen, R. (2005), “On the verisimilitude of Artificial Intelligence”, *Logique et Analyse* **189-192**: 323-50.
- Heiser, J.F., Colby, K.M., Faught, W.S. & Parkinson, R.C. (1980), “Can psychiatrists distinguish a computer simulation of paranoia from the real thing?”, *Journal of Psychiatric Research* **15**: 148-62.
- Hodges, A. (1992), *Alan Turing: The Enigma*. London: Vintage.
- Moor, J.H. (1976), “An Analysis of the Turing Test”, *Philosophical Studies* **30**: 249-57. Reprinted in S. Shieber (ed.), *The Turing Test: Verbal Behaviour as the Hallmark of Intelligence*. Cambridge, Mass.: MIT Press, pp. 297-306.
- _____ (1987), “Turing Test”, en S.C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence*, Vol. **2**. New York, Wiley, pp. 1126-30.
- Preston, J. & Bishop, M. (2002), *Views into the Chinese Room*. Oxford: Oxford University Press.
- Putnam, H. (1965), “Brains and behaviour”, reprinted in J. Heil (ed.), *Philosophy of Mind: A Guide and Anthology*. Oxford: Oxford University Press, pp. 96-104.
- Ryan, Marie-Laure (1991), *Possible Worlds, Artificial Intelligence and Narrative Theory*. Bloomington, Ind.: Indiana University Press.
- Saygin, A.P., Cickeli, I. & Akman, V. (2003), “Turing test: 50 years later”, en J.H. Moor (ed.), *The Turing Test: The Elusive Standard of Artificial Intelligence*. Dordrecht: Kluwer Academic Publishers, pp. 23-78.

- Searle, J. (1980), "Minds, Brains and Programs", *The Behavioral and Brain Sciences* **3**: 417-424. Reprinted in M. Boden (ed.), *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, pp. 67-88.
- _____ (1990), "Is the brain's mind a computer program?", *Scientific American*, January 1990, pp. 20-25.
- _____ (2004), *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Swade, D. (2000), *The Difference Engine: Charles Babbage and the Quest to build the First Computer*. London: Penguin.
- Turing, A.M. (1936), "On computable numbers, with an application to the *Entscheidungsproblem*", *Proceedings of the London Mathematical Society*, series 2, **42**: 231-65 (with corrections in **43**: 544-6). Reprinted in M. Davis (ed.), *The Undecidable*. New York: Raven Press.
- _____ (1948), "Intelligent Machinery", National Physical Laboratory Report, en B. Meltzer & D. Michie (eds.), *Machine Intelligence* **5**, Edinburgh: Edinburgh University Press, pp. 3-23.
- _____ (1950), "Computing intelligence and machinery" from *Mind* LIX, N° 2236, pp. 433-60. Reprinted in M.A. Boden (ed.), *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, pp. 40-66.
- Weizenbaum, J. (1984), *Computer Power and Human Reason: From Judgement to Calculation*. Harmondsworth: Pelican.