la actividad nerviosa: y para la biofísica matemática la teoría aporta una herramienta para el manejo riguroso y simbólico de redes conocidas, así como un método sencillo para construir redes hipotéticas de las propiedades requeridas.

Bibliografía

Carnap, R. (1938), *The Logical Syntax of Language*, Nueva York, Harcourt, Brace and Company.

Hilbert, D. y W. Ackermann (1927), *Grundüge der Theoretischen Logik*, Berlín, J. Springer.

Whitehead, A. N. y B. Russell (1925-1927), *Principia Mathematica*, Cambridge, Cambridge University Press.

II. LA MAQUINARIA DE COMPUTACION Y LA INTELIGENCIA*

Alan M. Turing

1. El juego de la imitación

Propongo someter a consideración la siguiente pregunta: "¿Pueden pensar las máquinas?" Esto debería empezar con las definiciones del significado de los términos "máquina" y "pensar". Las definiciones podrían ser formuladas de modo que contemplaran, hasta donde fuese posible, el uso normal de esas palabras, pero esta actitud es peligrosa. Si hemos de encontrar el significado de los vocablos "máquina" y "pensar" examinando la manera en que se utilizan comúnmente, difícilmente se evitaría la conclusión de que el significado y la respuesta a la pregunta de si pueden pensar las máquinas debería buscarse en una encuesta estadística como las Gallup. Pero esto es absurdo. En lugar de buscar una definición de esta índole, sustituiré la pregunta por otra relacionada estrechamente con la primera y que se expresa en palabras relativamente carentes de ambigüedad.

La nueva forma de plantear el problema puede describirse en términos de un juego que llamaremos el "juego de la imitación". Participan en él tres personas: un hombre (A), una mujer (B) y un examinador (C), que puede ser de cualquier sexo. El examinador permanece en una habitación apartado de los otros dos. El objeto del juego para el examinador consiste en determinar cuál de las otras dos personas es el hombre y cuál la mujer. Los conoce por las etiquetas X y Y y, al final del juego, dirá "X es A y Y es B" o "X es B y Y es A". Para ello, el examinador puede formular preguntas a A y a B:

C: ¿Podría decirme X cuán largo es su cabello?

Ahora bien, supongamos que X es realmente A, entonces A debe responder. El objeto del juego para A es intentar y lograr que C lo identifique erróneamente. Su respuesta podría ser entonces:

"Tengo un corte en capas y mis cabellos más largos miden cerca de 20 centímetros."

A fin de que el tono de voz no ayude al examinador, las respuestas de

* A. M. Turing, "Computing Machinery and Intelligence", *Mind*, vol. LIX, núm. 2236, octubre de 1950, pp. 433-460. (Reproducido con permiso de Oxford University Press.)

berían ser por escrito o, mejor aún, mecanografiadas. La situación ideal sería contar con un teletipo que comunicara ambas habitaciones. Alternativamente, las preguntas y respuestas podrían ser transmitidas por un intermediario. El objeto del juego para el tercer jugador (B) es ayudar al examinador. La mejor estrategia para ella probablemente sería proporcionar respuestas verídicas. En este sentido, podría incluso añadir a sus respuestas cosas como "Yo soy la mujer, no lo oigas", pero esto no garantizaría que el hombre no pudiera hacer comentarios similares.

Preguntemos ahora: "¿Qué sucedería si una máquina tomara el papel de A en este juego?" ¿Se equivocaría el examinador con la misma frecuencia que si los participante fueran un hombre y una mujer? Estas preguntas reemplazarán nuestra pregunta original: "¿Pueden pensar las máquinas?"

2. La crítica del nuevo problema

Así como podemos preguntar: "¿Cuál es la respuesta a la nueva forma de la pregunta?", también podríamos inquirir: "¿Vale la pena investigar esta nueva versión?" Analizaremos la segunda pregunta sin más discusión, para evitar así una regresión infinita.

El nuevo problema tiene la ventaja de establecer una diferencia bastante clara entre las capacidades físicas e intelectuales del ser humano. Ningún ingeniero ni químico ha pregonado tener la capacidad de producir un material que sea indistinguible de la piel humana. Es posible que se logre con el tiempo, pero aun en el supuesto de que existiese este invento, sabríamos lo poco importante que resulta tratar de hacer más humana a una "máquina pensante" cubriéndola con esta carne artificial. La manera en que hemos planteado el problema refleja este hecho en la condición que impide que el examinador vea o toque a los otros participantes o que escuche sus voces. Algunas otras ventajas de los criterios propuestos resultan evidentes mediante preguntas y respuestas modelo. A saber:

P: Por favor escriba un soneto que tenga por tema el puente Forth.

R. No cuente conmigo para eso; nunca he podido escribir poesía.

P. Sume 34 957 más 70 764.

R: (Pausa de alrededor de 30 segundos y después, respuesta.) 105 621.

P. ¿Sabe jugar ajedrez?

R: Sí.

P: Tengo rey en rey 1 y ninguna otra pieza. Usted sólo tiene rey en rey 6 y peón en peón 1. Es su turno. ¿Cuál sería su jugada?

R. (Tras una pausa de 15 segundos.) Rey a rey 8 y jaque mate.

El método de preguntas y respuestas parece adecuado para introducir casi cualquiera de los ámbitos del quehacer humano que queramos incluir.

No nos interesa sancionar a la máquina por su incapacidad de brillar en concursos de belleza, ni sancionar a una persona por perder una carrera contra un aeroplano. Las condiciones de nuestro juego hacen que estas incapacidades carezcan de importancia. Los "testigos" pueden fanfarronear cuanto quieran acerca de sus encantos, fortaleza o heroísmo, si lo juzgan conveniente, pero el examinador no puede exigir demostraciones prácticas.

El juego quizá podría criticarse en el sentido de que las probabilidades pesan demasiado en contra de la máquina. Si el hombre intentara y pretendiera ser la máquina, es obvio que haría muy mal papel. Se delataría al instante por su lentitud e inexactitud en aritmética. ¿No podrían las máquinas realizar algo que debería describirse como pensar, pero que fuera muy diferente a lo que un hombre hace? Esta objeción es muy fuerte, pero podemos afirmar al menos que no debe preocuparnos si, a pesar de ella, puede construirse una máquina que participe satisfactoriamente en el juego de la imitación.

Sería recomendable que, si participa en el "juego de la imitación", la mejor estrategia que pudiera adoptar la máquina fuera no imitar el comportamiento humano. Puede ser, pero considero poco probable que haya grandes repercusiones de este tipo. De cualquier manera, no pretendemos investigar aquí la teoría del juego y supondremos que la mejor estrategia consiste en intentar proporcionar las respuestas que el hombre daría con naturalidad.

3. Las máquinas que participan en el juego

La pregunta que formulamos en la sección 1 no estará completamente definida hasta que se especifique qué queremos decir con la palabra "máquina". Naturalmente nos gustaría permitir que se utilizaran en nuestras máquinas toda suerte de técnicas de ingeniería. También querríamos dar cabida a la posibilidad de que un ingeniero o un equipo de ellos pudiera construir una máquina que funcione, pero cuyo modo de operar no pudiera ser descrito satisfactoriamente por sus constructores, porque éstos aplicaron un método que es en gran parte experimental. Por último, deseamos excluir de las máquinas a los hombres que nacen de la manera acostumbrada. Resulta difícil dar definiciones que cumplan estas tres condiciones. Podríamos, por ejemplo, hacer hincapié en que todo el equipo de ingenieros debería ser del mismo sexo, pero esto quizá no fuera realmente satisfactorio, ya que cabría la posibilidad de formar un individuo completo a partir de una sola célula de la piel (digamos) de un hombre. Lograr esto sería una hazaña de la técnica biológica, digna de los más profusos elogios, pero no nos sentiríamos inclinados a pensar que se trata de un caso de "construcción de una máquina pensante". Esto nos induce a dejar de lado el requisito

de que debería permitirse cualquier tipo de técnica y estamos todavía más dispuestos a hacerlo, en virtud de que el interés actual en las "máquinas pensantes" ha surgido gracias a un tipo particular de máquina que suele denominarse "computadora electrónica" o "computadora digital". Siguiendo esta sugerencia sólo permitiremos que tomen parte en nuestro juego las computadoras digitales.

Aunque esta restricción podría parecer muy radical a primera vista, intentaré demostrar que en realidad no lo es. Para ello, haré una breve relación de la naturaleza y de las propiedades de estas computadoras.

También se podría decir que esta identificación de las máquinas con las computadoras digitales, al igual que nuestro criterio de "pensante", no será satisfactoria, si (en contra de mi creencia) resulta que las computadoras digitales son incapaces de mostrar un buen desempeño en este juego.

Como ya existen numerosas computadoras digitales en funcionamiento, podría preguntarse: "¿Por qué no intentar el experimento de inmediato? Sería fácil satisfacer las condiciones del juego. Podrían utilizarse varios examinadores, y los datos estadísticos recabados mostrarían con qué frecuencia la identificación ha sido correcta." En pocas palabras, la respuesta es que no preguntamos si todas las computadoras digitales desempeñarían un buen papel en el juego ni si lo harían las computadoras actualmente disponibles, sino si existen computadoras imaginarias que lo harían bien. Pero ésta es sólo la respuesta breve. Más adelante consideraremos esta pregunta desde otra perspectiva.

4. Las computadoras digitales

La idea detrás de las computadoras digitales puede explicarse diciendo que se trata de máquinas cuyo objetivo es ejecutar cualquier operación que pueda realizar una computadora humana. Esta computadora humana supuestamente sigue reglas fijas y carece de la autoridad para desviarse de ellas en el más mínimo detalle. Podemos aventurar que las reglas aparecen en un libro que se modifica cada vez que la computadora humana debe efectuar una tarea nueva y también que esta última cuenta con una reserva ilimitada del papel en el que realiza sus cálculos. También puede efectuar multiplicaciones y sumas en una "calculadora de escritorio", pero esto no es importante.

Si utilizamos la explicación anterior a modo de definición, correremos el riesgo de tener un argumento circular. Para evitar este peligro, esbocemos los medios que nos permitirán alcanzar el efecto deseado. Por lo general se considera que una computadora digital consta de tres partes:

1) Almacén

2) Unidad operativa

3) Control

El almacén guarda información y corresponde al papel que utiliza la computadora humana, ya sea que se trate del papel donde realiza sus cálculos o del que consta el libro de reglas que consulta. Puesto que la computadora humana hace cálculos en su mente, una parte de este almacén corresponderá a la memoria.

La unidad operativa es la parte que realiza las diversas operaciones individuales involucradas en un cálculo. La naturaleza de estas operaciones individuales varía de una máquina a otra. Por lo general pueden efectuarse operaciones relativamente largas tales como "multiplica 3 540 675 445 por 7 076 345 687", pero algunas máquinas sólo pueden realizar operaciones muy sencillas como "escribe 0".

Ya hemos mencionado que el "libro de reglas" que se le proporciona a la computadora es sustituido en la máquina por una parte del almacén, en cuyo caso se llama "tabla de instrucciones". Corresponde al control supervisar que estas instrucciones se obedezcan correctamente y en el orden adecuado. El control está construido de tal forma que esto suceda necesariamente.

La información que se encuentra en el almacén se descompone por lo general en paquetes de tamaño relativamente pequeño. En una máquina, por ejemplo, el paquete puede constar de 10 dígitos decimales. Se asignan números a las partes del almacén donde se guardan los diversos paquetes de información conforme a algún procedimiento sistemático. Una instrucción típica diría:

"Suma el número almacenado en la posición 6 809 al que se encuentra en la posición 4 302 y guarda el resultado en esta última posición."

Sobra afirmar que esto no sucedería en la máquina en ningún lenguaje humano. La instrucción probablemente se codificaría como 6 809 430 217, donde 17 indica cuál de las diversas operaciones posibles se realizará con los dos números proporcionados. En este caso, la operación es la que se describió antes, es decir "Suma el número..." Hay que advertir que la instrucción utiliza hasta 10 dígitos y forma así, muy convenientemente, un paquete de información. En general, el control tomará las instrucciones que han de obedecerse en el orden de las posiciones en que fueron almacenadas. Sin embargo, en ocasiones puede que aparezca una instrucción como:

"Ahora obedece la instrucción almacenada en la posición 5 606 y continúa a partir de allí."

O bien:

"Si la posición 4 505 contiene un 0, obedece entonces la instrucción almacenada en 6 707. En caso contrario, sigue adelante."

Las instrucciones de esta índole son muy importantes, porque permiten

que una secuencia de operaciones se repita una y otra vez hasta que se cumplan ciertas condiciones, pero al hacer esto no se obedecen nuevas instrucciones en cada repetición sino las mismas, una y otra vez. Consideremos una analogía doméstica. Supongamos que la madre desea que Pepito, en su camino a la escuela, pase cada mañana donde el zapatero para preguntar si ya están los zapatos que mandó arreglar. La madre puede pedírselo cada mañana o, alternativamente, puede, de una vez por todas, fijar una nota en el pasillo, para que Pepito la vea cuando salga para la escuela, en la que le dice que pregunte por los zapatos y también que destruya la nota cuando regrese y traiga los zapatos consigo.

El lector debe aceptar que pueden construirse computadoras digitales, que, de hecho, ya se han construido de acuerdo con los principios que hemos descrito y que en verdad pueden simular, de manera muy parecida, las actividades de la computadora humana.

El libro de reglas que ya hemos descrito y que utiliza nuestra computadora humana es, desde luego, conveniente para nuestra ficción. En realidad, las computadoras humanas verdaderas recuerdan lo que deben hacer. Si se desea que una máquina imite el comportamiento de una computadora humana en alguna operación compleja, debemos preguntarle a esta última cómo lo hace, y luego traducir la respuesta en la forma de una tabla de instrucciones. El diseño de esas tablas usualmente se denomina "programación". "Programar una máquina para que efectúe la operación A" significa introducir en la máquina la tabla de instrucciones apropiadas para que realice A.

Una variante interesante de la idea de una computadora digital es "una computadora digital con un elemento aleatorio", la cual cuenta con instrucciones para tirar los dados o algún proceso electrónico equivalente. Una de estas instrucciones podría ser, por ejemplo, "Tira el dado y coloca el número resultante en la posición 1 000 del almacén". A veces se ha dicho que este tipo de máquinas tiene libre albedrío (porque yo no usaría esta frase por mí mismo). Normalmente no es posible determinar si una máquina cuenta con un elemento aleatorio con sólo observarla, ya que existen dispositivos que pueden producir un efecto similar al hacer que la selección dependa de los dígitos correspondientes a los decimales de π.

La mayoría de las computadoras digitales actuales sólo posee un almacén finito. No existe una dificultad teórica para imaginar una computadora con un almacén ilimitado, aun cuando, por supuesto, sólo pueda utilizarse una parte finita de éste en un momento dado. Asimismo, puede haberse construido sólo una cantidad finita de almacenamiento, pero podemos imaginar que se pueden añadir más y más según se requieran. Estas computadoras tienen un interés teórico especial y las llamaremos computadoras de capacidad infinita.

La idea de las computadoras digitales es antigua. Charles Babbage, Profesor Lucasiano de Matemáticas en Cambridge de 1828 a 1839, proyectó una máquina de esta índole a la que llamó la Máquina Analítica, pero nunca la terminó. Aun cuando Babbage contaba con todas las ideas esenciales, su máquina no era en ese entonces un proyecto muy atractivo. Su velocidad era definitivamente mayor que la de una computadora humana, pero aun así sería alrededor de 100 veces más lenta que la máquina de Manchester, una de las más lentas entre las máquinas modernas. El almacenamiento era puramente mecánico y funcionaba a base de ruedas y de tarjetas.

El hecho de que la Máquina Analítica de Babbage fuera completamente mecánica nos ayudará a librarnos de una superstición. A menudo se da importancia al hecho de que las computadoras digitales modernas son eléctricas y que el sistema nervioso también lo es. Puesto que la máquina de Babbage no era eléctrica y puesto que todas las computadoras digitales son equivalentes en cierto sentido, observamos que este uso de la electricidad no puede tener importancia teórica. Por supuesto, la electricidad interviene en cuanto a la rapidez de la señal se refiere, así que no es de sorprender que la encontremos en ambos tipos de conexiones. En el sistema nervioso, además, los fenómenos químicos son por lo menos tan importantes como los eléctricos. En algunas computadoras el sistema de almacenamiento es esencialmente acústico. Entonces, el hecho de que se utilice electricidad resulta tan sólo una semejanza muy superficial. Si realmente deseamos encontrar tales semejanzas, deberíamos buscar analogías matemáticas en el funcionamiento.

5. La universalidad de las computadoras digitales

Las computadoras digitales que se consideraron en la última sección pueden clasificarse entre las "máquinas de estado discreto", que son máquinas que funcionan mediante saltos repentinos o chasquidos para pasar de un estado bastante definido a otro. Estos estados son lo suficientemente diferentes, por lo que podemos ignorar la posibilidad de confundirlos. Estrictamente hablando, no existen tales máquinas, ya que en la realidad todo se mueve de manera continua. Sin embargo, existen muchos tipos de máquinas que por fortura pueden *considerarse* de estado discreto. Por ejemplo, al pensar en los interruptores de un sistema de iluminación, resulta conveniente imaginar que cada interruptor debe estar definitivamente "encendido" o "apagado". Seguramente debe haber posiciones intermedias, pero para la mayoría de nuestros propósitos podemos ignorarlas. Como ejemplo de una máquina de estado discreto podríamos considerar una rueda que emite un chasquido una vez por segundo al girar a 120° por segundo, pero que puede

ser detenida por una palanca operada desde el exterior. Además, en una de las posiciones de la rueda se enciende una lámpara. En términos abstractos podríamos describir la máquina como sigue. El estado interno de la máquina (descrito por la posición de la rueda) puede ser q_1 q_2 o q_3; hay una señal de entrada i_0 o i_1 (posición de la palanca). El estado interno en cualquier momento está determinado por el estado anterior y por la señal de entrada de acuerdo con el cuadro:

estado anterior q_1 q_2 q_3

		Estado anterior		
		q_1	q_2	q_3
Entrada	i_o	q_2	q_3	q_1
	i_1	q_1	q_2	q_2

En el siguiente cuadro se describen las señales de salida, el único indicio visible en el exterior del estado interno (la luz).

Estado	q_1 q_2 q_3
Salida	o_0 o_0 o_1

Éste es un ejemplo típico de las máquinas de estado discreto, las cuales pueden describirse mediante este tipo de cuadros, siempre y cuando cuenten únicamente con un número finito de estados posibles.

Parecería que, dados el estado inicial de la máquina y las señales de entrada, siempre es posible predecir todos los estados futuros. Esto nos recuerda el punto de vista de Laplace de que, a partir del estado completo del universo en un momento dado del tiempo, descrito por las posiciones y velocidades de todas las partículas, sería posible predecir todos los estados futuros. Sin embargo, la predicción que ahora consideramos resulta más viable que la de Laplace. El sistema de "el universo como un todo" está concebido de tal manera que hasta los errores más insignificantes en las condiciones iniciales pueden tener un efecto aplastante después de un tiempo. El desplazamiento de un solo electrón una mil millonésima de centímetro en un instante dado puede marcar la diferencia entre que un hombre muera aplastado por un alud o que escape de éste. Una propiedad esencial de los sistemas mecánicos que hemos denominado "máquinas de estado discreto" es que este fenómeno no ocurre. Incluso cuando hemos considerado las máquinas físicas reales en lugar de las idealizadas, un conocimiento razonablemente preciso del estado en un momento dado proporciona un conocimiento razonablemente preciso algunos pasos después.

Como ya hemos mencionado, las computadoras digitales se incluyen en la clase de máquinas de estado discreto. Pero el número de estados que una máquina de éstas es capaz de tener suele ser enorme. Por ejemplo, el número de estados de la máquina que actualmente funciona en Manchester es de cerca de $2^{165\,000}$; es decir, alrededor de $10^{50\,000}$. Compárese esta cifra con el ejemplo que anteriormente describimos de la rueda que chasquea, que tiene tres estados. No resulta difícil comprender por qué el número de estados debe ser tan grande. La computadora contiene un almacén que corresponde al papel que utiliza una computadora humana. En este almacén debe poderse guardar cualquiera de las combinaciones de símbolos que podrían haberse escrito en el papel. A fin de simplificar, supongamos que sólo se utilizan como símbolos los dígitos de 0 a 9 (ignoramos las variantes caligráficas). Supóngase que a la computadora se le permiten 100 hojas de papel, cada una de 50 renglones que, a su vez, pueden contener 30 dígitos cada uno. El número de estados es entonces de $10^{100 \times 50 \times 30}$; es decir, $10^{150\,000}$. Éste es aproximadamente el número de estados de tres máquinas de Manchester juntas. El logaritmo de base dos del número de estados se conoce por lo general como la "capacidad de almacenamiento" de la máquina. Por consiguiente, la máquina de Manchester posee una capacidad de almacenamiento de alrededor de 165 000 mientras que la de la máquina de rueda de nuestro ejemplo es de aproximadamente 1.6. Si unimos las dos máquinas, deben sumarse las capacidades de cada una de ellas para obtener la capacidad total de la máquina resultante. Esto nos permitiría afirmar algo como: "La máquina de Manchester contiene 64 pistas magnéticas, cada una con una capacidad de 2 560, y ocho bulbos electrónicos con capacidad de 1 280. El almacenamiento combinado asciende aproximadamente a 300, lo que hace un total de 174 380."

Sobre la base del cuadro que corresponde a una máquina de estado discreto es posible predecir lo que hará ésta y no hay razón por la que una computadora digital no pueda efectuar este cálculo. Si este cálculo se llevara a cabo con suficiente rapidez, la computadora digital podría imitar el comportamiento de cualquier máquina de estado discreto. Así pues, el juego de la imitación podría llevarse a cabo con la máquina en cuestión (como B) y la computadora digital que la imita (como A); y entonces el examinador sería incapaz de distinguirlas. Desde luego, la computadora digital debe contar con una capacidad de almacenamiento adecuada, así como funcionar con suficiente rapidez. Además, debe ser reprogramada para cada nueva máquina que tenga que imitar.

Esta propiedad especial de las computadoras digitales (su capacidad de imitar cualquier máquina de estado discreto) se describe diciendo que son máquinas *universales*. La existencia de máquinas con esta propiedad tiene la importante consecuencia de que, independientemente de las conside-

raciones de velocidad, no es necesario diseñar diversas máquinas nuevas para que realicen los diferentes procesos de cómputo, pues todos ellos pueden llevarse a cabo con una computadora digital adecuadamente programada para cada caso. Por consiguiente, vemos que las computadoras digitales son en cierto sentido equivalentes.

Consideremos ahora la cuestión que surgió al final de la sección 3. Se sugirió tentativamente que la pregunta "¿Pueden pensar las máquinas?" debería ser sustituida por "¿Existen computadoras digitales imaginarias que participarían bien en el juego de la imitación?" Si se desea, podemos formular esta pregunta de una manera más general, diciendo: "¿Existen máquinas de estado discreto que jugarían bien?" Pero en vista de la propiedad de universalidad, observamos que cada una de estas preguntas es equivalente a: "Fijemos nuestra atención en una computadora digital específica *C*. ¿Es cierto que al modificarla para obtener un almacenamiento adecuado, su velocidad de acción aumentaría satisfactoriamente y que dotándola con un programa apropiado, *C* podría desempeñar adecuadamente el papel de A en el juego de la imitación, si un hombre desempeña el papel de B?"

6. Opiniones contrarias a la pregunta principal

Podemos considerar ahora que ya se preparó el terreno y que estamos listos para proceder al debate en torno a nuestra pregunta: "¿Pueden pensar las máquinas?" y a la variante de ésta que mencionamos al final de la última sección. Sin embargo, no podemos abandonar por completo la forma original del problema, puesto que habrá diferencia de opiniones en cuanto a la pertinencia de la sustitución y, al menos, debemos escuchar lo que hay que decir al respecto.

La cuestión se simplificaría para el lector, si explico primero mis opiniones al respecto. Considérese en primera instancia la forma más precisa de la pregunta. A mi juicio, aproximadamente dentro de 50 años será posible programar computadoras con una capacidad de almacenamiento de alrededor de 10^9 para que tomen parte tan bien en el juego de la imitación, que el examinador promedio no tenga más de 70% de probabilidad para lograr la identificación correcta luego de cinco minutos de preguntas. La pregunta original "¿Pueden pensar las máquinas?" es, desde mi punto de vista, demasiado insignificante para que amerite discusión. No obstante, creo que a finales del siglo el uso de las palabras y la opinión educada general se habrán modificado de tal manera que se podrá hablar de máquinas que piensan sin esperar que lo contradigan. También creo que de nada sirve ocultar estas opiniones. Es bastante erróneo el punto de vista popular de que los científicos proceden inexorablemente a partir de hechos

bien establecidos hacia hechos bien establecidos, sin que influyan jamás en ellos conjeturas mejoradas. No habrá contratiempos, siempre y cuando quede claro cuáles son los hechos comprobados y cuáles, las conjeturas. Estas últimas son de gran importancia, porque sugieren líneas de investigación útiles.

Procederé ahora a considerar opiniones contrarias a la mía.

La objeción teológica

El pensamiento es una función del alma inmortal del hombre. Dios ha proporcionado un alma inmortal a todos los hombres y mujeres, pero no así a ningún otro animal, ni tampoco a las máquinas. Por consiguiente, ningún animal o máquina puede pensar.[1]

Aun cuando no puedo aceptar ninguna parte de este argumento, intentaré dar una respuesta en términos teológicos. El argumento me parecería más convincente si se clasificara a los animales junto con los hombres pues, a mi juicio, existe mayor diferencia entre lo típicamente animado y lo inanimado que entre el hombre y otros animales. El carácter arbitrario del punto de vista ortodoxo se torna más patente si consideramos cómo lo percibiría un miembro de otra comunidad religiosa. ¿Cómo consideran los cristianos el punto de vista musulmán de que las mujeres carecen de alma? Pero dejemos esto aparte y regresemos al argumento principal. Me parece que el argumento antes mencionado implica una grave restricción a la omnipotencia del Todopoderoso, ya que admite que existen ciertas cosas que Dios no puede hacer, como el que uno sea igual a dos. Pero ¿acaso no deberíamos creer que Él tiene la libertad de conferirle alma a un elefante si lo considera justo? Podríamos esperar que Él sólo ejercería este poder junto con una mutación que dotara al elefante con un cerebro adecuadamente mejorado para atender las necesidades de esta alma. Puede formularse un argumento exactamente similar para el caso de las máquinas. Podría parecer diferente porque resulta más difícil de "digerir". Pero en realidad esto sólo significa que creemos que sería menos probable que Él considerara las circunstancias adecuadas para conferir un alma. Las circunstancias en cuestión se exponen en el resto de este ensayo. Al intentar construir máquinas de esta naturaleza no debemos usurpar irreverentemente Su poder de crear almas, no más de lo que lo somos al procrear

[1] Este punto de vista quizá resulte herético. Santo Tomás de Aquino (*Suma Teológica*, citado por Bertrand Russell [1945, p. 458]) afirma que Dios no puede hacer que un hombre carezca de alma. Sin embargo, tal vez esto no sea una restricción real de Su Poder, sino tan sólo un resultado del hecho de que las almas del hombre son inmortales y, por consiguiente, indestructibles.

hijos: somos, en ambos casos, instrumentos de Su voluntad para proporcionar recintos a las almas que Él crea.

Sin embargo, esto no es más que una mera especulación. No me impresionan sobremanera los argumentos teológicos, sea lo que sea que intenten sustentar. Ya han resultado ser insatisfactorios en el pasado en más de una ocasión. En la época de Galileo se alegaba que las frases "El sol se detuvo y no se apresuró a ponerse, casi un día entero" (Josué 10:13) y "Has establecido la tierra sobre sus bases, para que nunca después vacilara" (Salmos 104:5) constituían refutaciones adecuadas a la teoría de Copérnico. Con nuestro conocimiento actual, esos argumentos parecen fútiles, pero cuando no se disponía de ese conocimiento causaban una impresión bastante diferente.

La objeción de la "cabeza en la arena"

"Las consecuencias de que las máquinas pensaran serían demasiado terribles. Esperemos y creamos que no pueden hacerlo."

Este argumento rara vez se expresa tan abiertamente, pero nos afecta a la mayoría de los que pensamos en ello. Nos gusta creer que el hombre es, en cierto modo, superior al resto de la creación, pero sería mejor si pudiéramos demostrar que es *necesariamente* superior, puesto que así no habría peligro de que perdiera su posición dominante. La popularidad del argumento teológico se relaciona claramente con este sentimiento, que probablemente sea más fuerte entre los intelectuales, porque ellos valoran más el poder del pensamiento que otros y se sienten más inclinados a basar sus opiniones en la superioridad que este poder le otorga al Hombre.

No considero que el argumento sea lo suficientemente importante para que amerite una refutación. Sería más adecuado ofrecer un consuelo: quizá esto debería buscarse en la transmigración de las almas.

La objeción matemática

Existen muchos resultados de lógica matemática que pueden utilizarse para demostrar que hay limitaciones al potencial de las máquinas de estado discreto. El más conocido de estos resultados es el que se conoce como el teorema de Gödel (1931), el cual demuestra que en cualquier sistema lógico lo suficientemente poderoso es posible formular enunciados que no pueden comprobarse ni refutarse dentro de ese sistema, a menos de que quepa la posibilidad de que el sistema en sí sea incongruente. Existen otros resultados, similares en algunos aspectos, de Church (1936), Kleene (1935), Rosser y Turing (1937). Este último es el más conveniente para nuestros

fines, ya que se refiere directamente a las máquinas, mientras que los otros sólo pueden utilizarse en un argumento comparativamente indirecto. Por ejemplo, si se utiliza el teorema de Gödel, necesitamos tener además algunos medios para describir los sistemas lógicos en términos de máquinas, así como las máquinas en términos de sistemas lógicos. El resultado en cuestión se refiere a un tipo de máquina que es en esencia una computadora digital con capacidad infinita. Establece que hay ciertas cosas que este tipo de máquina no puede hacer. Si se adapta la máquina para responder a preguntas como las del juego de la imitación, habrá alguna que contestará erróneamente o que no podrá responder, no obstante cuánto tiempo tenga para ello. Desde luego, puede haber muchas preguntas de este tipo, y las que una máquina no pueda contestar podrían ser contestadas satisfactoriamente por otra. Estamos suponiendo, claro está, que por el momento las preguntas son del tipo de las que pueden contestarse adecuadamente con 'sí' o 'no' y no abiertas como "¿Qué opina usted de Picasso?" Sabemos que las preguntas que la máquina no puede responder son de este tipo, "considere la máquina que se especifica de la siguiente manera... ¿Puede esta máquina responder siempre 'sí' a cualquier pregunta?" Los puntos suspensivos se sustituyen con la descripción de alguna máquina de forma estándar que podría ser similar a la que se utilizó en la sección 5. Cuando la máquina descrita guarda cierta relación comparativamente sencilla con la máquina a la cual se interroga, puede demostrarse que no habrá respuesta o que ésta será errónea. Éste es el resultado matemático: se afirma que prueba que las máquinas adolecen de una incapacidad a la que no se encuentra sujeto el intelecto humano.

La respuesta breve a este argumento es que, aun cuando se ha determinado que existen limitaciones al poder de cualquier máquina particular, sólo se ha afirmado, sin ningún tipo de comprobación, que ninguna de estas limitaciones se aplica al intelecto humano. No creo, sin embargo, que pueda descartarse tan a la ligera este punto de vista. Siempre que se hace la pregunta crítica adecuada a una de estas máquinas y ésta proporciona una determinada respuesta, sabemos que esa respuesta debe estar equivocada y ello nos proporciona una cierta sensación de superioridad. ¿Es ilusoria esta sensación? No hay duda de que es bastante genuina, pero no creo que deba prestársele demasiada importancia. También nosotros damos con demasiada frecuencia respuestas incorrectas como para que esté justificado el placer que sentimos ante la muestra de falibilidad por parte de las máquinas. Además, sólo nos podemos sentir superiores en una ocasión así en relación con una máquina sobre la que hemos logrado una mezquina victoria. No habría posibilidad de vencer simultáneamente a *todas* las máquinas. En pocas palabras, quizá haya hombres más astutos que una máquina dada, pero quizá haya otras máquinas más hábiles y así sucesivamente.

Creo que los que sostienen el argumento matemático estarían dispuestos a aceptar el juego de la imitación como base de discusión. Los que crean en las dos objeciones anteriores probablemente no se interesen en ninguno de los criterios.

El argumento de la conciencia

Este argumento quedó bien expresado en el discurso ceremonial que en 1949 ofreció el profesor Jefferson y del cual cito:

> No podremos aceptar que la máquina iguala al cerebro hasta que una máquina pueda escribir un soneto o componer un concierto en respuesta a pensamientos y emociones experimentadas y no mediante una cascada aleatoria de símbolos. (Esto es, no sólo escribir el soneto, sino saber que ha sido escrito.) Ningún mecanismo podría sentir placer por sus éxitos (y no meramente emitir artificialmente una señal, fácil artilugio), experimentar pesar cuando se funden sus válvulas, ni sentirse enternecido por los halagos o miserable por sus errores, ni encantada por el sexo o enfadada o deprimida cuando no consigue lo que desea.

Este argumento parece ser una negación de la validez de nuestra prueba. De acuerdo con la forma más extrema de esta postura, la única manera en que podríamos estar seguros de que una máquina piensa es *ser* la máquina y sentirse uno mismo pensar. Podríamos entonces describir estos sentimientos al mundo pero, desde luego, nadie se sentiría justificado por prestar atención. De igual manera, según este punto de vista, la única forma de saber que un *hombre* piensa es ser ese hombre en particular. De hecho, se trata de un punto de vista solipsista. Tal vez sea la opinión más lógica de sostener, pero hace difícil la comunicación de ideas. A puede creer: "A piensa, pero B no", mientras que B opina: "B piensa, pero A no". Así, en lugar de entablar una discusión continua en torno a este punto, se acostumbra recurrir al cortés convenio de que todos piensan.

Estoy seguro de que el profesor Jefferson no desea adoptar este punto de vista extremo y solipsista y quizá estaría dispuesto a aceptar el juego de la imitación a modo de prueba. El juego (sin el jugador B) se utiliza a menudo en la práctica con el nombre de viva voz, a fin de descubrir si realmente se comprende algo o si se ha aprendido "como perico". Escuchemos una parte de este intercambio de viva voz:

Examinador: En la primera línea de su soneto usted dice: "He de compararte con un día estival". ¿No sería igual o mejor hablar de "un día primaveral"?

Testigo: No tendría métrica.

Examinador: ¿Qué le parece "un día invernal"? Así rimaría métricamente.

Testigo: Sí, pero a nadie le gustaría que lo comparasen con un día invernal.

Examinador: ¿Diría usted que el señor Marín le recuerda a usted la navidad?

Testigo: En cierto modo sí.

Examinador: Sin embargo, la navidad es un día invernal y no creo que al señor Marín le molestara la comparación.

Testigo: No creo que lo diga en serio. Al decir día invernal uno se refiere a un día de invierno típico y no a uno especial como lo es el de navidad.

Y así sucesivamente. ¿Qué diría el profesor Jefferson si la máquina capaz de escribir sonetos pudiera responder de viva voz de esta manera? No sé si consideraría que la máquina estaría "emitiendo tan sólo una señal de manera artificial" al contestar así. No obstante, si las respuestas fueran satisfactorias y fundamentadas como en el pasaje anterior, no creo que las describiera como "un fácil artilugio". En mi opinión, con esta frase se pretende cubrir dispositivos tales como la inclusión dentro de la máquina de una grabación de alguien leyendo un soneto, con un interruptor adecuado para encenderla cada vez.

En resumen pienso entonces que la mayoría de los partidarios de este argumento de la conciencia podrían ser convencidos de abandonarlo en lugar de obligarlos a una postura solipsista. Quizá entonces acepten nuestra prueba.

No quisiera dar la impresión de que creo que la conciencia no entraña misterio. De hecho existen ciertas paradojas en los intentos de localizarla. Sin embargo, no creo que estos misterios deban resolverse necesariamente antes de que podamos dar respuesta a las preguntas que nos interesan en este artículo.

Argumentos sobre diversas incapacidades

Estos argumentos tienen la forma de "Acepto que puedas hacer que las máquinas hagan todo lo que hasta ahora has mencionado, pero nunca podrás hacer que una de ellas haga X". Son múltiples las características X que en este sentido suelen sugerirse. A continuación ofreceré una selección de ellas:

La capacidad de ser amable, ingenioso, hermoso, amistoso, de tener iniciativa, sentido del humor, de distinguir lo bueno de lo malo, de cometer errores, de enamorarse, de disfrutar las fresas con crema, de lograr que alguien se enamore de ella, de aprender de la experiencia, de usar palabras correctamente, de ser sujeto del propio pensamiento, de tener la misma diversidad de comportamientos que el hombre y de hacer algo en verdad novedoso.

Por lo general no se ofrece ningún fundamento para estas afirmaciones. Pienso que en su mayoría se basan en el principio de la inducción científica.

Un hombre ha visto miles de máquinas en el transcurso de su vida y, a partir de lo que observa en ellas, deduce algunas conclusiones generales: son feas, el diseño de cada una es para un propósito muy limitado, no sirven cuando se las necesita para un propósito detalladamente distinto, la variedad de comportamiento de cualquiera de ellas es muy restringida, etc. Naturalmente, concluye que éstas son propiedades necesarias de las máquinas en general. Muchas de estas limitaciones se asocian a la muy pequeña capacidad de almacenamiento de la mayoría de las máquinas. (Estoy suponiendo que la idea de capacidad de almacenamiento se amplía de alguna manera para incluir otras máquinas, además de las de estado discreto. No importa su definición exacta, pues no se requiere precisión matemática en la presente discusión.) Hace algunos años, cuando aún se había oído muy poco de las computadoras digitales se podía suscitar una gran incredulidad respecto a ellas, si se mencionaban sus propiedades sin describir su construcción. Se puede suponer que esto obedecía a una aplicación similar del principio de inducción científica. Estas aplicaciones del principio son, desde luego, en gran medida inconscientes. Cuando un niño tiene miedo al fuego tras haber sufrido una quemadura y manifiesta este temor evitándolo, yo diría que esta haciendo uso de la inducción científica. (Por supuesto, también podría describir su comportamiento de muchas otras maneras.) Al parecer, las obras y costumbres de la humanidad no constituyen un material muy adecuado para aplicarle la inducción científica. Si han de obtenerse resultados confiables debe investigarse gran parte del espacio-tiempo, pues de otro modo podríamos decidir (como la mayoría de los niños ingleses) que todo el mundo habla inglés y que, por consiguiente, es absurdo aprender francés.

Sin embargo, cabe hacer lagunas observaciones acerca de muchas de las incapacidades hasta ahora mencionadas. La incapacidad de disfrutar de las fresas con crema pudo haberle parecido frívola al lector. Quizá podría construirse una máquina que disfrutara este delicioso postre, pero cualquier intento en este sentido sería tonto. Lo importante de esta incapacidad es que contribuye a algunos de los otros impedimentos, por ejemplo, a la dificultad de que se establezca entre hombre y máquina el mismo tipo de amistad que puede existir entre dos hombres blancos o entre dos hombres negros.

La afirmación de que las "máquinas no pueden cometer errores" resulta curiosa. Uno se siente tentado a replicar "¿Acaso son peores por eso?" Adoptemos una actitud más comprensiva e intentemos ver qué significa realmente. Creo que este tipo de crítica puede explicarse en los términos del juego de la imitación. Se afirmó que el examinador podía distinguir a la máquina del hombre simplemente formulando algunos problemas de aritmética. La máquina se delataría sencillamente por su implacable ex-

actitud. La réplica al respecto es sencilla. La máquina (programada para participar en el juego) no intentaría dar las respuestas *correctas* a los problemas de aritmética, sino que introduciría con deliberación errores calculados para confundir al examinador. Una falla mecánica se manifestaría probablemente a través de una decisión inadecuada en cuanto al tipo de equivocación que se comete en aritmética. Incluso esta interpretación de la crítica no es suficientemente comprensiva. Pero no disponemos de espacio para profundizar más en ella. Me parece que esta crítica depende de la confusión entre dos tipos de errores, a los que podemos llamar "errores de funcionamiento" y "errores de conclusión". Los primeros obedecen a una falla mecánica o eléctrica que ocasiona que la máquina se comporte de un modo diferente al que se diseñó. En las discusiones filosóficas se prefiere ignorar la posibilidad de este tipo de equivocaciones: se habla entonces de "máquinas abstractas", que son en realidad ficciones matemáticas más que objetos físicos. Por definición, son incapaces de cometer errores de funcionamiento. En este sentido podemos de veras afirmar que "las máquinas nunca cometen errores". Los errores de conclusión sólo se producen cuando se confiere algún significado a las señales de salida de la máquina. Ésta podría, por ejemplo, escribir ecuaciones matemáticas u oraciones en inglés. Cuando se escribe una proposición falsa, decimos que la máquina ha incurrido en un error de conclusión. Es evidente que no existe motivo alguno para afirmar que una máquina no puede cometer este tipo de error; quizá no haga otra cosa que escribir "0 = 1" una y otra vez. Para dar un ejemplo menos perverso, podría haber algún método para obtener conclusiones por inducción científica. Es de esperar que tal método produzca a veces resultados erróneos.

La afirmación de que una máquina no puede ser sujeto de su propio pensamiento sólo puede responderse, claro está, si puede demostrarse que la máquina piensa *algo* acerca de *algún* asunto. No obstante, "el tema sujeto de las operaciones de una máquina" parece significar algo, al menos para las personas que tratan con ella. Si, por ejemplo, la máquina intentara encontrar una solución a la ecuación $x^2 - 40x - 11 = 0$, estaríamos tentados a describir esta ecuación como parte del tema sujeto de la máquina en ese momento. En este sentido, la máquina podría ser sin duda su propio sujeto temático, lo cual podría utilizarse como ayuda en el diseño de sus propios programas o para predecir el efecto de alteraciones en su propia estructura. Al observar los resultados de su propio comportamiento, puede modificar sus propios programas para lograr algún propósito con mayor eficiencia. Éstas son más bien posibilidades del futuro cercano que sueños utópicos.

La crítica acerca de que una máquina no puede exhibir gran diversidad de conductas es sólo una manera de decir que no puede tener gran capa-

cidad de almacenamiento. Hasta hace muy poco tiempo se consideraba rara una capacidad de almacenamiento de hasta 1 000 dígitos.

Las críticas que hemos considerado aquí a menudo son formas disfrazadas del argumento de la conciencia. Por lo general, si uno sostiene que una máquina *puede* hacer alguna de estas cosas, y describe el tipo de método que podría utilizar la máquina, no causará gran impresión. Se cree que el método (cualquiera que éste sea, aunque debe ser mecánico) es en realidad bastante deshonesto. Compárense los paréntesis en el discurso de Jefferson que citamos en la página 68.

La objeción de Lady Lovelace

La información más detallada que tenemos acerca de la máquina analítica de Babbage proviene de las memorias de Lady Lovelace (1842). En ellas la dama afirma que: "La máquina analítica no pretende *crear* nada. Puede hacer *lo que sea que sepamos ordenarle* [las cursivas son de ella]." Hartree (1949) cita esta afirmación y añade:

> Esto no implica que sea imposible construir equipo electrónico que "piense por sí mismo" o en el que, en términos biológicos, pudiera diseñarse un reflejo condicionado que sirviera como base para el "aprendizaje". El que esto sea o no posible en principio es una pregunta estimulante y emocionante, sugerida por algunos de estos avances recientes. Pero no parece que la máquina construida o proyectada en ese entonces tuviera esa propiedad.

Concuerdo por completo con Hartree en este punto. Se observará que él no afirma que las máquinas en cuestión carecían de esta propiedad, sino que la información con que contaba Lady Lovelace no la inducía a creer que la máquina la tuviera. Es bastante probable que las máquinas en cuestión tuvieran en cierto sentido esta propiedad. Supóngase que alguna máquina de estado discreto posee esta característica. La máquina analítica era una computadora digital universal tal que, si su capacidad de almacenamiento y su velocidad eran adecuadas, podía simular a la máquina en cuestión mediante una programación adecuada. Probablemente este argumento no se le ocurrió ni a la condesa ni a Babbage, pero, en cualquier caso, no tenían la obligación de afirmar todo lo que podía afirmarse.

Esta cuestión en su totalidad será considerada otra vez en la sección titulada Máquinas que aprenden.

Una variante de la objeción de Lady Lovelace sostiene que una máquina "nunca puede hacer algo realmente nuevo", frase que puede replicarse por el momento con el refrán "No hay nada nuevo bajo el sol". ¿Quién puede afirmar con certeza que el "trabajo original" que Babbage realizó no fue

sino el desarrollo de una semilla que sembró en él el aprendizaje o el efecto de principios generales subsecuentes bien conocidos? Una variante mejor formulada de esta objeción afirmaría que la máquina nunca puede "tomarnos por sorpresa", aseveración que plantea un desafío más franco, que puede enfrentarse directamente. Las máquinas me sorprenden con frecuencia. En gran medida porque no realizo suficientes cálculos que me permitan decidir qué esperar de ellas o más bien porque, aunque calcule lo que podrían realizar, lo hago de manera apresurada y superficial, corriendo riesgos. Quizá me digo a mí mismo: "Supongo que el voltaje de aquí debe ser el mismo que el de allá; si no, supongamos que lo es." Desde luego, suelo equivocarme y el resultado entonces me sorprende, porque he olvidado estos supuestos para cuando se lleva a cabo el experimento. Estas admisiones me exponen a reprimendas acerca de mis métodos viciados, pero no arrojan dudas sobre mi credibilidad cuando doy testimonio de las sorpresas que he experimentado.

No espero que esta réplica aplaque a mi crítico, quien tal vez responda que las sorpresas de esta naturaleza obedecen a un acto creativo mental de mi parte y no confieren crédito alguno a la máquina. Esto nos remite de regreso al argumento de la conciencia, lejos de la idea de sorpresa. Esta línea de argumentación debe considerarse cerrada, pero quizá valga la pena advertir que la apreciación de algo tan sorprendente requiere otro tanto de "actividad mental creadora", sea que el suceso sorpresivo provenga de un hombre, de un libro, de una máquina o de cualquier otra cosa.

A mi juicio, el punto de vista de que las máquinas no pueden sorprender obedece a la falacia a la que se encuentran particularmente sujetos los filósofos y los matemáticos: la suposición de que tan pronto se presenta un hecho a la mente, todas las consecuencias de ese hecho irrumpirán simultáneamente en la mente junto con él. Esta suposición resulta de gran utilidad en muchas circunstancias, pero solemos olvidar con demasiada facilidad que es falsa. Una consecuencia natural es suponer que no hay virtud en el mero cálculo de las consecuencias a partir de datos y principios generales.

El argumento de la continuidad del sistema nervioso

Ciertamente el sistema nervioso no es una máquina de estado discreto. Un pequeño error en la información acerca de las dimensiones del impulso nervioso que incide en una neurona puede marcar una gran diferencia en las dimensiones del impulso de salida. Podría argüirse que, siendo así, no podemos esperar ser capaces de imitar el comportamiento del sistema nervioso con un sistema de estado discreto.

Es cierto que una máquina de estado discreto debe ser diferente de una

máquina continua. No obstante, si nos apegamos a las condiciones del juego de la imitación, el examinador no tendría ninguna ventaja con esta diferencia. La situación se aclara más si consideramos otras máquinas continuas más sencillas. Un analizador diferencial serviría bien para nuestros propósitos. (Un analizador diferencial es cierto tipo de máquina que no es del tipo de estado discreto que se utiliza para algunos tipos de cálculo.) Algunos de ellos proporcionan sus respuestas en forma mecanográfica y por ello son adecuados para participar en el juego. Aun cuando una computadora digital no podría predecir exactamente las respuestas que daría a un problema el analizador diferencial, sí sería capaz de ofrecer el tipo correcto de respuesta. Por ejemplo, si se le pide que dé el valor de π (aproximadamente 3.1416), sería razonable seleccionar aleatoriamente entre los valores 3.12, 3.13, 3.14, 3.15, 3.16 con probabilidad de 0.05, 0.15, 0.55, 0.19, 0.06 (por ejemplo). En estas circunstancias sería muy difícil que el examinador distinguiera el analizador diferencial de la computadora digital.

El argumento de la informalidad del comportamiento

No es posible producir un conjunto de reglas que pretenda describir lo que una persona debe hacer en cada grupo de circunstancias concebible. Podría, por ejemplo, haber una regla que dictara que debemos detenernos al ver la luz roja de un semáforo y avanzar cuando la luz cambie a verde. Empero, ¿qué sucedería si por algún desperfecto ambas aparecieran al mismo tiempo? Tal vez se decidiría que lo más seguro sería detenerse. No obstante, más adelante podría surgir otra dificultad a raíz de esta decisión. Intentar proporcionar reglas de conducta que cubran cualquier eventualidad, incluso las que surjan a partir de las luces de los semáforos, parecería imposible. Concuerdo con todo esto.

A partir de lo dicho se alega que no podemos ser máquinas. Aunque temo que difícilmente le haré justicia, intentaré reproducir el argumento, el cual al parecer discurre así: "Si cada hombre contara con un conjunto definido de reglas de conducta mediante las cuales normara su vida, no sería mejor que una máquina. Sin embargo, puesto que no existen tales reglas, los hombres no pueden ser máquinas." Es evidente que el centro no está distribuido. No creo que el argumento haya sido planteado en esos términos, pero pienso que aun así, éste es el argumento que se utiliza. No obstante, puede surgir cierta confusión entre "reglas de conducta" y "leyes del comportamiento" que enturbie el asunto. Por "reglas de conducta" me refiero a preceptos como: "deténgase cuando vea la luz roja", a partir de los cuales uno puede actuar y de los cuales se está consciente. Por "leyes del comportamiento" me refiero a las leyes de la naturaleza que se aplican

al organismo humano como "si lo pellizcas, chillará". Si sustituimos "leyes de conducta mediante las cuales normara su vida "por "leyes del comportamiento que norman su vida" en el argumento citado, el centro sin distribuir dejaría de ser insalvable, ya que consideramos que no sólo es cierto que el ser normado por leyes del comportamiento implica ser algún tipo de máquina (aunque no necesariamente de estado discreto), sino también que ser una máquina implica ser normado por esas leyes. Sin embargo, no podemos convencernos a nosotros mismos tan fácilmente de la ausencia de leyes cabales del comportamiento como de la de leyes cabales de conducta. La única manera que conocemos para encontrar dichas leyes es la observación científica y sabemos con certeza que no hay circunstancia alguna en la que podamos afirmar: "Hemos buscado lo suficiente. No existen tales leyes."

Podemos demostrar de manera más concluyente que cualquier afirmación de esta naturaleza sería injustificada. Supongamos que estuviéramos seguros de encontrar esas leyes, si es que existen. Entonces, dada una máquina de estado discreto, ciertamente sería posible descubrir por observación lo suficiente acerca de ella para predecir su comportamiento futuro en un tiempo razonable, digamos unos 1 000 años. No obstante, no parece ser éste el caso. He instalado en la computadora de Manchester un pequeño programa que sólo utiliza 1 000 unidades de almacenamiento, mediante el cual la máquina responde a un número de 16 dígitos con otro en un lapso de dos segundos. Yo desafiaría a cualquiera a que a partir de estas réplicas aprendiera lo suficiente del programa para poder predecir cualquier respuesta a valores no procesados.

El argumento de la percepción extrasensorial

Supongo que el lector se encuentra familiarizado con la idea de la percepción extrasensorial y con el significado de sus cuatro manifestaciones principales: telepatía, clarividencia, precognición y psicocinesis. Estos fenómenos inquietantes parecen negar todas nuestras ideas científicas comunes. ¡Cómo nos gustaría desacreditarlos! Por desgracia, la información estadística, al menos en lo que a la telepatía se refiere, es abrumadora. Resulta muy difícil reordenar nuestras ideas para que incorporen estos nuevos hechos. Una vez que los aceptamos, no parece que nos falte mucho para creer en fantasmas y duendes. Así pues, una de las primeras ideas que desaparecerían sería la de que nuestros cuerpos se mueven sencillamente de acuerdo con las leyes conocidas de la física, y con algunas otras aún no descubiertas, pero similares.

Este argumento es a mi juicio bastante sólido. Puede replicarse que

muchas teorías científicas siguen funcionando en la práctica, pese a que se encuentren en conflicto con la percepción extrasensorial (PES). De hecho, podemos arreglárnosla muy bien si nos olvidamos de ella. Sin embargo, esto ofrece poco consuelo y sentimos temor de que el pensamiento sea precisamente el tipo de fenómeno en el que la PES resultara especialmente importante.

Un argumento más específico basado en la PES podría decir:

> Jugemos el juego de la imitación utilizando como testigos a un hombre que sea bueno para la recepción telepática y a una computadora digital. El examinador puede formular preguntas como: "¿A qué palo corresponde la baraja que tengo en la mano derecha?" Ya sea por telepatía o por clarividencia, el hombre proporciona la respuesta correcta 130 veces en 400 barajas. La máquina sólo puede adivinar al azar y quizá sólo acierte 104 veces, por lo que el examinador logra la identificación correcta.

Aquí se abre una posibilidad interesante. Supongamos que la computadora digital contiene un generador de números aleatorios. Entonces resultará natural utilizarlo para decidir las respuestas que hay que dar. No obstante, entonces el generador estará sujeto a los poderes psicocinéticos del examinador y quizá esta psicocinesis provoque que la máquina acierte con mayor frecuencia que lo esperado según un cálculo de probabilidades, así que el examinador seguiría sin poder hacer la identificación correcta. Por otra parte, el examinador podría adivinar acertadamente sin preguntar, recurriendo a la clarividencia, pues con la PES todo puede suceder.

Si se admite la telepatía, sería necesario hacer más rigurosa nuestra prueba. La situación podría considerarse análoga a la que ocurriría si el examinador estuviese hablando consigo mismo y uno de los competidores lo escuchara a través de la pared. Para llenar todos los requisitos satisfactoriamente, habría que situar a los participantes en una "habitación a prueba de telepatía".

7. MÁQUINAS QUE APRENDEN

El lector habrá anticipado que no poseo argumentos muy convincentes ni positivos para apoyar mis opiniones. Si los tuviera, no me habría esmerado en señalar las falacias de las opiniones contrarias a las mías. A continuación proporcionaré la información que poseo.

Volvamos por un instante a la objeción de Lady Lovelace, que afirmaba que la máquina sólo pude hacer lo que le decimos que haga. Podría decirse que un hombre puede "inyectar" una idea en la máquina y que ésta responderá hasta cierto punto y luego quedará inmóvil, como la cuerda de un piano a la que se ha propinado un martillazo. Otro símil sería una pila

atómica menor que el tamaño crítico; una idea inyectada corresponde a un neutrón que entra en la pila desde el exterior. Cada uno de estos neutrones producirá una cierta perturbación que, a la larga, se extingue. Sin embargo, si el tamaño de la pila se incrementa lo suficiente, es muy probable que la perturbación causada por el neutrón que entra se extienda y aumente hasta que se destruya toda la pila. ¿Existe un fenómeno correspondiente para las mentes y existe alguno para las máquinas? Efectivamente, parece que existe uno para la mente humana. La mayoría de éstas son, al parecer, "subcríticas"; es decir, corresponden en esta analogía a las pilas de tamaño subcrítico. Una idea presentada a una de estas mentes daría origen, en promedio, a por lo menos una idea como respuesta. Una proporción bastante pequeña es supercrítica. Una idea presentada a una de estas mentes podría dar origen a toda una "teoría" de ideas secundarias, terciarias y más remotas. La mente de los animales parece ser definitivamente subcrítica. Si aceptamos esta analogía preguntaremos: "¿Puede lograrse que una máquina sea supercrítica?"

La analogía de "la cáscara de cebolla" también nos es útil. Al considerar las funciones de la mente o del cerebro encontramos ciertas operaciones que pueden explicarse en términos puramente mecánicos. Lo que decimos no corresponde a la mente real: es una especie de cáscara que debemos quitar si hemos de encontrar la mente real. Empero, entonces, en lo que queda encontramos otra cáscara que hay que quitar, y así sucesivamente. Si procedemos así ¿llegaremos alguna vez a la mente "real" o, finalmente, nos toparemos con una cáscara que no tiene nada? En este último caso, toda la mente es mecánica. (Sin embargo, no sería una máquina de estado discreto. Ya hemos analizado esto.)

Los dos últimos párrafos no pretenden ser argumentos convincentes. Más bien deberían describirse como "recitaciones que tienden a producir crédito".

El único respaldo realmente satisfactorio que se puede dar a la opinión que expresamos al principio de la sección 6, sería, el que nos proporcionara el aguardar al fin del siglo y entonces realizar el experimento descrito. No obstante ¿qué podemos decir mientras tanto? ¿Qué pasos deben darse ahora para que tenga éxito el experimento?

Como ya expliqué antes, el problema es principalmente de programación. También habrá que hacer avances en la ingeniería, pero parece improbable que éstos no satisfagan los requisitos. Las estimaciones acerca de la capacidad de almacenamiento del cerebro varían entre 10^{10} y 10^{15} dígitos binarios. Yo me inclino por los valores más bajos y creo que sólo una fracción muy pequeña se utiliza para los tipos más elevados de pensamiento. Es probable que la mayoría se utilice para retener impresiones visuales. Me sorprendería que se requiriera más de 10^9 de esta capacidad

para jugar de una manera satisfactoria el juego de la imitación, si acaso contra un ciego. (Nota: la capacidad de la *Encyclopaedia Britannica*, 11a. edición, es de 2×10^9.) Una capacidad de almacenamiento de 10^7 sería una posibilidad muy real, incluso con las técnicas actuales. Es probable que no sea necesario aumentar la velocidad de operación de las máquinas. Las partes de las máquinas modernas que pueden considerarse análogas a las células nerviosas funcionan casi 1 000 veces más rápido que estas últimas, lo que proporcionaría un "margen de seguridad" que podría compensar las pérdidas de velocidad ocasionadas por diversos motivos. Nuestro problema, por consiguiente, consiste en descubrir cómo programar estas máquinas para que participen en el juego. A mi ritmo actual de trabajo, produzco cerca de 1 000 dígitos de programa al día, de modo que unos 60 trabajadores, trabajando duramente durante 50 años, podrían consumar la tarea y eso, si nada fuera a dar al bote de la basura. Sería mejor contar con un método algo más expedito.

Durante el proceso de intentar imitar la mente humana adulta inevitablemente se piensa en el proceso que la ha llevado al estado en que se encuentra. Podemos advertir tres componentes:

1) El estado inicial de la mente; digamos cuando se nace.

2) La educación a la cual se ha sometido.

3) Otra experiencia a la que se haya sometido, que no se describa como educación.

En vez de intentar producir un programa que simule la mente adulta, ¿por qué no tratar de producir uno que simule la mente del niño? Si ésta se sometiera entonces a un curso educativo adecuado se obtendría el cerebro de adulto. Supuestamente el cerebro humano es algo parecido a una libreta que se adquiere en la papelería: muy poco mecanismo y muchas hojas en blanco. (Mecanismo y escritura son casi sinónimos desde nuestro punto de vista. Nuestra esperanza es que el cerebro infantil tiene un mecanismo tan reducido que algo como él pueda programarse fácilmente. Como una primera aproximación podemos suponer que la cantidad de trabajo invertida en educación sea la misma que la que se requiere para el niño humano.

Por consiguiente, hemos dividido nuestro problema en dos partes: el programa infantil y el proceso educativo. Ambos se encuentran estrechamente relacionados. No podemos esperar que encontremos una buena máquina infantil al primer intento. Tenemos que experimentar instruyendo a una de estas máquinas y ver qué tan bien aprende. Luego podemos intentarlo con otra y ver si es mejor o peor. Existe una relación obvia entre este proceso y la evolución, mediante las identificaciones:

Estructura de la máquina infantil	=	material hereditario
Cambios en la máquina infantil	=	mutaciones
Selección natural	=	juicio del experimentador

Sin embargo, sería de esperarse que este proceso resulte más expedito que la evolución. La supervivencia del más apto es un método lento para medir ventajas. El experimentador, mediante el ejercicio de la inteligencia, debería ser capaz de acelerarlo. De igual importancia es el hecho de que las mutaciones aleatorias no restrinjan este proceso. Si puede rastrear la causa de alguna debilidad, posiblemente podrá imaginar el tipo de mutación que la mejore.

No será posible aplicar exactamente el mismo proceso de enseñanza a la máquina que a un niño normal. Por ejemplo, no se le podrán proporcionar piernas, por lo que tampoco se le podría pedir que salga y llene el balde de carbón. Es muy posible que tampoco tenga ojos. Pero aunque estas deficiencias puedan ser superadas mediante un astuto diseño de ingeniería, no podemos enviar a la escuela a esta criatura sin que los demás niños se burlen demasiado de ella. Pero alguna instrucción debe recibir. No hay que preocuparse demasiado por las piernas, ojos, etc. El ejemplo de Helen Keller muestra que la educación puede llevarse a cabo siempre que la comunicación en ambas direcciones entre maestro y alumno se establezca por alguno u otro medio.

Normalmente asociamos los castigos y las recompensas con el proceso de enseñanza. Algunas máquinas infantiles sencillas pueden construirse o programarse sobre este tipo de principio. La máquina debe construirse de tal manera que no sea probable que se repitan los sucesos que preceden brevemente a la ocurrencia de una señal de castigo, mientras que una señal de recompensa aumentaría la probabilidad de que se repitieran los sucesos que la ocasionaron. Estas definiciones no presuponen sentimiento alguno por parte de la máquina. He realizado algunos experimentos con una máquina infantil de esta índole y he logrado enseñarle algunas cosas, pero el método de enseñanza era demasiado poco ortodoxo para considerar que el experimento realmente haya tenido éxito.

El uso de castigos y recompensas puede, en el mejor de los casos, formar parte del proceso de enseñanza. Hablando a grandes rasgos, si el profesor no cuenta con otros medios para comunicarse con el alumno, la cantidad de información que éste recibe no excede el número total de recompensas y castigos aplicados. Para cuando el niño hubiera aprendido a repetir "Casablanca", probablemente estaría muy adolorido, si el texto sólo hubiera podido ser descubierto mediante la técnica de "veinte preguntas" y cada "No" tomara la forma de un golpe. Por consiguiente, es necesario contar con otros canales de comunicación "no emocionales". Si se dispone de estos canales, es posible enseñar a una máquina, mediante castigos y re-

compensas, a obedecer órdenes dadas en algún lenguaje, por ejemplo, en un lenguaje simbólico. Estas órdenes se transmitirían a través de esos canales "no emocionales". El uso de este lenguaje disminuirá entonces en gran medida el número de castigos y recompensas requeridos.

Las opiniones pueden variar en cuanto a la complejidad que resulte adecuada para la máquina infantil. Podría intentarse hacerla tan sencilla como sea posible, en congruencia con los principios generales. De manera alternativa, podría contarse con un sistema completo de inferencia lógica "integrado" en la máquina.[2] En este caso, el almacenamiento estaría ocupado en gran parte con definiciones y proposiciones. Las proposiciones tendrían varios tipos de *status*, por ejemplo, hechos bien establecidos, conjeturas, teoremas demostrados matemáticamente, enunciados provenientes de una autoridad, expresiones que pese a presentar la forma lógica de una proposición carecen de credibilidad, e incluso algunas proposiciones que podrían describirse como "imperativas". La máquina debería construirse de tal manera que tan pronto como se clasifique una proposición imperativa como "bien establecida" ocurra automáticamente la acción adecuada. Para ilustrar esto, supongamos que el profesor le dice a la máquina: "Haz tu tarea escolar ahora." Esto puede causar que el enunciado "El profesor dice: 'haz tu tarea escolar ahora'" se incluya entre los hechos bien establecidos. Otro de estos hechos podría ser: "Todo lo que el profesor dice es cierto." Si se combinan estas dos aseveraciones se podría llegar, a la larga, a que al imperativo "Haz tu tarea escolar ahora" se incluya entre los hechos bien establecidos, lo cual, por la construcción de la máquina, significará que la tarea escolar en efecto se empieza a realizar, pero el efecto es muy satisfactorio. Los procesos de inferencia que utiliza la máquina no tienen que satisfacer a los logicistas más exigentes. Podría, por ejemplo, no haber jerarquía de tipos. Pero ello no significa necesariamente que las falacias de tipo ocurran en mayor proporción que el riesgo que corremos de caernos en un precipicio. Los imperativos adecuados (expresados *dentro* de los sistemas, sin formar parte de las reglas *del* sistema) como "No uses una clase a menos que se trate de una subclase de alguna de las que haya mencionado el profesor" pueden tener un efecto similar a "No te acerques demasiado al borde del precipicio".

Los imperativos que puede obedecer una máquina carente de miembros están destinados a tener un carácter más bien intelectual (como en el ejemplo de hacer la tarea). Entre estos imperativos tendrán importancia los que regulan el orden en que se aplican las reglas del sistema lógico que se va a aplicar, ya que cuando se utiliza un sistema lógico, existe en cada etapa

[2] O mejor, "programado en (la máquina)", ya que nuestra máquina infantil se programaría en una computadora digital, pero el sistema lógico no tendría que aprenderse.

un número muy grande de pasos alternativos, cualquiera de los cuales puede aplicarse en lo que a la obediencia a las reglas del sistema lógico se refiere. Estas opciones marcan la diferencia entre un argumentador brillante y uno inepto y no entre uno correcto y uno falaz. Las proporciones que conducen a imperativos de este tipo podrían ser: "Cuando se mencione a Sócrates, utiliza el silogismo en Bárbara" o "Si un método ha demostrado ser más rápido que otro, no utilices el método más lento". Algunos de ellos pueden ser "dados por una autoridad", pero otros quizá sean producidos por la propia máquina, por inducción científica, por ejemplo.

La idea de una máquina que aprende quizá parezca paradójica a algunos lectores. ¿Cómo pueden cambiar las reglas de operación de la máquina? Deberían describir por completo cómo reaccionará la máquina cualquiera que sea su historia, independientemente de los cambios que pueda experimentar. Las reglas son, por consiguiente, casi invariables en el tiempo. Esto es muy cierto. La explicación de la paradoja es que las reglas que se modifican en el proceso de aprendizaje son de un tipo mucho menos pretencioso, que sólo exige una validez efímera. El lector podría trazar un paralelo con la Constitución de Estados Unidos.

Una característica importante de una máquina que aprende es que con frecuencia su profesor ignorará gran parte de lo que sucede en el interior, aunque sea capaz de predecir en cierta medida el comportamiento de su alumno. Su principal aplicación correspondería a la educación más reciente de una máquina derivada de una máquina infantil con un diseño (o programa) bien probado. Esto está en claro contraste con el procedimiento normal de utilizar una máquina para efectuar cómputos, pues el objetivo que se tiene entonces es obtener una imagen mental clara del estado de la máquina en cada momento de la computación. Este objetivo sólo puede lograrse con esfuerzo. El punto de vista de que "la máquina solamente puede hacer lo que sabemos cómo ordenarle que haga"[3] resulta extraño frente a esto. La mayoría de los programas que podemos introducir en la máquina ocasionarán que haga algo que puede no tener sentido alguno para nosotros o que nos parecerá un comportamiento totalmente aleatorio. El comportamiento inteligente supuestamente consiste en apartarse del comportamiento completamente disciplinado, que entraña la computación, aunque de manera sutil, sin dar lugar a conductas aleatorias o a iteraciones repetitivas sin sentido. Otro resultado importante de la preparación de nuestra máquina para su participación en el juego de la imitación mediante un proceso de enseñanza y aprendizaje es que probablemente se omita de una manera bastante natural la "falibidad humana", es decir, sin un "entrenamiento" especial. (El lector debe reconciliar

[3] Compárese con la afirmación de Lady Lovelace, la cual no incluye la palabra "sólo".

esto con el punto de vista expresado en las páginas 70-72.) Los procesos que se parecen no producen resultados 100% ciertos; si lo fueran, no podrían desaprenderse.

Sería sensato incluir un elemento aleatorio en una máquina que aprende. Un elemento aleatorio resulta bastante útil cuando se busca la solución de un problema. Supongamos, por ejemplo, que nos interesa encontrar un número entre 50 y 200 que sea igual al cuadrado de la suma de sus dígitos. Podríamos empezar con 51, luego intentar el 52, etc., hasta obtener un número que funcione. Alternativamente, podríamos seleccionar números al azar hasta obtener uno bueno. Este método tiene la ventaja de que no es necesario llevar un registro de los valores probados, pero tiene la desventaja de que puede probarse dos veces el mismo número, aunque esto no es muy importante si hay varias soluciones. El método sistemático tiene la desventaja de que puede haber un enorme bloque sin solución en la región que hay que investigar primero. Ahora bien, el proceso de aprendizaje debe considerarse como la búsqueda de una forma de comportamiento que satisfaga al profesor (o algún otro criterio). Puesto que es probable que exista un número muy grande de soluciones satisfactorias, el método aleatorio parece mejor que el sistemático. Cabe señalar que éste se utiliza en el proceso análogo de la evolución, pero en este caso, el método sistemático no es posible. ¿Cómo se podría llevar la cuenta de las distintas combinaciones genéticas que se han intentando para evitar el probarlas de nuevo?

Podríamos esperar que, con el tiempo, las máquinas lleguen a competir con el hombre en todos los campos puramente intelectuales. No obstante, ¿cuáles son las mejores para comenzar? Incluso ésta resulta una decisión difícil. Mucha gente piensa que lo mejor sería una actividad muy abstracta, como jugar ajedrez. También puede sostenerse que lo mejor sería dotar a la máquina con los mejores órganos sensoriales que el dinero pueda comprar, y luego enseñarla a comprender y a hablar inglés. Este proceso podría seguir el proceso normal de enseñanza de un niño. Se le podrían señalar cosas y nombrarlas, etc. Reitero que desconozco la respuesta correcta, pero considero que hay que intentar ambos enfoques.

Aunque nuestra visión hacia adelante es muy corta, podemos darnos cuenta de que hay mucho por hacer.

BIBLIOGRAFÍA

Church, A. (1936), "An Unsolvable Problem of Elementary Number Theory", *American J. Mathematics*, 58, pp. 345-363.

Gödel, K. (1931), "Uber Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme, I", *Monatshefte für Mathematica and Physics*, pp. 173-189.

Hartree, D. R. (1949), *Calculating Instruments and Machines*, Urbana, University of Illinois Press.

Kleene, S. C. (1935), "General Recursive Functions of Natural Numbers", *American J. Mathematics*, 57, pp. 153-157, 219-244.

Russell, B. (1945), *History of Western Philosophy*, Nueva York, Simon and Schuster.

Turing, A. M. (1937), "On Computable Numbers, with an Application to the *Entscheidungsproblem*", *Proc. London Math. Soc.*, 43, p. 544; (1936), 42, pp. 230-265.