Maquinaria computacional e Inteligencia

Alan Turing, 1950
Traductor: Cristóbal Fuentes Barassi, 2010,
Universidad de Chile.

1. El juego de la imitación.

Propongo considerar la siguiente pregunta: "¿Pueden pensar las máquinas?". Se debiera comenzar definiendo el significado de los términos 'máquina' y 'pensar'. Estas definiciones deberían ser elaboradas de manera tal que reflejen lo mejor posible el uso normal de estas palabras, pero una actitud así es peligrosa. Si el significado de las palabras 'máquina' y 'pensar' proviene del escrutinio de cómo son usadas comúnmente, se hace difícil escapar de la conclusión de que el significado y respuesta a la pregunta "¿pueden las máquinas pensar?" debiera ser buscado en una encuesta estadística, tal como la encuesta Gallup. Pero eso es absurdo. En vez de intentar una definición así, propondré reemplazar esa pregunta por otra, la cual se encuentra estrechamente relacionada y que se puede expresar en palabras relativamente poco ambiguas.

La nueva forma del problema puede ser descrita en términos de un juego, el cual llamaremos "el juego de la imitación". Se juega con 3 personas, un hombre (A), una mujer (B), y un interrogador (C) de cualquier sexo. El interrogador se encuentra en una habitación distinta a la de los otros dos participantes. El objetivo del juego para el interrogador es determinar cuál de los participantes es el hombre y cuál es la mujer. Él los identifica con las etiquetas X y Y, y al final del juego él dice si "X es A y Y es B", o "X es B e Y es A". Al interrogador se le permite hacer preguntas tanto a A como B del tipo:

C: X, ¿Sería tan amable de decirme el largo su cabello?

Ahora, suponga que X es de hecho A, por lo que A debe responder. El objetivo de A en el juego es tratar de que C haga una identificación falsa. Por lo que su respuesta podría ser:

"Mi pelo está cortado en capas, y los mechones más largos tienen unos 20 centímetros".

Para que los tonos de voz no ayuden al interrogador, las respuestas deben ser escritas, o mejor aún, tecleadas. Las condiciones ideales deberían incluir un teletipo que comunique ambas habitaciones. De manera opcional, las preguntas y respuestas podrían ser repetidas por un intermediario. El objetivo de B en el juego es ayudar al interrogador. Probablemente, la mejor estrategia para ella sea dar respuestas verdaderas. Ella puede incluir en sus respuestas cosas tales como "Yo soy la mujer, ¡no lo escuches!", pero aquello no garantizaría nada ya que el hombre podría decir cosas similares.

Ahora hacemos la pregunta: "¿qué pasaría si una máquina asume el rol de A en este juego?" ¿Discriminaría equivocadamente el interrogador con la misma frecuencia con la que lo hace cuando el juego se juega con un hombre y una mujer? Estas preguntas reemplazan la pregunta original "¿pueden las máquinas pensar?".

2. Crítica del nuevo problema.

Así como es posible preguntar "¿cuál es la respuesta para esta nueva pregunta?", uno podría preguntar "¿vale la pena investigar esta nueva

pregunta?". Esta última la investigamos sin más preámbulos, para así evitar regresiones infinitas.

El nuevo problema posee la ventaja de trazar una línea muy clara entre las capacidades físicas y las intelectuales del hombre. Ningún ingeniero o químico afirma ser capaz de producir un material que sea indistinguible de la piel humana. Es posible que en algún momento aquello se pueda hacer, pero aún suponiendo la disponibilidad de esta invención deberíamos sentir que no tiene mucho sentido en tratar de hacer más humana a una "máquina pensante" a través del revestimiento de piel artificial. La manera en la cual el problema ha sido planteado refleja esta condición, la que impide que el interrogador pueda ver o tocar a los otros competidores, o escuchar sus voces. Algunas otras ventajas del criterio que se ha propuesto pueden ser ejemplificadas con una muestra de preguntas y respuestas:

Q: Por favor escribe un soneto con el tema del Puente de Forth.
A: No cuentes conmigo para eso. Nunca pude escribir poesía.
Q: Suma 34957 y 70764.
A: (Pausa de 30 segundos y luego da la respuesta) 105621.
Q: ¿Juegas ajedrez?
A: Sí.
Q: Tengo mi K en mi K1, y ninguna otra pieza. Tú solo tienes K en K6, y R en R1. Es tu turno. ¿Qué jugada harías?
A: (Pausa de 15 segundos) R-R8 mate.

El método de la pregunta y respuesta pareciera ser adecuado para introducir casi cualquiera de los campos de estudio humano que quisiéramos incluir. No queremos sancionar a la máquina por su incapacidad de brillar en concursos de belleza, ni tampoco castigar a un hombre por perder una carrera contra un aeroplano. Las condiciones de nuestro juego hacen que estas incapacidades sean irrelevantes. Los "testigos" pueden presumir, si lo consideran aconsejable, todo lo que quieran acerca de sus encantos, fuerza o heroísmo, pero el interrogador no puede pedir demostraciones prácticas.

El juego podría quizás ser criticado en base al hecho de que las probabilidades están demasiado en contra de la máquina. Si el hombre intentase pretender ser una máquina, éste haría claramente una muy mala demostración. Lo delataría rápidamente su lentitud e inexactitud en aritmética. ¿Podría ser que las máquinas realizaran algo que pudiera ser descrito como pensar, pero que es muy distinto de lo que un hombre hace? Esta objeción es muy sólida, pero al menos se puede decir que si, a pesar de todo, una máquina puede ser construida para jugar satisfactoriamente el juego de la imitación, no necesitamos preocuparnos por esta objeción.

Es posible sugerir que cuando se juegue el "juego de la imitación", la mejor estrategia para la máquina podría ser hacer algo distinto que imitar la conducta de un hombre. Podría ser, pero creo que es poco probable que se produjera un efecto como este. En cualquier caso, no hay ninguna intención de investigar la teoría del juego, y daremos por asumido que la mejor estrategia es tratar de dar respuestas que un hombre daría de manera natural.

3. Las máquinas involucradas en el juego.

La pregunta realizada en la sección 1 no será lo suficientemente específica hasta que delimitemos qué queremos decir con la palabra "máquina". Es natural que queramos permitir el uso de cualquier técnica de ingeniería en nuestras máquinas. También queremos admitir la posibilidad que un ingeniero o un grupo de ingenieros pueda construir una máquina que funcione, pero cuya forma de operar no puede ser descrita satisfactoriamente por sus constructores debido a que ellos usan un método que fuera experimental en gran medida. Finalmente, queremos excluir de las máquinas a los hombres nacidos de manera normal. Es difícil elaborar las definiciones de manera tal que estas tres condiciones sean cumplidas. Por ejemplo, uno podría insistir que el equipo de ingenieros deba ser de un sexo únicamente, pero esto no sería realmente satisfactorio, ya que es probablemente posible crear a un individuo completo a partir de una sola célula de la piel (por ejemplo) de un hombre. Hacer aquello sería una hazaña para la técnica biológica, que además merecería grandes alabanzas, pero no nos inclinaría a pensar que sea un caso en el cual se está "construyendo una máquina pensante". Lo anterior nos lleva a abandonar el requerimiento de que cualquier técnica deba ser permitida. Estamos más que dispuestos para hacerlo considerando el hecho de que el interés actual en las "máquinas pensantes" ha sido despertado por un tipo particular de máquina, usualmente llamada "computador electrónico" o "computador digital". Siguiendo esta sugerencia, sólo permitimos que los computadores digitales tomen parte del nuestro juego.

A primera vista, esta restricción se ve muy drástica. Trataré de demostrar que no lo es en la realidad. Hacer esto necesita una explicación breve de la naturaleza y las propiedades de estos computadores.

También podría decirse que la identificación de las máquinas con los computadores digitales, tal como nuestro criterio de "pensar", sólo será insatisfactorio si (contrario a lo que yo creo) los computadores digitales en definitiva son incapaces de hacer una buena demostración en el juego.

Ya hay un gran número de computadores digitales que se encuentran funcionando, y se podría preguntar "¿por qué no tratar de experimentar inmediatamente? Sería fácil satisfacer las condiciones el juego. Se podría usar un número de interrogadores, y se podrían compilar estadísticas para demostrar con qué frecuencia se hicieron las identificaciones correctas". La respuesta breve es que no estamos preguntando si todos los computadores digitales lo harían bien en el juego, ni tampoco si los computadores disponibles en la actualidad lo harían bien, sino que si hay computadores imaginables que lo harían bien. Pero ésta es sólo una respuesta breve. Nos aproximaremos a esta pregunta desde otro ángulo después.

4. Computadores digitales.

La idea detrás de los computadores digitales podría ser explicada diciendo que estas máquinas pueden llevar a cabo cualquier operación que pudiera ser realizada por un computador humano. El computador humano debiera seguir reglas fijas; él no tiene ninguna autoridad para desviarse en nada de ellas. Podemos suponer que estas reglas se encuentran en un libro, el cual es alterado cada vez que a él se le da un nuevo trabajo. Tam-

bién posee un suministro ilimitado de papel sobre el cual realiza sus cálculos. De igual manera, él podría realizar sus multiplicaciones y sumas en una "máquina escritorio", pero eso no es importante.

Si usamos las explicaciones recién mencionadas como una definición, estaríamos en peligro de exhibir circularidad argumentativa. Evitamos esto dando un esbozo de los medios por los cuales el efecto deseado es alcanzado. Se puede considerar que un computador digital consiste en tres partes:

(i) Almacenamiento.
(ii) Unidad ejecutiva.
(iii) Control.

El almacenamiento es almacenamiento de información, y corresponde al papel que utiliza el computador humano, y corresponde tanto el papel para hacer los cálculos como al libro en el cual se encuentran impresas las reglas. En la medida en que el computador humano hace los cálculos en su cabeza, una parte del almacenamiento corresponderá a su memoria.

La unidad ejecutiva es la parte en la cual se llevan a cabo las diversas operaciones individuales involucradas en un cálculo. Qué es lo que son estas operaciones individuales variará de una máquina a otra. Usualmente, operaciones relativamente largas como "multiplique 3540675445 por 7076345687" pueden ser realizadas, pero en algunas máquinas, sólo las operaciones simples del tipo "escriba 0" son posibles.

Hemos mencionado que "el libro de reglas" que se le proporcionó al computador es reemplazado en la máquina por una parte del almacenamiento. Es entonces cuando se le denomina "tabla de instrucciones". Es el deber del Control de procurar que estas instrucciones sean obedecidas adecuadamente y en el orden correcto. El Control está construido de tal manera, que lo anterior ocurra necesariamente.

La información en el almacenamiento es usualmente dividida en paquetes de tamaño moderadamente pequeño. En una máquina, por ejemplo, un paquete podría consistir en 10 dígitos decimales. Se asignan números a las partes del almacenamiento en la que los varios paquetes de información son almacenados, de manera sistemática. Una instrucción típica podría ser:

"Sume el número almacenado en la posición 6809 con el que está en la 4302 y ubique el resultado de vuelta en esta última posición de almacenamiento."

No hay necesidad de decir que esto no ocurriría con expresiones en inglés. Sería probablemente codificado con una forma tipo 6809430217. Aquí, 17 significa cuál de las variadas operaciones posibles se realizará en los dos números en este caso la operación que se describió recién, es decir "sume el número...". Es posible notar que la instrucción requiere 10 dígitos, y así se forma un paquete de información de manera muy conveniente. El control tomará normalmente instrucciones para ser obedecidas en el orden en el que están almacenadas, pero ocasionalmente una instrucción como

"Ahora obedezca la instrucción almacenada en la posición 5606, y continúe desde ahí."

podría ocurrir, o también

"Si la posición 4505 contiene 0 obedezca a continuación la instrucción almacenada en 6707, en caso contrario, continúe normalmente."

Las instrucciones de este tipo son muy importantes porque hacen posible que una secuencia

de operaciones sea repetida una y otra vez hasta que una condición se cumpla, pero mientras la obedece, no hay instrucciones nuevas en cada repetición, sino las mismas una y otra vez. Para considerar una analogía doméstica, suponga que Madre quiere que Tommy llame al zapatero todas las mañanas mientras vaya saliendo al colegio para saber si ya le repararon sus zapatos. Ella le puede preguntar nuevamente todas las mañanas. O bien, ella puede poner un aviso de una vez por todas en la entrada, la cual él verá cuando se vaya al colegio y que dice que llame preguntando por los zapatos, y también destruir el aviso cuando vuelva si ya los tiene.

El lector debe aceptar como un hecho que los computadores digitales pueden ser construidos, y de hecho han sido construidos, de acuerdo a los principios que hemos descrito, y que de hecho pueden imitar las acciones de un computador humano con bastante similitud.

El libro de reglas que hemos descrito y que es usado por nuestro computador humano es una ficción conveniente, por supuesto. Los computadores humanos reales en realidad recuerdan lo que deben hacer. Si uno quiere hacer que una máquina imite el comportamiento de un computador humano en alguna tarea compleja, se le debe preguntar cómo lo hace, y luego traducir la respuesta a una tabla de instrucciones. Construir tablas de instrucciones es usualmente conocido como "programación". "Programar una máquina para que lleve a cabo la operación A" significa poner la tabla de instrucciones apropiada en la máquina de manera tal que la realice.

Una variante interesante sobre la idea de un computador digital es un computador digital con algún elemento aleatorio. Éstos poseen instrucciones con respecto al lanzamiento de un dado o algún proceso electrónico equivalente; una instrucción de este tipo podría ser: Lance el dado y anote el resultado en el lugar de almacenamiento 1000. Algunas veces, una máquina como ésta es descrita como si tuviera libre albedrío (aunque yo no ocuparía esta frase). Normalmente, no es posible determinar a través de la observación de una máquina si posee un elemento aleatorio, dado que un efecto similar se puede producir por un dispositivo que haga elecciones dependiendo de los decimales de π.

La mayoría de los computadores digitales en la actualidad sólo tienen un almacenamiento finito. No hay ningún problema teórico con la idea de que un computador tenga un almacenamiento ilimitado. Por supuesto, sólo una parte finita puede ser usada en cada momento. De manera similar, sólo una cantidad finita puede haber sido construida, pero podemos imaginar que se agreguen y más y más en la medida en que sea requerido. Tales computadores tienen intereses teóricos especiales y serán llamados computadores de capacidad infinita.

La idea de un computador digital es muy antigua. Charles Babbage, profesor Lucasiano de matemáticas en Cambridge desde 1828 a 1839, planeó una máquina así, la que llamó "Máquina Analítica", pero que nunca pudo terminar. Aunque Babbage poseía todas las ideas esenciales, su máquina nunca fue un prospecto muy atractivo para su época. La velocidad con la cual habría funcionado habría sido mucho más rápida que un computador humano, pero 100 veces más lenta que la máquina de Manchester. El almacenamiento era puramente mecánico, y utilizaba tarjetas y rodamientos.

El hecho de que la Máquina Analítica de Babbage fuera a ser completamente mecánica nos

ayudará a deshacernos de una superstición. Se le otorga importancia al hecho de que los computadores digitales son eléctricos, y que el sistema nervioso también es eléctrico. Dado que la máquina de Babbage no era eléctrica, y dado que los computadores digitales son en cierto sentido equivalentes, podemos ver que el uso de electricidad no tiene una importancia teórica. Por supuesto que se usa la electricidad cuando se necesita una transmisión rápida de señales, así que no es sorprendente que los encontremos en ambas conexiones. En el sistema nervioso, los fenómenos químicos son, al menos, igual de importantes que los eléctricos. En algunos computadores, el sistema de almacenamiento es básicamente acústico. Por lo tanto, el rasgo de utilizar electricidad puede ser visto como similitud superficial. Si queremos encontrar esas similitudes, debemos entonces buscar analogías matemáticas de función.

5. Universalidad de los computadores digitales.

Los computadores digitales considerados en la última sección podrían ser clasificados entre las "máquinas de estados discretos". Estas son la máquinas que se mueven usando saltos (o clicks) repentinos de un estado definido a otro. Estos estados son lo suficientemente distintos como para descartar la posibilidad de confundirlos. Estrictamente hablando, no existen tales máquinas. Todo se mueve realmente de manera continua. Pero hay muchos tipos de máquinas que pueden ser provechosamente consideradas como máquinas de estados discretos. Por ejemplo, al considerar los interruptores de un sistema de iluminación, es una ficción conveniente que cada interruptor está definitivamente prendido o apagado. Debe haber posiciones intermedias, pero para la mayoría de nuestros propósitos, nos podemos olvidar de ellos. Como ejemplo de una máquina de estados discretos podríamos considerar una rueda que gira 120 grados una vez por segundo, pero que se puede detener con una palanca externa; además, una lámpara se enciende en una de las posiciones de la rueda. La máquina puede ser descrita de manera abstracta como sigue: El estado interno de la máquina (la cual es descrita por la posición de la rueda) podría estar en 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 último estado y la señal de entrada de acuerdo a esta tabla:

		Último Estado		
		q_1	q_2	q_3
Entrada	i_0	q_2	q_3	q_1
	i_1	q_1	q_2	q_3

Las señales de salida, las únicas indicaciones visibles externamente de los estados internos (la luz), es descrita por la tabla de estados:

Estado:	q_1	q_2	q_3
Salida:	o_0	o_0	o_1

Este es un ejemplo típico de una máquina de estados discretos. Éstas pueden ser descritas por estas tablas, siempre y cuando posean un número finito de estados posibles.

Pareciera que dado el estado inicial de la máquina y las señales de entrada siempre sería posible predecir los estados futuros. Esto es una reminiscencia de la visión de Laplace que establecía

que a partir de el estado completo del universo en un momento en el tiempo, descrito por las posiciones y velocidades de todas sus partículas, debería ser posible predecir todos los estados futuros. La predicción que estamos considerando es, sin embargo, más cercana a la práctica que a la visión de Laplace. El sistema del "universo como un todo" es tal, que pequeños errores en las condiciones iniciales pueden tener un efecto inmenso en un tiempo posterior. El desplazamiento de un solo electrón en una milmillonésima de centímetro en determinado momento podría generar la diferencia entre que un hombre muera en una avalancha un año después, o que escape de ella. Es una propiedad esencial de los sistemas mecánicos que hemos llamado "máquinas de estados discretos" que este fenómeno no ocurra. Incluso cuando consideramos las máquinas físicas reales en vez de las idealizadas, el conocimiento razonablemente preciso de un estado en determinado momento produce un conocimiento razonablemente preciso luego de una cierta cantidad de pasos.

Como hemos mencionado, los computadores digitales caben dentro de la clase de máquinas de estados discretos. Pero el número de estados de los cuales una máquina es capaz es, con frecuencia, tremendamente grande. Por ejemplo, el número de estados para la máquina trabajando en Manchester es aproximadamente de $2^{165,000}$ es decir, cerca de $10^{50,000}$. Compare eso con nuestro ejemplo de la rueda antes mencionado, el cual tenía tres estados. No es difícil visualizar por qué el número de estados debiera ser tan inmensamente grande. El computador incluye un almacenamiento correspondiente a la cantidad de papel usado por un computador humano. Debe ser posible escribir en el lugar de almacenamiento cualquiera de las combinaciones de símbolos que hubieran sido escritas en papel. En términos prácticos, suponga que sólo los dígitos entre 0 y 9 son usados como símbolos. Las variaciones en la caligrafía son ignoradas. Suponga que al computador se le entregan 100 hojas de papel, con 50 líneas cada una, y a su vez, cada línea con 30 dígitos. Por lo tanto, el número de estados es $10^{100\times50\times30}$–es decir, $10^{150,000}$. Este es el número aproximado de estados de tres máquinas de Manchester juntas. El logaritmo de base 2 con el número de estados es conocido como la "capacidad de almacenamiento" de la máquina. De esta manera, la máquina de Manchester tiene una capacidad de almacenamiento de aproximadamente 165.000, y la máquina-rueda de nuestro ejemplo alrededor de 1.6. Si las dos máquinas son unidas, sus capacidades deben ser sumadas para obtener la capacidad de la máquina resultante. Esto lleva a la posibilidad de aseveraciones tales como "la máquina de Manchester contiene 64 pistas cada una con una capacidad de 2,560, ocho tubos electrónicos con una capacidad de 1,280. El almacenamiento heterogéneo alcanza a 300, alcanzando un total de 174,380.

Dada la tabla correspondiente a una máquina de estados discretos, es posible predecir qué es lo que hará. No hay razón para creer que este cálculo no pudiera ser llevado a cabo por un computador digital. Dado que podría ser llevado a cabo lo suficientemente rápido, el computador digital podría imitar el comportamiento de cualquier máquina de estados discretos. El juego de la imitación podría entonces ser jugado con la máquina en cuestión (como si fuera B), y el computador digital que imita (como A) y el interrogador sería incapaz de distinguirlos. Por supuesto, el computador digital debe tener una capacidad de almacenamiento adecuada así como también funcionar lo suficientemente rápido. Además, debe ser progra-

mado de nuevo para cada máquina que se desea que imite.

Se puede aludir a esta propiedad especial de los computadores digitales, que puedan imitar una máquina de estados discretos, diciendo que son máquinas universales. La existencia de máquinas con esta propiedad tiene la consecuencia importante de que, consideraciones de velocidad dejadas de lado, es innecesario diseñar varias máquinas nuevas para que hagan variados procesos computacionales. Todo ellos pueden ser realizados con un solo computador digital que se encuentre programado adecuadamente para cada caso. Se puede vislumbrar como consecuencia de esto que todos los computadores digitales son equivalentes en cierto sentido. Ahora podríamos considerar nuevamente el punto mencionado al final de la sección 3. Se sugirió tentativamente que la pregunta "¿pueden las máquinas pensar?" debiera ser reemplazada por "¿hay computadores digitales imaginables que tendrían un buen desempeño en el juego de la imitación?". Pero en virtud de la propiedad de universalidad podemos ver que cualquiera de estas preguntas es equivalente a lo siguiente: "Permitámonos fijar nuestra atención en un computador C en particular. ¿Sería cierto que si se modifica este computador de manera tal que exhiba un almacenamiento adecuado, un aumento en su velocidad de acción, y dado que se le entregue el programa apropiado, podría C jugar satisfactoriamente la parte de A en el juego de la imitación, mientras que B sea llevada a cabo por un hombre?".

6. Perspectivas contrarias sobre la pregunta principal.

Ahora podemos considerar que el terreno ha sido delimitado, y nos encontramos listos para proceder con el debate sobre la pregunta "¿pueden pensar las máquinas?" y la variante de esta pregunta presentada al final de la última sección. No podemos abandonar del todo la versión original de problema, dado que las opiniones variarán con respecto a la idoneidad de la substitución y debemos al menos escuchar lo que se pueda decir en conexión con esto.

Se simplificarán ciertos asuntos para el lector si primero explico mis propias ideas con respecto al tema. Consideremos la versión más precisa de la pregunta. Creo que en un periodo de tiempo de 50 años será posible programar computadores, con una capacidad de almacenamiento de alrededor de 10^9, para que puedan jugar el juego de la imitación de tal manera que el interrogador promedio no pueda obtener más de un 70 por ciento de posibilidades de hacer la identificación acertada luego de cinco minutos de preguntas. Con respecto a la pregunta original, "¿pueden las máquinas pensar?", creo que no tiene mucho sentido como para merecer discusión. No obstante, creo que cuando lleguemos a finales de siglo, el uso de las palabras y la opinión educada general habrán cambiado tanto, que uno podrá ser capaz de hablar de máquinas pensantes sin esperar ser contradicho. Creo además que ningún propósito útil se puede lograr al ocultar estas ideas. La visión popular que los científicos proceden inexorablemente desde los hechos bien establecidos hacia otros hechos bien establecidos sin nunca ser influenciados por alguna conjetura no probada es bastante equivocada. Dado que se hace claro cuá-

les son los hechos probados y cuáles son conjeturas, no puede haber ningún daño. Las conjeturas poseen una gran importancia debido a que éstas sugieren líneas útiles de investigación.

Ahora procedo a considerar opiniones opuestas a las mías.

6.1. La objeción teológica.

Pensar es una función del alma inmortal del hombre. Dios le ha otorgado un alma inmortal a cada hombre y mujer, pero no a otros animales o máquinas. Por lo tanto, ningún animal o máquina puede pensar.

Soy incapaz de aceptar ninguna parte de esto, pero intentaré responder en términos teológicos. Encontraría más convincente el argumento si los animales fueran clasificados junto con los hombres, ya que según mi criterio, hay una diferencia aún más grande entre el típico ser animado y el inanimado que entre el hombre y los otros animales. El carácter arbitrario de la visión ortodoxa se hace más claro si consideramos cómo podría parecer para el miembro de otra comunidad religiosa. ¿Cómo consideran los cristianos la visión musulmana de que las mujeres no tienen alma? Pero dejemos este punto de lado y volvamos a la discusión central. Me parece que el argumento antes mencionado implica una restricción seria a la omnipotencia del Todopoderoso. Es admisible que haya cosas que Él no pueda hacer, tal como igualar uno con dos, pero ¿no deberíamos creer que Él tiene la libertad de conferir un alma a un elefante si a Él le parece? Podríamos esperar que Él usaría este poder sólo en conjunción con una mutación que provea al elefante con un cerebro apropiadamente mejorado para suministrar las necesidades de esta alma. Una discusión con la misma forma podría hacerse en el caso de las máquinas. Podría parecer distinta, porque es más difícil de "tragar". Pero esto sólo significa que pensamos que sería menos probable que Él pudiera considerar apropiadas estas circunstancias para conferir un alma. Las circunstancias en cuestión son discutidas en el resto de este artículo. En el intento de construir estas máquinas, no deberíamos estar usurpando irreverentemente su poder para crear almas, en mayor medida que cuando procreamos niños: más bien, en cada caso somos instrumentos de su voluntad al proveer mansiones para las almas que Él crea.

Sin embargo, esto es mera especulación. No me impresionan mucho las discusiones teológicas, sea cual sea el tema. Tales discusiones han sido frecuentemente insatisfactorias en el pasado. En los tiempos de Galileo, se discutía que los textos "Y el sol se quedó quieto... y consideró no bajar por un día" (Josué Cap. 10, v.13) y "Él puso los cimientos de la tierra, para que no se moviera en ningún momento" (Salmos Cap. 105, v.5) eran una refutación adecuada para la teoría de Copérnico. Con nuestro actual conocimiento, tal argumento parece fútil. Cuando ese conocimiento no estaba disponible, causaba una impresión muy diferente.

6.2. La objeción de las "cabezas en la arena".

"La consecuencia de que las máquinas piensen sería demasiado espantosa. Esperemos y creamos que no lo pueden hacer".

Este argumento es rara vez expresado tan abiertamente como recién se plantea. Pero afecta a la mayoría de los que pensamos en ello. Nos gusta creer que el hombre es sutilmente superior al res-

to de la creación. Es mejor si se puede demostrar que es necesariamente superior, ya que entonces no hay peligro de que pierda su posición de mando. La popularidad del argumento teológico está claramente conectada con esta sensación. Es probable que sea bastante fuerte entre los intelectuales, dado que ellos valoran el poder de pensar más que otros, y están más inclinados a basar su creencia en la superioridad del Hombre en base a este poder.

No creo que este argumento sea lo suficientemente sustancial como para requerir refutación. El consuelo sería más apropiado; quizás eso debiera buscarse en la trasmigración de las almas.

6.3. La objeción Matemática.

Hay un varios resultados en la lógica matemática que puede ser usado para mostrar que existen limitaciones para los poderes de una máquina de estados discretos. El más conocido de estos resultados es el teorema de Gödel (1931), y demuestra que en cualquier sistema lógico lo suficientemente poderoso se pueden formular aseveraciones que no se pueden ni probar ni desaprobar dentro del sistema, a menos que el sistema en sí sea inconsistente. Algunos resultados similares en algunos aspectos se encuentran en el trabajo de Church (1936), Kleene (1935), Rosser (1936), y Turing (1937). El último resultado es el más conveniente de considerar, dado que refiere directamente a las máquinas mientras que otros sólo pueden ser usados comparativamente en discusiones indirectas; por ejemplo, si el teorema de Gödel es usado, necesitamos tener además algún medio para describir sistemas lógicos en términos de máquinas, y máquinas en términos de sistemas lógicos. El resultado en cuestión se refiere a un tipo de máquina que es esencialmente un computador digital con una capacidad infinita. Establece que hay ciertas cosas que una máquina con estas características no puede hacer. Si se le arma para que de respuestas a preguntas como las del juego de la imitación, habrá algunas preguntas a las cuales responderá erróneamente, o simplemente no responderá, no importa cuánto tiempo se le de para responder. Por supuesto, podría haber muchas de esas preguntas, y preguntas que no podrían ser respondidas por una máquina, podrían ser respondidas satisfactoriamente por otra. Estamos suponiendo en este caso que las preguntas son del tipo "si" y "no", y no del tipo "¿qué piensas sobre Picasso?". Las preguntas que las máquinas deberían fallar son del tipo "considere la máquina especificada de la siguiente manera... ¿podrá esta máquina responder alguna vez 'si' a alguna pregunta?". Los puntos deben ser reemplazados por la descripción habitual de una máquina, que podría ser algo así como la usada en la sección 5. Cuando la máquina descrita posee cierta relación relativamente simple con la máquina que esta siendo interrogada, se puede demostrar que la respuesta es equivocada o no disponible. Este es el resultado matemático: se argumenta que hay una discapacidad en las máquinas que el intelecto humano no posee.

La respuesta breve a este argumento es que, aunque se establece que hay limitaciones para el poder de cualquier máquina, solo se ha dicho, sin ningún tipo de prueba, que tales limitaciones no se aplican al intelecto humano. Pero no creo que esta visión pueda ser descartada a la ligera. Siempre que se le haga la respuesta crítica apropiada a una de estas máquinas, y den una respuesta definitiva, sabemos que esta respuesta debe estar equivocada, y eso nos da cierta sensación de superioridad. ¿Es esta sensación ilusoria? No ca-

be duda que es genuina, pero no creo que se le deba dar demasiada importancia. Con bastante frecuencia respondemos equivocadamente como para justificar algún tipo de satisfacción por tener evidencia de la falibilidad de las máquinas. Además, nuestra superioridad sólo se puede sentir en tales ocasiones en relación con la máquina sobre la que nos anotamos esa victoria pírrica. No habría la menor oportunidad de triunfar simultáneamente sobre todas las máquinas. En pocas palabras, puede que haya hombres más inteligentes que cualquier máquina dada, pero de nuevo, puede que haya otras máquinas aún más inteligentes, y así consecutivamente.

Creo que aquéllos que se aferran a la discusión matemática deberían estar dispuestos a aceptar el juego de la imitación como una base para la discusión. Aquéllos que creen en las dos primeras objeciones probablemente no estén interesados en ningún criterio.

6.4. El argumento desde la conciencia.

Este argumento se encuentra muy bien expresado en la disertación de la Medalla de Lister del profesor Jefferson en 1949, de donde cito: "Hasta que una máquina pueda escribir un soneto o componer un concierto debido a las emociones y pensamientos que tuvo, y que no sea debido al uso de símbolos al azar, podremos estar de acuerdo que máquina es igual a cerebro es decir, no sólo que lo escriba, sino saber que lo escribió. Ningún mecanismo podría sentir (y no sólo una mera señal artificial, cosa fácil de hacer) placer por sus éxitos, sentir pesar cuando se le funde una válvula, sentirse bien con un halago, sentirse miserable por sus errores, estar encantado por el sexo, estar enojado o deprimido cuando no consigue lo que quiere."

Este argumento parece ser una negación a la validez de nuestra prueba. De acuerdo a la forma más extrema de esta visión, la única manera con la cual uno podría estar seguro de que una máquina piensa es ser la máquina y sentir su propio pensamiento. Por tanto, se podría describir estas sensaciones al mundo, pero por supuesto nadie estaría justificado de poder percatarse. De manera similar, de acuerdo a esta visión, la única manera de saber que un hombre piensa es ser ese hombre en particular. Éste es el punto de vista de solipsista. Esa sería la visión más lógica de sostener, pero hace difícil la comunicación de las ideas. A está inclinado a creer "A piensa pero B no", mientras que B cree "B piensa pero A no". En vez de argumentar continuamente contra este punto, lo usual es tener la convención educada de que todos pensamos.

Estoy seguro de que el profesor Jefferson no quiere adoptar el punto de vista extremista del solipsismo. Probablemente él se encontraría dispuesto a aceptar el juego de la imitación como una prueba. El juego (con el jugador B omitido) es usado frecuentemente en la práctica con el nombre de *viva voce* para descubrir si alguien realmente entiende algo o ha "aprendido como perico". Escuchemos una parte de un *viva voce*:

Interrogador: En la primera línea de su soneto, el cual dice "deberíais compararles con un día de verano", ¿no sería igual o mejor decir "un día de primavera"?

Testigo: no tendría la métrica correcta.

Interrogador: y si usamos "un día de invierno". Ese sí la tendría.

Testigo: sí, pero nadie quiere ser comparado con un día de invierno.

Interrogador: ¿se podría decir que el Sr. Pickwick le recuerda la navidad?

Testigo: de cierta manera.

Interrogador: pero la navidad es un día de invierno, y no creo que al Sr. Pickwick le importe la comparación.

Testigo: no creo que estés hablando en serio. Cuando uno dice día de invierno se refiere a un día típico de invierno más que a uno especial como la navidad.

Y se podría continuar. ¿Qué diría el profesor Jefferson si la máquina escritora de sonetos fuera capaz de responder de esta manera en el *viva voce*? No sé si él consideraría a la máquina como si estuviera mandando "meras señales artificiales" para responder, pero si las respuestas fueran satisfactorias y sostenidas como en el pasaje anterior, no creo que las describiría como un "cosa fácil de hacer". Creo que esta frase tiene por intención cubrir a dichos dispositivos en los cuales a una máquina se le incluye la grabación de alguien que lee un soneto, con un interruptor apropiado para prenderlo de cuando en cuando.

En pocas palabras, creo que aquellos que apoyan el argumento de la conciencia podrían ser persuadidos de abandonarlo en vez de forzar una posición solipsista. Probablemente, estarán dispuestos a aceptar nuestra prueba.

No quisiera dar la impresión de que creo no hay un misterio con respecto a la conciencia. Hay, por ejemplo, algo así como una paradoja en conexión con cualquier intento de localizarla. Pero no creo que estos misterios deban ser resueltos antes de que podamos responder la pregunta que concierne a este artículo.

6.5. Argumentos desde las discapacidades múltiples.

Estos argumentos toman la forma "Te aseguro que puedes hacer máquinas que hagan todas las cosas que dices, pero nunca podrás hacer una que sea capaz de X". Varias características de X se sugieren en relación a esto. Yo mismo ofrezco una selección:

Ser amable, ingenioso, hermoso, amigable, tener iniciativa, tener sentido del humor, diferenciar lo correcto de lo incorrecto, cometer errores, enamorarse, disfrutar las fresas con crema, hacer que alguien se enamore de él, aprender de la experiencia, usar las palabras apropiadamente, ser sujeto de sus propios pensamientos, tener tanta diversidad de conductas como un hombre, hacer algo realmente novedoso.

Usualmente no se ofrece ningún soporte a estos enunciados. Creo que la mayoría se basan en el principio de inducción científica. Un hombre ve miles de máquinas en su vida. Y de acuerdo a lo que ve, saca una cantidad de conclusiones generales. Son feas, se diseñan con un propósito muy limitado, y cuando se les requiere para una cosa ligeramente diferente, no sirven, la variedad de conductas es muy poca, y así sucesivamente. Naturalmente, un hombre concluye que estas son en general propiedades necesarias de las máquinas. Muchas de estas limitaciones se encuentran asociadas a la poca capacidad de almacenamiento de la mayoría de ellas (asumo que la idea de capacidad de almacenamiento incluye de alguna manera a otras máquinas distintas a las máquinas de estados discretos. La definición exacta no importa dado que no se sostiene ninguna precisión matemática en esta discusión). Algunos años atrás, cuando no se sabía mucho sobre los compu-

tadores digitales, había mucha incredulidad con respecto a ellos, si es que alguien mencionaba sus propiedades sin describir su construcción. Presumiblemente, eso se debía a una aplicación similar del principio de inducción científica. Estas aplicaciones del principio son, por supuesto, inconscientes en su mayoría. Cuando un niño con quemaduras le teme al fuego y demuestra su miedo evitándolo, yo podría decir que él está aplicando inducción científica. (Por supuesto podría describir su conducta de muchas otras maneras). Los trabajos y costumbres de la humanidad no parecen ser material apropiado sobre el cual aplicar la inducción científica. Una gran parte del espacio-tiempo debe ser investigada si se busca obtener resultados confiables. De otra manera, podríamos (como la mayoría de los niños ingleses lo hacen) decidir que todo el mundo habla inglés, y que por lo tanto es tonto aprender francés.

Sin embargo, se pueden hacer comentarios especiales acerca de las muchas discapacidades que han sido mencionadas. La incapacidad de disfrutar las fresas con crema podría sorprender al lector dada su frivolidad. Posiblemente, se podría hacer una máquina que disfrute este delicioso plato, pero cualquier intento de crearla sería estúpido. Lo que es importante con respecto a esta discapacidad es que contribuye a otras discapacidades, por ejemplo, a la dificultad del mismo tipo de amabilidad entre hombre y máquinas como entre hombre blanco con hombre blanco, o entre hombre negro y hombre negro.

La declaración de que "las máquinas no pueden cometer errores" parece curiosa. Uno es tentado a responder "¿son peores por eso?". Pero adoptemos una actitud un poco más comprensiva, y veamos que quiere decir realmente. Creo que esta crítica puede ser explicada en términos del juego de la imitación. Se sostiene que interrogador podría distinguir simplemente la máquina del hombre al plantearle una cantidad de problemas aritméticos. La máquina sería desenmascarada dado su altísima eficacia. La respuesta a esto es simple. La máquina (programada para jugar el juego) no intentaría dar la respuesta correcta a los problemas aritméticos. Deliberadamente, cometería errores de manera calculada para confundir al interrogador. Una falla mecánica podría probablemente delatar a la máquina con una decisión poco apropiada con respecto al tipo de error que cometa en el cálculo. Incluso esta interpretación de la crítica no es lo suficientemente comprensiva. Pero no podemos dedicarle el espacio para ahondar en ello. Me parece que esta crítica se sostiene en una confusión entre dos tipos de errores. Podríamos etiquetarlos como "errores de funcionamiento" y "errores de conclusión". Los errores de funcionamiento se deben a alguna falla mecánica o eléctrica que produce que la máquina se comporte de una manera distinta con respecto a la cual fue diseñada. En la discusión filosófica a uno le gusta ignorar la posibilidad de tales errores; uno se encuentra por lo tanto discutiendo sobre "máquinas abstractas". Estas máquinas abstractas son ficciones matemáticas más que objetos físicos. Por definición, incapaces de presentar errores de funcionamiento. En este sentido, podemos realmente decir que "las máquinas nunca cometen errores". Errores de conclusión sólo se pueden producir cuando se adjunta algún significado a las señales de salida de la máquina. La máquina podría, por ejemplo, escribir ecuaciones matemáticas, u oraciones en inglés. Cuando escribe una proposición falsa, decimos que la máquina ha cometido un error de conclusión. Claramente, no hay razón para decir que la máquina no pueda cometer este tipo de error. Podría solamente escribir repetidas veces "0 = 1". Al considerar un ejemplo menos rebuscado, se podría

tener algún método para dar conclusiones a través de la inducción científica. Debemos esperar que un método así nos lleve a resultados erróneos ocasionalmente.

La aseveración de que una máquina no puede ser objeto de su propio pensamiento sólo puede ser respondida si se puede mostrar que la máquina posee algún pensamiento con algún objeto. No obstante, "el objeto de las operaciones de una máquina" pareciera significar algo, al menos para las personas que tratan con ella. Si, por ejemplo, la máquina estuviera tratando de encontrar una solución para la ecuación $x^2 - 40x - 11 = 0$, uno se sentiría tentado a describir estas ecuación como parte del objeto del pensamiento de la máquina en ese momento. En este sentido, la máquina puede sin lugar a dudas ser el objeto de su propio pensamiento. Podría ser usada para ayudar a crear sus propios programas, o para predecir el efecto de las alteraciones en su propia estructura. A través de la observación de resultados de su propia conducta, podría modificar sus programas de manera tal de alcanzar algún propósito de manera más efectiva. Éstas son más bien posibilidades para el futuro cercano que sueños utópicos.

La crítica que refiere a que una máquina no puede tener una gran variedad de conductas es sólo una manera de decir que no puede tener una gran capacidad de almacenamiento. Hasta hace poco, la capacidad de almacenamiento de incluso mil dígitos era muy rara.

Las críticas que estamos considerando acá son con frecuencia formas disfrazadas del argumento desde la conciencia. Generalmente, si uno sostiene que una máquina puede hacer una de estas cosas, y describe el tipo de método que la máquina podría usar, uno no produciría una gran impresión. Se cree que el método (cualquiera que sea, dado que debe ser mecánico) es realmente deshonesto. Compare con el paréntesis de la afirmación de Jefferson mencionada más arriba.

6.6. La objeción de Lady Lovelace.

La información más detallada de la Máquina Analítica de Babbage proviene de una de las memorias de Lady Lovelace (1842). En ésta, ella sostiene que "la Máquina Analítica no tiene pretensiones de *originar* nada. Puede hacer *cualquier cosa que sepamos ordenarle* que haga" (su cursiva). Esta declaración es citada por Hartree, que añade: "esto no implica que no sea posible construir un equipamiento electrónico que podrá 'pensar por sí mismo', o sobre el cual, en términos biológicos, uno pudiera construir un reflejo condicionado, que podría servir como la base del 'aprendizaje'. Si es que esto es posible en principio o no es una pregunta estimulante y apasionante, y es sugerida debido a algunos de los desarrollos recientes. Pero no parece que las máquinas construidas o proyectadas en ese momento pudieran tener esta propiedad".

Estoy completamente de acuerdo con Hartree. Uno podrá notar que él no asevera que las máquinas en cuestión no tuvieran la propiedad, sino que la evidencia disponible para Lady Lovelace no la alentaba a creer que la tuvieran. Es muy posible que las máquinas en cuestión tuvieran esta propiedad en cierto sentido. Suponga que alguna máquina de estados discretos tiene la propiedad. La Máquina Analítica era un computador universal digital, así que, si la capacidad de almacenamiento y velocidad fueran adecuadas, se podría a través de una programación apropiada hacer que se imitara a la máquina en cuestión. Probablemente este argumento no se le ocurrió

ni a la Condesa ni a Babbage. En cualquier caso, no tenían ninguna obligación de afirmar todo lo que se puede afirmar.

Toda esta pregunta será considerada nuevamente bajo la perspectiva de las máquinas que aprenden.

Una variante de la objeción de Lady Lovelace afirma que una máquina "nunca hace nada realmente nuevo". Esto podría ser aludido desde la perspectiva de "no hay nada nuevo bajo el sol". Quién puede tener certeza que el "trabajo original" que alguien haya hecho no fue solamente el crecimiento de una semilla plantada en él a través de la enseñanza, o el efecto de seguir principios generales bien sabidos. Una variante mejor a la objeción dice que una máquina nunca puede "sorprendernos". Esta declaración es un desafío más directo y puede ser enfrentada más directamente. Las máquinas me sorprenden con gran frecuencia. Esto se debe en gran medida a que no hago el cálculo suficiente para decidir qué puedo esperar de ellas, o más bien porque, aunque hago una estimación, lo hago apurado, con descuido, tomando riesgos. Quizás me digo a mi mismo, "Y creería que el voltaje acá debiera ser el mismo que allá: bueno, supongamos que así es". Naturalmente, con frecuencia me equivoco, y el resultado es una sorpresa, pues para cuando el experimento se lleva a cabo, estos supuestos ya se han olvidado. Reconocer lo anterior me deja expuesto a críticas sobre mis modos descuidados de proceder, pero no arrojan ninguna duda sobre mi credibilidad cuando testifico las sorpresas que experimento.

No espero que esta respuesta silencie a mi crítico. Él probablemente dirá que esas sorpresas se deben a algún acto mental creativo de mi parte, y que no otorga ningún crédito a la máquina. Esto nos lleva de vuelta a la discusión sobre la conciencia, y lejos de la idea de la sorpresa. Esta es una línea argumental que debemos considerar cerrada, pero quizás valga la pena destacar que la apreciación de algo como sorpresivo requerirá un "acto mental creativo", sin importar si el evento sorpresivo es originado por un hombre, un libro, una máquina o cualquier otra cosa.

La idea de que las máquinas no producen sorpresas se debe, creo yo, a una falacia a la cual se encuentran especialmente sujetos los filósofos y los matemáticos. Es el supuesto de que tan pronto como se presente un hecho a una mente, todas las consecuencias de ese hecho florecen en ella simultáneamente con el hecho. Es un supuesto muy útil en muchas circunstancias, pero uno olvida demasiado fácilmente que es falso. Una consecuencia natural de hacer esto es que uno asume que no hay mérito en la simple búsqueda de consecuencias a partir de datos y principios generales.

6.7. Argumento desde la continuidad en el sistema nervioso.

Con toda certeza, el sistema nervioso no es una máquina de estados discretos. Un pequeño error en la información sobre el tamaño del impulso nervioso que afecte a una neurona, podría hacer una gran diferencia en el tamaño del impulso saliente. Se podría sostener que, siendo esto así, no se puede esperar sea posible imitar la conducta del sistema nervioso con un sistema de estados discretos.

Es verdad que una máquina de estados discretos debe ser diferente de una máquina continua. Pero si nos adherimos a las condiciones del juego de la imitación, el interrogador no será capaz sacar

provecho de esta diferencia. La situación se puede hacer más clara si consideramos a otras máquinas continuas más simples. Un analizador diferencial servirá bien. (Un analizador diferencial en un tipo de máquina, que no es del tipo de estados discretos, y que se usa para algunos tipos de cálculos). Algunos de éstos generan sus respuestas de manera escrita, por lo que son adecuadas para formar parte del juego. No sería posible para un computador digital predecir exactamente qué respuestas daría el analizador diferencial para determinado problema, pero sería muy capaz de dar el tipo correcto de respuesta. Por ejemplo, si se le pidiera que de el valor de π (alrededor de 3.1416), sería razonable elegir al azar entre valores 3.12, 3.13, 3.14, 3.15, 3.16 con probabilidades de 0.05, 0.15, 0.55, 0.19, 0.06 (por ejemplo). Bajo estas circunstancias, sería muy difícil para el interrogador distinguir entre el analizador diferencial y el computador digital.

6.8. El argumento desde la informalidad de la conducta.

No es posible producir una lista de reglas que pretendan describir lo que un hombre debiera hacer en cada circunstancia concebible. Uno podría, por ejemplo, tener una regla en la que se debe parar cuando alguien ve un semáforo en rojo, y de seguir si alguien ve la luz verde; pero, ¿qué pasa si por alguna falla, ambas luces aparecen juntas? Uno quizás decidiría detenerse ya que es más seguro. Pero una nueva dificultad bien podría producirse más tarde por esta decisión. El intento de proveer reglas de conducta para cubrir cada eventualidad, incluso aquellas que se producen por los semáforos, parece imposible. Estoy de acuerdo con todo esto.

Dado lo anterior, se dice que no podemos ser máquinas. Trataré de reproducir el argumento, pero me temo que no podré hacerle justicia. Pareciera que es algo así: "Si cada hombre tuviera un grupo determinado de reglas de conducta por las cuales él regulara su vida, no sería más que una máquina. Pero no existen tales reglas, así que los hombres no pueden ser máquinas". La falacia argumental es flagrante. No creo que el argumento se haya propuesto así alguna vez, no obstante creo que ese es el argumento que se da. Sin embargo, podría hacer una cierta confusión entre las "reglas de conducta" y las "leyes de comportamiento". Se entiende por "reglas de conducta" los preceptos tales como "detenerse si uno ve una luz roja", sobre las cuales alguien puede actuar, y de las cuales se está consciente. Por "leyes de comportamiento" me refiero a las leyes de la naturaleza aplicables al cuerpo de un hombre, tales como "si lo pinchas, va a chillar". Si substituimos "reglas de conducta que regulan su vida" por "leyes de comportamiento que regulan su vida" en el argumento citado con anterioridad, la falacia deja de serlo. Pues no solamente creemos que ser regulados por leyes de comportamiento implica ser cierto tipo de máquina (aunque no necesariamente una máquina de estados discretos), sino que recíprocamente, ser tal máquina implica ser regulado por tales leyes. Sin embargo, no podemos convencernos tan fácilmente de la ausencia de leyes completas del comportamiento, como de la ausencia de reglas completas de la conducta. La única manera que sabemos para descubrir tales leyes es la observación científica, y ciertamente conocemos de circunstancia alguna en la que podamos decir: "hemos buscado lo suficiente. No existen tales leyes".

Podemos demostrar más convincentemente que cualquier declaración de este tipo sería injustifi-

cada: Suponga que estuviéramos seguros de encontrar tales leyes cuando existieran. Entonces, dada una máquina de estados discretos, sería con certeza posible descubrir tales leyes a través suficientes observaciones de ella, y así predecir su conducta futura, y todo esto dentro de un tiempo razonable, digamos, mil años. Pero este no parece ser el caso. He instalado en el computador de Manchester un programa pequeño usando solo 1000 unidades de almacenamiento, tal que cuando a la máquina se le provee con un número de 16 dígitos, responde con otro número de 16 dígitos en dos segundos. Desafiaría a cualquiera a aprender lo suficiente de las respuestas del programa para ser capaz de predecir cualquier respuesta a valores no probados con anterioridad.

6.9. El argumento desde la percepción extra-sensorial.

Asumo que el lector se encuentra familiarizado con la idea de percepción extra-sensorial, y del significado de sus cuatro términos, que son telepatía, clarividencia, precognición y psicoquinesia. Estos fenómenos perturbadores parecieran ir en contra de todas nuestras ideas científicas usuales. ¡Cómo nos gustaría desacreditarlos! Desafortunadamente, la evidencia estadística, al menos para la telepatía, es abrumadora. Es muy difícil reordenar nuestras ideas de manera tal de que encajen con estos datos nuevos. Una vez que uno las ha aceptado, no parece un gran avance creer en fantasmas y cocos. La idea de que nuestros cuerpos se mueven simplemente debido a las reglas físicas conocidas, junto a otras que aún no han sido descubiertas pero que sin embargo son similares, sería una de las primeras en ser abandonadas.

Para mi criterio, este argumento es sólido. Uno puede fácilmente responder que muchas teorías científicas parecen ser útiles en la práctica, a pesar del conflicto con la P.E.S.; de hecho, uno puede olvidarse de ello sin mayor problema. Esto es consuelo muy pobre, y me temo que el pensar sea el tipo de fenómeno donde la P.E.S. podría ser especialmente relevante.

Un argumento más específico basado en la P.E.S. podría ser: "Déjanos jugar el juego de la imitación, usando como testigo un hombre que es un buen receptor telepático, y un computador digital. El interrogador puede preguntar cosas tales como '¿qué figura tiene la carta que tengo en mi mano derecha?'. El hombre, a través de telepatía o clarividencia, responde correctamente 130 veces de un total de 400 cartas. La máquina sólo puede adivinar azarosamente, y obtiene 104 respuestas correctas, por lo que el interrogador realiza la identificación correcta". Hay una posibilidad interesante que se abre aquí. Suponga que el computador digital contiene un generador de números aleatorios. Entonces, será natural usar esto para decidir qué respuesta dar. Pero entonces el generador de números aleatorios estará sujeto a los poderes psicoquinéticos del interrogador. Quizás esta psicoquinesia podría hacer que la máquina adivine correctamente con más frecuencia de la esperada por un cálculo de probabilidad, por lo que el interrogador sería incapaz aún de realizar la identificación correcta. Por otro lado, él podría ser capaz de adivinar correctamente sin hacer preguntas, sino a través de la clarividencia. Con la P.E.S. cualquier cosa podría pasar.

Si se admite la telepatía, será necesario afinar nuestra prueba. La situación podría ser considerada como análoga a lo que ocurriría si el interrogador estuviera hablando consigo mismo en

voz alta y uno de los competidores estuviera escuchando con su oído en la pared. Poner a los competidores en una "habitación a prueba de telepatía" bastaría para cumplir todos los requerimientos.

7. Máquinas que aprenden.

El lector habrá anticipado que no tengo argumentos de naturaleza positiva muy convincentes para apoyar mis puntos de vista. Si los tuviera, no me habría tomado el trabajo de destacar las falacias de perspectivas contrarias. Ahora presentaré la evidencia que tengo.

Volvamos por un momento a la objeción de Lady Lovelace, la cual afirma que la máquina sólo puede hacer lo que le digamos que haga. Uno podría decir que un hombre puede "inyectar" una idea en la máquina, y que esta responderá hasta cierto punto y luego caerá en un estado de quietud, tal como la cuerda de un piano cuando la golpea el martillo. Otra comparación sería una carga atómica cuyo tamaño es menor al tamaño crítico: una idea inyectada correspondería al neutrón que entra desde fuera. Tal neutrón causará cierta alteración que eventualmente se acabará. Sin embargo, si el tamaño de la carga atómica es lo suficientemente incrementada, la alteración causada por la entrada de ese neutrón muy probablemente continuará sucesivamente, aumentando hasta que la carga se destruya. ¿Existe algún fenómeno análogo para las mentes y existe uno análogo para las máquinas? Al parecer sí hay uno para la mente humana. La mayoría de ellas parecen ser "subcríticas", es decir, en la analogía corresponderían a cargas de tamaño subcrítico. La idea que se le presenta a una mente así generará en promedio menos que una idea en respuesta. Una pequeña proporción de ellas son supercríticas. Una idea presentada a una de tales mentes podría generar toda una nueva "teoría" que contenga ideas secundarias, terciarias, y otras más remotas. Las mentes de los animales parecieran ser absolutamente subcríticas. En relación a esta analogía se puede preguntar, "¿se puede hacer una máquina para que sea supercrítica?"

La analogía de las "capas de cebolla" también es útil. Al considerar las funciones de la mente o el cerebro, encontramos ciertas operaciones que se pueden explicar en términos puramente mecánicos. Decimos que esto no corresponde a la mente real: es un cierto tipo de capa que debemos sacar si queremos encontrar la mente real. Pero luego en lo que queda, encontramos otra capa que se puede remover, y así sucesivamente. Si se procede de esta manera, ¿podemos alcanzar en algún momento la mente real, o eventualmente llegamos a una capa que no contiene nada? En este último caso, toda la mente sería mecánica. (Sin embargo, no sería una máquina de estados discretos. Ya discutimos esto.)

Los dos últimos párrafos no aseguran ser necesariamente argumentos convincentes. Deberían en vez ser descritos como "recitados que tienden a producir creencia".

El único fundamento realmente satisfactorio que se puede dar para la visión expresada al principio de la sección 6 será que, debemos esperar el final de siglo y recién entonces, hacer el experimento descrito. Pero, ¿qué podemos decir mientras tanto? ¿Qué pasos deberían darse para que el experimento sea exitoso?

Como he explicado, el problema es esencialmente un problema de programación. También tendrán que haber avances en la ingeniería, pero parece improbable que estos no vayan a ser adecuados

para los requerimientos. Las estimaciones para la capacidad de almacenamiento del cerebro varían entre 10^{10} a 10^{15} dígitos binarios. Yo me inclino por los valores más bajos y creo que sólo una pequeña fracción se usa en los tipos de pensamiento superior. La mayor parte se usa probablemente para la retención de impresiones visuales. Me sorprendería si más de 10^9 se requiera para jugar satisfactoriamente el juego de la imitación, al menos, si se juega contra un ciego. (Nota: la capacidad de la Enciclopedia Britannica, undécima edición, es 2×10^9.) Una capacidad de almacenamiento de 10^7 sería una posibilidad factible incluso con las técnicas actuales. Es probable que no sea necesario incrementar la velocidad de las operaciones de las máquinas en lo absoluto. Las partes de las máquinas modernas, que podrían ser consideradas como análogas a células nerviosas, trabajan cerca de mil veces más rápido que estas últimas. Esto debiera otorgar un "margen de seguridad" el cual pudiera cubrir pérdidas de velocidad que se produzcan de muchas maneras. Por lo tanto, nuestro problema es descubrir cómo programar estas máquinas para jugar el juego. En mi actual tasa de trabajo, produzco cerca de mil dígitos de programa al día, por lo que cerca de sesenta trabajadores, que trabajen regularmente a través de cincuenta años podrían lograr el objetivo, si es que nada se debe desechar. Algún método más expedito parece deseable.

Durante el proceso de intento de imitación de la mente de un humano adulto tendemos a pensar bastante sobre el proceso que produjo el estado en el que se encuentra. Podríamos mencionar tres componentes:

(a) El estado inicial de la mente, digamos al momento de nacer;
(b) La educación a la cual fue sujeta;
(c) Otra experiencia, no descrita como educación, a la que haya sido sujeta.

En vez de tratar de producir un programa similar a la mente adulta, ¿por qué no tratar en vez de producir una que simule la mente de un niño? Si ésta fuera luego sujeta al curso apropiado de educación, uno obtendría el cerebro adulto. Presumiblemente, el cerebro de un niño es algo así como un cuaderno que uno compra en una tienda. Un mecanismo más bien simple, con muchas hojas en blanco. (Mecanismo y escritura son casi sinónimos para nuestro punto de vista.) Nuestra esperanza es que haya tan poco mecanismo en el cerebro del niño, que algo así pueda ser programado fácilmente. Podemos asumir en una primera aproximación que la cantidad de trabajo en la educación sería muy similar a la de un niño humano.

Hemos por tanto dividido nuestro problema en dos partes: el programa - niño y el proceso educativo. Éstos permanecen estrechamente relacionados. No podemos esperar encontrar un buen niño-máquina al primer intento. Uno debe experimentar enseñando a una máquina así y ver qué tan bien aprende. Luego, se puede intentar con otra y ver si es mejor o peor. Hay una conexión obvia entre este proceso y la evolución, a través de estas identificaciones:

Estructura del niño-máquina = material hereditario
Cambios de niño-máquina = Mutaciones
Selección natural = Juicio del experimentador

Uno podría esperar, no obstante, que este proceso será más expedito que la evolución. La sobrevivencia del más apto es un método lento para medir ventajas. El experimentador, a través del ejercicio de la inteligencia, debiera ser capaz de acelerarlo. Igualmente importante es el hecho de que él no se encuentra restringido a las muta-

ciones aleatorias. Si puede rastrear la causa para una debilidad, puede probablemente pensar en el tipo de mutación para mejorarla.

No será posible aplicar exactamente los mismos métodos de enseñanza a una máquina que los que se le aplicarían a un niño normal. No se le puede proveer piernas, por ejemplo, por lo que no se le pedirá que vaya a buscar el cubo para el carbón. Posiblemente no tendría ojos. Pero sea como sea que estas deficiencias sean superadas por un diseño inteligentemente planeado, uno no podría enviar a la criatura a la escuela sin que los niños no se burlaran excesivamente de ella. Se le debe dar cierta instrucción. No debemos preocuparnos tanto de las piernas, ojos, y otra cosas. El ejemplo de la señorita Helen Keller demuestra que la educación puede producirse siempre y cuando la comunicación en ambos sentidos entre maestro y pupilo se produzca de una u otra manera.

Normalmente, se asocian los castigos y recompensas con proceso educativo. Se podría construir o programar ciertos niño-máquinas simples basados en este principio. La máquina debe ser construida de tal manera que los eventos previos a una señal de castigo sean poco probables de volver a ocurrir, mientras que la señal de recompensa aumente la probabilidad de repetición de los eventos que llevaron a ella. Estas definiciones no presuponen ningún sentimiento por parte de la máquina. He hecho algunos experimentos con uno de estos niño-máquinas, y tuve éxito en ensañarle algunas cosas, pero el método de enseñanza fue demasiado poco ortodoxo como para que el experimento sea considerado realmente exitoso.

El uso de castigos y recompensas puede, en el mejor de los casos, ser sólo una de las partes del proceso de enseñanza. En términos muy generales, si el profesor no tiene otros medios de comunicarse con el pupilo, la cantidad de información que le puede llegar no excede el número total de recompensas y castigos aplicados. Para el momento en el cual el niño ha aprendido a repetir "Casabianca", probablemente estará muy adolorido, si el texto sólo se pudiera descubrir a través una técnica tipo "Veinte Preguntas", donde cada "No" toma la forma de un golpe. Es necesario por lo tanto tener algunos otros canales de comunicación "no emocionales". Si estos estuvieran disponibles, sería posible enseñarle a una máquina a través de castigos y recompensas a obedecer órdenes en algún lenguaje, por ejemplo un lenguaje simbólico. Estas órdenes serían transmitidas a través de los canales "no emocionales". El uso de este lenguaje disminuirá en gran medida el número de castigos y recompensas requeridos.

Las opiniones pueden variar respecto a la complejidad que es adecuada en el niño-máquina. Uno podría tratar de hacerlo lo más simple posible consistentemente con los principios generales. O bien, uno podría tener un sistema completo de inferencias lógicas "construido dentro" de él. En este último caso, el almacenamiento sería mayormente ocupado con definiciones y proposiciones. Las proposiciones tendrían varios tipos de estatus, tales como hechos bien establecidos, conjeturas, teoremas matemáticos probados, declaraciones hechas por una autoridad, y expresiones que tengan la forma lógica de una proposición pero no un valor de creencia. Ciertas proposiciones pueden ser descritas como "imperativas". La máquina debiera ser construida de tal manera que tan pronto como una proposición imperativa sea clasificada como "bien establecida", la acción apropiada se produce inmediatamente. Para ilustrar esto, supongamos que el profesor le dice a una máquina "haz tu tarea ahora". Esto podría hacer que "El profesor dice 'haz tu tarea

ahora"' sea incluida entre los hechos bien establecidos. Otro hecho como este podría ser "todo lo que dice el profesor es verdad". Al combinar estos dos se podría llegar al fin a que el imperativo "haz tu tarea ahora" sea incluida entre los hechos bien establecidos, y esto, dada la construcción de la máquina, implicará que la tarea de hecho comienza a ser realizada; pero el efecto es muy satisfactorio. El proceso de inferencia usado por la máquina no necesita satisfacer a los lógicos más exactos. Podría no existir una jerarquía de tipos, por ejemplo. Pero esto no necesariamente significa que las falacias de tipo tienen que ocurrir más que lo que nosotros estamos destinados a caer en los acantilados sin baranda. Imperativos apropiados (expresadas dentro del sistema, no que formen parte de las reglas del sistema) tales como "no use una clase a menos que sea una subclase de otra que ha sido mencionada por el profesor" pueden tener un efecto similar a "no acercarse demasiado al borde".

Los imperativos que pueden ser obedecidos por una máquina que no tenga extremidades están destinados a tener un carácter más bien intelectual, como en el ejemplo (hacer la tarea) dado arriba. Los imperativos que serán más importantes son aquéllos que regulen el orden en el cual las reglas del sistema lógico utilizado sean aplicadas. Pues en cada etapa en la que se use un sistema lógico, existe un gran número de pasos alternativos, cualquiera de los cuales se puede aplicar en lo que concierne a la obediencia de las reglas del sistema lógico. Estas elecciones hacen la diferencia entre un razonador brillante y uno inútil, no la diferencia entre uno consistente y uno falaz. Las proposiciones que lleven a los imperativos de este tipo podrían ser "cuando Sócrates sea mencionado, use el silogismo en Bárbara" o "si un método demuestra ser más rápido que otro, no utilice el lento". Algunos de éstos se podrían ser "hechos por una autoridad", pero otros podrían ser producidos por la propia máquina, a través, por ejemplo, de una inducción científica.

La idea de una máquina que aprende podría parecer paradójica para ciertos lectores. ¿Cómo pueden las reglas de funcionamiento de una máquina cambiar? Estas debieran describir completamente cómo la máquina reaccionará sin importar cuál sea su historia, o los cambios por los que pudo haber pasado. Las reglas son entonces bastante invariantes en el tiempo. Esto es muy cierto. La explicación de la paradoja es que las reglas que son cambiadas en el proceso de aprendizaje son de un tipo menos pretencioso, afirmando solamente tener una validez efímera. El lector puede hacer un paralelo con la constitución de los Estados Unidos.

Una característica importante de una máquina que aprende es que su profesor será con bastante frecuencia muy ignorante sobre qué es lo que ocurre dentro de ella, aunque aún así podría ser capaz de predecir la conducta de su pupilo, hasta cierto punto. Este debería ser el caso en mayor medida en la educación tardía de una máquina que surja de un niño-máquina con un diseño (o programa) bien probado. Esto se encuentra en directo contraste con el procedimiento normal al utilizar una máquina para que haga cálculos: en ese caso, el objetivo es tener una imagen mental clara del estado de la máquina en cada momento de la cálculo. Este objetivo se puede lograr solamente con dificultad. La visión de que "la máquina solamente puede hacer lo que sepamos ordenarle que haga", se vuelve extraña de cara a esto. La mayoría de los programas que podamos otorgarle a una máquina resultarán en que ésta haga algo que no tiene ningún sentido para nosotros, o lo consideraríamos como comportamiento

completamente aleatorio. El comportamiento inteligente consiste presumiblemente es una desviación de la conducta completamente disciplinada involucrada en los cálculos, pero una desviación moderada, que no genera conductas aleatorias, o secuencias repetitivas sin sentido. Otro resultado importante de preparar a nuestra máquina para su parte en el juego de la imitación a través de un proceso de enseñanza y aprendizaje es que la "falibilidad humana" probablemente sea omitida de manera natural, es decir, sin un adiestramiento especial. (El lector debiera reconciliar esto con el punto de vista en páginas anteriores) Los procesos que son aprendidos no producen una certeza del cien por ciento en el resultado; si así fuera, no podrían ser desaprendidos.

Sería aconsejable incluir un elemento aleatorio en una máquina que aprende. Un elemento aleatorio es bastante útil cuando buscamos la solución a un problema. Suponga, por ejemplo, que quisiéramos encontrar un número entre 50 y 200 el cual es igual al cuadrado de la suma de sus dígitos; podríamos empezar con 51 y luego tratar con 52 y continuar hasta que encontremos un número que funcione. Como alternativa, podríamos elegir números al azar hasta que encontremos uno que sirva. Este método tiene la ventaja de que es innecesario mantener registro de los valores que han sido probados, pero la desventaja de que se podría intentar con el mismo número dos veces; pero eso no es muy importante si hay varias soluciones. El método más sistemático tiene la desventaja de que puede haber un enorme bloque sin soluciones en la región que se investiga primero. Ahora el proceso de aprendizaje podría ser considerado como la búsqueda de una forma de comportamiento que satisfaga al profesor (o algún otro criterio). Dado que probablemente haya un gran número de soluciones satisfactorias, el método aleatorio parece ser mejor que el sistemático. Debería señalarse que este es el que se usa en el proceso análogo de la evolución, ya que el método sistemático no es posible. ¿Cómo se podría mantener registro de las distintas combinaciones genéticas que se han probado, de manera de evitar realizarlas de nuevo?

Podríamos esperar que las máquinas eventualmente compitan con los hombres en todos los campos puramente intelectuales. Pero, ¿cuáles son los mejores para comenzar? Incluso eso es una decisión difícil. Mucha gente cree que una actividad muy abstracta, como jugar ajedrez, sería lo mejor. También se puede sostener que lo mejor es proveer a la máquina con los mejores órganos sensoriales que el dinero pueda comprar y enseñarle a comprender y hablar inglés. Este proceso podría seguir la enseñanza normal de un niño. Las cosas podrían ser señaladas y nombradas, y así sucesivamente. Nuevamente, no sé cuál sea la respuesta correcta, pero creo que ambas aproximaciones debieran intentarse.

Sólo podemos ver una corta distancia delante de nosotros, pero podemos ver ahí, muchísimo de lo que se necesita hacer.