Publicat en *Quaderns de filosofia i ciència,* nº 21-22, pàgs: 45-54.
Societat de Filosofia del País Valencià, València, 1993.

¿NOS DAN LAS GRACIAS LAS MÁQUINAS DE TABACO?

Antoni Defez i Martín

> BORDEU: ¿Habéis visto en el jardín del rey, dentro de una jaula de verdor, un orangután que tiene el aspecto de un San Juan que predica en el desierto?
> MLLE DE LESPINASSE: Sí, lo he visto.
> BORDEU: El cardenal de Polignac le decía un día: ¡habla y te bautizo!
>
> Denis DIDEROT

> ¡Pero una máquina seguramente no puede pensar!
> -¿Es éste un enunciado empírico? No. Sólo decimos que piensa un ser humano y aquello que se le parece...
>
> L. WITTGENSTEIN, *Investigaciones filosóficas,* § 360

El título de este trabajo -"¿Nos dan las gracias las máquinas de tabaco?"- no representa, en realidad, una correcta descripción de lo que pretendo en estas páginas. Mi intención no es únicamente, como podría parecer, tratar la cuestión de si las máquinas, o mejor los ordenadores, pueden llegar a ser conscientes, tener intenciones y sentimientos o pensar. Además de la cuestión de si las máquinas pueden llegar a tener mente, me gustaría decir algo también sobre el problema de si eso que llamamos mente (por supuesto, mente humana) puede ser considerada como un ordenador, como una máquina de Turing o, si se quiere, como un programa informático. Ambos problemas, siendo distintos, se encuentran, sin embargo, íntimamente relacionados, ya que una respuesta afirmativa al primero de ellos debe descansar, como hace la tesis de la Inteligencia Artificial, en una respuesta afirmativa al segundo. No obstante, comprometerse con la idea de que la mente humana total o parcialmente funciona como un ordenador, o es funcionalmente equivalente a un programa informático, no exige

necesariamente apostar por la posibilidad de que los ordenadores lleguen *de facto* a tener mente.

Dado que estas cuestiones son complejas de analizar y se necesita transitar a través de argumentaciones muy enmarañadas, procederé de entrada, y para que todo quede claro desde el principio, a exponer cuál es la tesis que pretendo defender. Por las razones que después se verán, considero imposible que bajo ninguna circunstancia un ordenador pueda llegar a poseer mente; y en segundo lugar, tampoco creo que se pueda decir que la mente humana funcione esencialmente como un ordenador, o que *in toto* no sea más que un sistema de procesamiento de información "definido por" y "explicable como" un programa informático. Para decirlo contestando la pregunta del título: ni las máquinas nos dan las gracias cuando nos dicen "encantada de servirle", ni tampoco es el caso que los seres humanos nos limitemos a proferir la oración “encantada de servirle", o alguna otra similar, cuando damos las gracias a otro ser humano. No obstante, no pretendo negar lo evidente: en concreto, que los ordenadores o que las máquinas de Turing puedan "simular", y a veces incluso hacer con más eficacia, algunas de las tareas humanas, incluidas algunas de las llamadas "mentales". Igualmente, encuentro también sugerente que algunas de las actividades mentales de los seres humanos puedan ser contempladas bajo el modelo explicativo de los programas informáticos o de los sistemas de procesamiento de información.

Mi posición respecto a estas cuestiones tiene como soporte una tesis filosófica de alcance más general, tesis que creo indiscutible a no ser que estemos dispuestos, de una parte, a renunciar a todo aquello que mejor tenemos para saber qué tipo de objetos somos -incluyendo nuestra "mente"-, a saber, nuestro conocimiento científico de la realidad; y de otra, que estemos dispuestos a abrazar "una mística de ciencia-ficción". Se trata de la tesis que paso a formular

en los dos siguientes enunciados: (i) Los seres humanos son organismos biológicos que filogenéticamente y ontogenéticamente son respectivamente el resultado de largos procesos de evolución natural y de desarrollo físico; (ii) Pero no sólo son los seres humanos seres biológicos; también, y de forma esencial para lo que son y lo que hacen, son producto tanto a nivel filogenético como ontogenético de procesos históricos e individuales de socialización. Dicho con otras palabras: un ser humano, con su mente incluida, es un organismo biológico -Y, por tanto, físico-, así como un ser histórico y socializado que interactúa con el medio natural y con el medio histórico y social que le rodea.

Como puede apreciarse, tengo interés en resaltar el carácter actuante e histórico-social de los seres humanos: somos organismos, es cierto; pero también se puede decir por ello que somos acción biológica y acción social históricamente determinada. Éste es un detalle que, en mi opinión, pasa frecuentemente desapercibido a aquellos que se ocupan de los problemas de la filosofía de la mente, cuando, por el contrario, parece que sea un dato que deba jugar un papel central en tales discusiones. A fin de comprobar su rendimiento filosófico, comenzaré considerando los análisis que recientemente ha hecho J. Searle y sus discusiones con algunos de los defensores de la I.A., por ejemplo con los Churchland[1].

De acuerdo con Searle, la tesis de la I.A. contiene en germen un "dualismo residual", en la medida que ofrece una interpretación de la mente humana en términos exclusivamente de

[1] Vid. J. Searle, "¿Es la mente un programa informático?", en *Investigación y Ciencia*, nº 162, marzo de 1990, págs. 10 y ss. (Este artículo forma parte del debate que Searle mantiene con el matrimonio Churchland en el citado número de la revista reseñada). Vid. también: J. Searle, "Minds, Brains and Programs", en Haugeland, J. (ed.), *Mind Design,* The MIT Press, Cambridge, 1981, págs. 283 y ss., y *Minds, Brains and Science,* British Boadeasting Co., Londres, 1984, Caps. 1-3.
Como introducciones históricas y filosóficas al problema, vid, p.e., P. Lacasa López, *La Psicología hoy: ¿organismos o máquinas?,* Edit. Cincel, Madrid, 1985, Cap. 8; J. Bouveresse, "¿Son inteligentes las máquinas?", J. Pitrat, "El nacimiento de la inteligencia artificial", J. G. Ganascia, "La concepción de los sistemas expertos", J. P. Sansonnet, "Las máquinas de la inteligencia artificial", en *Mundo Científico* (La Recherche), n.' 53, diciembre de 1985, págs. 11 94 y ss.

conjuntos de operaciones, funciones o conductas definibles formalmente, olvidando así el hecho de que los seres humanos y, en concreto, nuestros cerebros son organismos biológicos. De esta manera la I.A. estaría apostando por la idea de que lo mental constituye algo separado e independiente del cerebro o de cualquier otro sistema físico específico. Como dice él mismo, refiriéndose a los defensores de la I.A., "la mente, suponen, es algo formal y abstracto, no una parte de la sustancia húmeda y escurridiza que ocupa nuestros cráneos" [2]. En contra de esta separación entre lo mental y lo natural, Searle propone un análisis de la mente que podemos resumir de la siguiente manera: (i) los fenómenos mentales "están causados por" procesos cerebrales; y (ii) los fenómenos mentales son rasgos de los cerebros o, si se quiere, "están realizados" en el cerebro[3]. La idea de Searle parece clara: contrariamente a lo que piensan los defensores de la I.A., no es indiferente ni irrelevante la materia con la que estamos hechos o la materia que constituye nuestro cerebro o nuestro Sistema Nervioso Central.

Desde un principio, los defensores de la I.A. han venido afirmando que análogamente a lo que sucede con los ordenadores entre el nivel del *hardware* y el nivel del *software,* acaece entre el cerebro y la mente, a saber, que así como el material con que está construido un ordenador es irrelevante al programa informática que ejecuta, igualmente la pasta que constituye nuestro cerebro es irrelevante a la mente o al programa informática que nos define[4]. Como en cierta ocasión llegó a decir H. Putnam de una manera un tanto láctea: "Podríamos estar hechos de queso suizo y esto no cambiaría las cosas" [5].

---

[2] J. Searle, "¿Es la mente un programa informático?", *op. cit.,* pág. 16.

[3] J. Searle, *op. cit.,* págs. 14 y ss.

[4] Para comprender el alcance de esta afirmación vid., p.e., el conjunto de artículos que H. Putnam dedica a la cuestión recopilados en *Mind, Language and Reality,* Cambridge University Press. Londres, 1975.

[5] H. Putnam, "Philosophy and our mental life" (1973), *op. cit.,* pág. 291.

Ahora bien, ¿en qué se sustenta esta analogía entre los ordenadores y los seres humanos? Para entenderlo necesitamos exponer las guías maestras por las que se ha desarrollado la tesis de la I.A. En este sentido, resulta paradigmático la propuesta metodológica de A. M. Turing. Para este autor, auténtico iniciador de la I.A., deberíamos sustituir la pregunta absurda o carente de sentido de si las máquinas (los ordenadores) pueden llegar a tener mente por la pregunta significativa de si las máquinas pueden "imitar" o "simular" la conducta humana llamada mental[6]. Para ello Turing ideó una serie de "juegos de imitación" en los que intervenían seres humanos y máquinas, y en los cuales la conducta de las máquinas resultaba, bajo ciertas condiciones de observación, totalmente indistinguible e indiferenciable de la conducta mental humana. Ante un intercambio de preguntas-respuestas, un observador, que no pudiese ver a los autores de las respuestas, no podría de ninguna manera saber cuándo la respuesta provenía de una persona o de un ordenador[7]. Sin embargo, es necesario señalar, como indicó K. Gunderson, que pese a sus intenciones Turing se deslizaba a contestar la pregunta absurda mediante la pregunta significativa, y de esa manera a considerar que si los ordenadores pueden simular la conducta mental humana, entonces cabe decir que tienen mente en la misma medida que lo decimos de los seres humanos[8].

Y a lugar parecido, podríamos decir, nos conducen, por ejemplo, los análisis posteriores de A. Newell o de J. A. Fodor. Así, Newell, pese a decirnos que la I.A. o bien es irrelevante, o bien ofrece soluciones inaceptables al problema filosófico de la mente, y pese a afirmar también que la I.A. habla de la naturaleza de la mente sólo desde el interior de su propia

---

[6] A. M. Turing, "Maquinaria computadora e inteligencia" (1950), en Alan Ross Anderson (ed.), *Controversia sobre mentes y máquinas,* Tusquets Eds., Barcelona, 1984, pág. 11.

[7] A. M. Turing, *op. cit.,* págs. 12 y ss.

[8] K. Gunderson, "El juego de imitación", en A. R. Anderson (ed.), *op. cit.,* págs. 96-97.

perspectiva, no obstante no renuncia a concebir la mente como un sistema de procesamiento de información[9]. Por su parte Fodor, que ha propuesto sustituir el concepto de "simulación" entendido, como hizo Turing, en términos de simples analogías de conducta observable por el de "equivalencia funcional", piensa que del hecho empírico de una tal equivalencia entre la conducta de los ordenadores y los organismos pueden razonablemente seguirse dos conclusiones: (i) que un determinado programa informático sea la teoría psicológica explicativa de la conducta de los organismos en cuestión; y (ii) que, como hecho de política lingüística, podamos aplicar racionalmente nuestros conceptos mentales a las máquinas en que se dé tal programa informática, cosa equivalente en su posición a decir que los ordenadores puedan poseer mente, ya que nos recomienda no distinguir entre "corrección lingüística" y "verdad" [10].

Contra estas interpretaciones de la mente como programa informático, Searle ha reaccionado correctamente afirmando que las mentes o los cerebros tienen no sólo sintaxis sino también semántica, a diferencia de los ordenadores o los programas informáticos. Estos últimos son definibles y explicables como conjuntos de instrucciones sintácticas o formales; para ellos parece regir lo que se podría llamar el principio de la indiferencia del sustrato material. No ocurre, sin embargo, lo mismo con los seres humanos. Lo que llamanos mente no es definible o explicable sólo, ni tampoco siempre parcialmente, como un conjunto de instrucciones formales o sintácticas. Si queremos ofrecer una explicación de lo que es la mente humana resulta irrenunciable su carácter semántico, lo cual en Searle conduce a la consideración de nuestro sustrato material, en concreto, nuestro cerebro o nuestro S.N.C. como aquella instancia donde podemos encontrar una explicación de los contenidos cualitativos de los llamados estados mentales, del carácter subjetivo de la experiencia, de la intencionalidad de

[9] Vid. A. Newell, *Inteligencia artificial y el concepto de Mente* (1973). Teorema, Valencia, 1980. (Un razonamiento similar lo ofrece H. Putnam en "Minds and machines" (1960), *op. cit.,* págs. 362 y ss.)
[10] J.A. Fodor, *La explicación psicológica* (1968), Cátedra, Madrid, 1980.

lo mental, de la autoconciencia, y de las relaciones causases en que se encuentran lo mental y lo físico.

Consideremos, por ejemplo, el caso de la percepción. Siguiendo las tesis de la I.A. podríamos definir el percibir los colores como un conjunto de operaciones, conductas o funciones discriminatorias de diversas longitudes de ondas electromagnéticas. Esto sería una especie de programa informático: dados unos determinados *inputs* el sistema reacciona con unos determinados *outputs,* donde es irrelevante o ha desaparecido lo que llamamos el contenido cualitativo de la percepción. Ahora bien, esto representa una incorrecta explicación de lo que es "percibir colores". El propio Putnam posteriormente lo ha reconocido mediante el análisis de casos de espectro invertido, mostrando así las fallas de sus concepciones anteriores. Percibir un color no es simplemente ofrecer un *output* determinado ante determinado *input*, esto es, percibir no es definible como una conducta discriminatoria; resulta para ello necesario tener en consideración el contenido cualitativo correspondiente. En un caso de espectro invertido una misma persona en tiempos diferentes, a pesar de discriminar igualmente su cerebro la misma longitud de onda, percibirá colores diferentes: *ex hypothesi,* puede en un caso ver azul y en otro rojo. Y ello muestra bien a las claras la necesidad de hacer referencia al sustrato material que nos define, es decir, a los concretos estados o procesos cerebrales que acaecen en nosotros[11].

Ejemplos como éste, y por los motivos señalados por Searle, muestran la insostenibilidad de la tesis de I.A.: somos organismos semánticos y no sólo artefactos sintácticos. Ahora bien, ¿de dónde procede el error de los defensores de la tesis de la I.A.?

[11] H. Putman, "Mind and body", en *Reason, Truth and History,* Cambridge University Press, Londres, 1988.

Diversos autores han señalado, y creo que acertadamente, que en esta línea de pensamiento existe una confusión categorial o lógica inicial. Se trata de asimilar, como dice el propio Searle, los conceptos de "simular" o de "imitar" al concepto de "duplicar" [12]. O como la describió Gunderson mediante su "juego del pisotón": no haber distinguido entre el concepto de "hacer lo mismo" y el de "lograr el mismo resultado final" [13]. Y es que como indicó M. Seriven, los argumentos basados en la emulación de la conducta no sirven a estos propósitos[14]. Se podría decir, por tanto, que los ordenadores (las máquinas de Turing) pueden "imitar", "simular" o "lograr el mismo resultado final" que los seres humanos con respecto a gran cantidad de actividades que realizan estos últimos; podemos incluso aceptar que lo hacen a veces con mayor eficacia que nosotros. No obstante, podemos afirmar también que los ordenadores (las máquinas de Turing) no pueden ni podrán "duplicarnos" o "hacer lo mismo" que hacemos nosotros. Y la razón de ello es la indicada por Searle: los cerebros causan mentes y los fenómenos mentales son rasgos o están realizados en el cerebro. Y así, si deseamos afirmar que cualquier otro sistema posee mente, deberá suceder que tal sistema posea poderes causases equivalentes a nuestro cerebro. Volviendo a nuestro ejemplo anterior, las máquinas de Turing podrán hacer las discriminaciones que hacemos los humanos cuando percibimos; sin embargo, no por ello percibirán colores.

En este sentido, de nada vale apelar a la totalidad formal o sintáctica del sistema, como hace el matrimonio Churchland. Los Churchland oponen al artificio conceptual de la "Sala china", ideado por Searle para demostrar que "simular" y "duplicar" son cosas distintas con respecto a entender chino, el artificio de un "Gimnasio Chino" en el cual, si bien es cierto que

[12] J. Searle, *op. cit.,* págs. 13 y 16.

[13] K. Gunderson, *op. cit.,* págs. 102 y ss.

[14] M. Scriven, "El concepto mecánico de la mente" (1953), en A. R. Anderson (ed.), págs. 51 y ss. Vid. P. M. Churchland y P. S. Churchland, "¿Podría pensar una máquina?", en *Investigación y ciencia,* nº 162, marzo 1990,

nadie habla ni entiende una palabra de chino, sin embargo es posible que de la totalidad del gimnasio -de un gimnasio, dicen, suficientemente grande y complejo- pueden darse o emanar fenómenos mentales. Para decirlo de otra manera: los gimnastas no entenderán chino, pero el gimnasio como un todo sí[15]. Y digo que de nada vale porque el sistema como tal sigue siendo definido únicamente como un sistema formal o sintáctico que simplemente "simula" o "imita" un determinado conjunto de operaciones, funciones o conductas.

Ahora bien, la apelación a la semántica o a los contenidos mentales como algo distintivo de la mente humana no parece que sea el único aspecto en que ésta se diferencie de los ordenadores. Es necesario también no olvidar lo que podemos llamar la vertiente pragmática de los seres humanos y de sus mentes, esto es, lo que Searle ha presentado como el rasgo de la "intencionalidad" de la mente humana[16]. Podríamos introducir este carácter de nuestra mente afirmando que además de ser seres biológicos somos también "organismos" que de forma necesaria actuamos tanto biológica como socialmente. Consideremos esto último detenidamente. Me ocuparé de la acción humana bien en la medida que hay fenómenos mentales -como, por ejemplo, "dar las gracias"- que no pueden ser reducidos semánticamente a puros estados cerebrales y también en la medida que aquello que parece central o esencial a la mente humana -la llamada "autoconsciencia"- tampoco no parece explicable si nos olvidamos que somos seres que actuamos.

Comencemos por el problema de "dar las gracias". De entrada, hay que señalar que la apelación a contenidos mentales o a estados cerebrales parece plausible para aquellos

---

págs. 18 y ss.

[15] Vid. P. M. Churchland y P. S. Churchland, "¿Podría pensar una máquina?", en *Investigación y ciencia,* n.' 162, marzo 1990, págs. 18 y ss.

[16] Vid. J. Searle, *Minds, Brains and Science,* págs. 57 y ss.

fenómenos mentales que tendrían una realización cerebral. Sin embargo, hay fenómenos mentales que exigen de forma necesaria la existencia de un contexto social y lingüístico, tanto para que se den de forma efectiva como para que sean adquiridos por los seres humanos a través de los procesos de socialización a que son sometidos. Dicho de otra manera: así como para sentir dolor o percibir colores nos basta con ser seres biológicos y no hay necesidad de haber sido socializados, con respecto a fenómenos mentales tales como "dar las gracias", "prometer" o "emocionarse con una obra de arte" nos resulta imprescindible haber sido introducidos en determinadas prácticas sociales, culturales y lingüísticas. Sólo bajo ciertas condiciones contextuales es posible realizar una promesa, dar las gracias o emocionarse estéticamente. Por ejemplo, y tal como indicó hace tiempo J. L. Austin, para "prometer" no basta con decir "prometo", debemos estar dispuestos también, a cumplir lo prometido, a aceptar reproches si no lo hacemos y, por último, a ofrecer excusas si fuese necesario[17]. Y lo mismo sucedería con "dar las gracias".

No somos "cerebros en una cubeta" [18]. Por el contrario, somos organismos socializados que interactuamos con el medio natural y social que nos rodea. Y esto último significa, como ha sugerido Wittgenstein, participar y ser competentes en concretas prácticas sociales establecidas[19]. Sólo por referencia a ellas, sólo en el interior de las totalidades que ellas constituyen tienen lugar y, por ende, resultan significativos tales fenómenos llamados mentales. En conclusión, los ordenadores no pueden "dar las gracias", ni "prometer", así como tampoco las cajetillas de tabaco -que en realidad no son sino otro tipo de máquinas- no nos advierten "que fumar perjudique seriamente a la salud". En todo caso, quien nos da las gracias cuando

[17] Vid. J. L. Austin, "'Other Minds' (1946) y "A plea of Excuses" (1956), en *Philosophical Papers,* Oxford University Press, Londres, 1970.

[18] Vid. H. Putnam, "Brains in a vat", en *Reason, Truth and History,* Cambridge University Press, Londres, 1988.

[19] Vid. L. Wittgenstein, *Investigacions Filosófiques* (1958), Laia, Barcelona, 1983.

compramos tabaco o quien nos avisa de sus peligros no es sino la Tabacalera, el Estado o, siendo más precisos, los seres humanos responsables de esas instituciones.

Llegados a este punto quisiera señalar que estas conclusiones no pretenden ser afirmaciones sintéticas mediatizadas por los resultados actuales de las investigaciones, tal que pudiesen ser modificables en un futuro. Si de lo que se trata es de demostrar que los ordenadores tienen o no tienen mente, debemos hacer uso de argumentos lógicos o conceptuales y no de argumentos empíricos. La cuestión, por tanto, no es demostrar que los ordenadores "ahora" no tienen mentes, sino que "jamás" podrán tenerlas pase lo que pase en el futuro. En este sentido a veces se dice "que vale, que en la actualidad no son las cosas de esa manera, pero quién sabe si en un futuro...". Observaciones de este estilo, que reducen la dificultad a un mero problema de "complejidad técnica", son las que nos conducen a lo que llamé al principio "la mística de la ciencia-ficción". Y para combatirla los argumentos de Searle no parecen que sean suficientes. Veámoslo.

Consideremos a tal efecto un argumento que frecuentemente ha sido usado para esclarecer la analogía entre mentes y máquinas de Turing. Se trata del teorema mediante el cual K. Gödel demostró que todo sistema formal o es completo pero inconsistente, o bien es consistente pero incompleto[20]. De acuerdo con Gödel, todo sistema axiomático siempre e inevitablemente dará lugar a teoremas o fórmulas que, perteneciendo a dicho sistema, serán, sin embargo, indecidibles o indemostrables dentro de este sistema. Consecuentemente, si el sistema es completo contendrá teoremas o fórmulas indemostrables y, por tanto, será inconsistente. Y al revés: el precio de la consistencia para todo sistema formal es que sea

---

[20] Vid. K. Gödel, "Sobre proposiciones formalmente indecidibles de los Principia Mathematica y sistemas afines" (1931), Teorema, Valencia, 1977.

incompleto, esto es, que no incluya todos aquellos teoremas o fórmulas que, derivándose correctamente de sus axiomas y reglas de transformación, no obstante sean indemostrables desde sí mismo. Evidentemente, la solución consiste en elaborar para un sistema dado otro sistema que lo abarque y desde el cual sea posible demostrar los teoremas que se resistán al primero. Ahora bien, sucede que por los mismos motivos en este segundo sistema, y en cualquier otro que queramos construir sucesivamente, volverán a plantearse los mismos problemas que se presentaban con respecto al primero. En sentido estricto, las dificultades reaparecerán tantas veces como sistemas vayamos creando, es decir, podemos crear sistemas *ad libitum* pero las dificultades reaparecerán *ad infinitum.*

Y bien, ¿por qué es interesante el teorema de Gödel para nuestro problema? Muy simple: un ordenador (una máquina de Turing) es un sistema formal o sintáctico finito con respecto al cual, como con respecto a cualquier otro sistema formal, el teorema de Gödel se cumplirá. Dicho de otra manera: los ordenadores son máquinas gödelianas para las cuales siempre habrán teoremas pertenecientes a sus programas que les serán indemostrables. Y el problema es que no podemos construir un sistema infinito de ordenadores, así como no podemos construir infinitos sistemas formales. Y he aquí la conclusión importante: un sistema formal o sintáctico por su propia esencia está incapacitado para la autorreferencia o, si se quiere, para la autoconsciencia. Ningún sistema formal y, por tanto, ningún ordenador puede referirse a sí mismo o tomarse a sí mismo en consideración de una forma completa: ni podrá demostrar, ni alterar todo aquello que lo define.

La gracia de esta argumentación estriba, evidentemente, en afirmar que la mente humana no es gödeliana. Y ello parece cierto, al menos, en algún sentido: la mente humana puede lo que un ordenador no puede, tanto con respecto a sí misma como con respecto a

cualquier ordenador. La mente humana tendría una capacidad recursiva infinita o, si se quiere, una capacidad infinita de autorreferencia o de autoconsciencia, ya que siempre se tiene la posibilidad de crear un sistema formal *n* que abarque a cualquier sistema dado y que demuestre lo que este sistema es incapaz de demostrar de sí mismo. Con otras palabras: la mente humana puede de forma infinita ser consciente de sí misma y tomarse en consideración a sí misma de forma infinita, cosa que no pueden hacer los ordenadores.

La que acabo de exponer es, en resumen, la argumentación que J.R. Lucas ha realizado en contra de las valoraciones de Turing y Putnam del poder demostrativo del teorema de Gödel con respecto a las diferencias entre la mente humana y los ordenadores[21]. Para Turing, por ejemplo, el teorema de Gödel no demostraba nada porque, en su opinión, no es cierto que los seres humanos no sean sistemas gödelianos. Para Turing, somos seres con limitaciones gödelianas; y además, así como hay seres humanos más inteligentes que otros, también hay o pueden haber ordenadores más inteligentes que otros y más inteligentes que muchos seres humanos[22]. A su vez, Putnam argüía, de forma parecida, que un ordenador puede poseer en su programa el teorema de Gödel y que la supuesta superioridad de la mente humana no debe consistir en poseer tal teorema, sino en demostrar para todo sistema dado que o es completo pero inconsistente, o bien que es consistente, pero incompleto. Y esto es precisamente lo que resulta, según Putnam, bastante inalcanzable para la mayoría de los seres humanos, si el sistema formal en cuestión es suficientemente complejo[23].

[21] J.R. Lucas, "Mentes, máquinas y Gódel" (1 96 l), en A. R. Anderson (ed.), *op. cit.,* págs. 69 y ss.

[22] A. M. Turing, *op. cit,* págs. 27-29.

[23] H. Putnam, "Mind and machines", en *op. cit.,* pág. 366.

Bueno, ¿y quién lleva razón?, ¿somos o no somos sistemas gödelianos? Creo que la discusión está viciada. Por una parte, parece que lleven razón Turing y Putnam al afirmar que la mente humana es tan gödeliana como cualquier ordenador: de hecho, la mayoría de los seres humanos no somos ni matemáticos, ni mucho menos matemáticos brillantes. El problema, sin embargo, es que se ha planteado, en mi opinión, en un terreno estéril, cosa que convierte a la discusión, como acabo de decir, en una disputa viciada. Creo que el problema debe ser tratado desde la perspectiva de nuestra conducta en tanto organismos que somos y no estrictamente desde nuestra posible conducta formal. Vayamos a ello.

Lo que Gödel viene a decimos es que un sistema formal llegará un momento en que quedará, digamos, "bloqueado", ya que habrán enunciados que le pertenecen y que le son indemostrables. Ahora bien, los seres humanos no somos sistemas formales y, por tanto, no somos sistemas gödelianos, aunque un sistema formal pueda ser un buen modelo explicativo para algunas de las cosas que hacemos. Por el contrario, los seres humanos somos organismos biológicos socializados y, como tales, la acción nos es esencial. Y he aquí lo importante: como organismos no podemos dejar o abstenernos de actuar. Dicho de otro modo: la posibilidad de quedarnos parados, quietos o "bloqueados" o en una tesitura de indecidibilidad nos es imposible, so pena de dejar de ser organismos y cesar. En cualquier situación que se nos presente, aun en el caso que no sepamos qué es lo que más nos conviene o lo que es más correcto o lo verdadero, siempre actuaremos bien o mal saliendo del *impasse,* modificando la situación, o intentando solucionar la dificultad que nos apremiaba, etc. En realidad es ésta la manera como la especie humana ha ido evolucionando biológica y socialmente hasta llegar a ser lo que ahora somos. Y lo mismo, *mutatis mutandis,* puede ser dicho con respecto a la constitución individual de un ser humano: su desarrollo físico y mental se lleva a cabo por la

acción. En resumen: nos es esencial e inevitable interactuar con el medio que nos rodea. Ya lo decía el *Fausto* de Goethe: *"Al principio fue la acción"*.

No plantear la cuestión desde este punto de vista puede dejar abierta la puerta a los llamados procesos de "emergencia" a que antes aludíamos. Si la "autoconsciencia" no es otra cosa que la capacidad de un sistema formal para la recursividad infinita, entonces ¿por qué no imaginar que tal capacidad pueda emerger a partir de cierto grado de complejidad del sistema? Esta posibilidad fue, de hecho, contemplada por Turing; pero también por su contrincante Lucas[24]. Ahora bien, de acuerdo con este último en ese caso, y a pesar de todo, la tesis de la I.A. continuaría siendo incorrecta, ya que no tendríamos todavía derecho a afirmar que una mente es un ordenador. La mente humana sería una "emergencia" de sistemas formales supercríticos, pero no un sistema formal en sí mismo. Dicho de otra manera: los ordenadores no serían mentes, aunque podrían causar mentes.

Esta posibilidad da lugar a una "mística de ciencia-ficción", la cual se basaría en la tesis de la I.A., pero que además descansaría en una concepción semejante a la de los Churchland cargada, eso sí, de "emergentismo" a partir de una supuesta complejidad creciente de los sistemas. Vendría a decir algo como lo siguiente: de un sistema físico dotado de un programa informático lo suficientemente complejo puede "emerger" una mente. Y así, nos encontramos en la literatura, en el cine, y en la imaginación de muchos ciudadanos 'ascensores asesinos', 'ordenadores malvados', 'máquinas que se enamoran' y otras cosas por el estilo; o encontramos, sin caer en truculencias semejantes, con visiones espiritualistas del futuro como la de R. Jastrow, para quien no sólo el siguiente eslabón evolutivo al cerebro del hombre serán los ordenadores, sino que a través de su concurso podremos alcanzar la inmortalidad y

[24] A. M. Turing, *op. cit.,* págs. 41 y ss; y J. R. Lucas, *op. cit,* págs. 90 y ss.

solucionar de paso los males del mundo, ya que en tal ascenso evolutivo podremos zafamos de nuestro cerebro reptiliano[25]. Pero, ¿tiene todo esto algún fundamento? En mi opinión, no.

Existe respecto al concepto de "emergencia" una notable confusión. Es evidente que los sistemas físicos y los sistemas biológicos llegan, en función de sus complejidades estructurales, a situaciones críticas que dan lugar a emergencias de propiedades que anteriormente no existían. Ahora bien, una "propiedad" no es una "entidad", ni la emergencia de una propiedad tampoco es lo mismo que la emergencia de una entidad. Por ejemplo: no se trata de la misma emergencia cuando decimos que de once hombres con pantalón corto "emerge" un equipo de fútbol, que cuando afirmamos que de Antonio y Adelina "emergió" Adelineta. No obstante, siendo emergencias de tipo distinto, en ambos casos se da una continuidad ontológica entre lo que emerge y aquello que, digamos como sustrato, lo posibilita o produce. Muy al contrario ocurre con la tesis que defiende la emergencia de la autoconsciencia para las máquinas: aquí, como si de una creación *ex novo* se tratase, se defiende que lo que emerge es una nueva propiedad o entidad distinta y discontinuo de aquel sustrato que la ha posibilitado, cuando en realidad, y como acabamos de decir, ocurre lo contrario: que lo que "emerge", en absoluto, no es indiferente del sistema que lo produce, ya que esa propiedad emergente es como es en función de su sustrato material. De otra manera: los estados mentales y la autoconsciencia no son algo distinto y separado del cerebro, ya que, como diría Searle, "están causados por" el cerebro y "realizados" en el cerebro mismo.

Pero no olvidemos en este punto el hecho de que somos organismos biológicos socializados. No creo que lo que llamamos estados mentales ni lo que llamamos

---

[25] R. Jastrow, *El telar mágico* (1981), Salvat Ed., Barcelona, 1988.

autoconsciencia sean propiedades que un cerebro pueda poseer de suyo. En su aparición, aparte del necesario sustrato material, también interviene la interacción natural y social que dicho organismo mantiene con el medio que le rodea. Es un truismo de la psicología evolutiva que el desarrollo de las capacidades y destrezas mentales de un individuo en buena medida depende de la eficacia del proceso de socialización a que ha sido sometido. Y no sólo esto sería cierto a nivel ontogenético, también al nivel de la especie humana parece razonable decir que el cerebro tiene una historia evolutiva, y que los estados mentales y la autoconsciencia son logros evolutivos de la humanidad en su adaptación natural, esto es, una de las maneras como la especie a la que pertenecemos ha resuelto, de momento exitosamente, sus problemas de supervivencia.

Todo ello falla en el caso de los ordenadores. Una vez que comprobamos la dependencia causal en que se encuentran las propiedades emergentes respecto a sus sustratos materiales, resulta claro que un ordenador no puede causar o hacer emerger una mente. Carece del potencial causal que tienen nuestros sustratos materiales -nuestros cerebros- y carecen igualmente de la historia evolutiva y social a que nuestros cerebros han estado sometidos. Para decirlo siguiendo la sugerencia de R Ziff: existe una diferencia esencial e infranqueable entre la mente humana y los ordenadores[26]. Desde luego que en este punto alguien podría preguntar: *¿y si construimos un ordenador con material orgánico?* Entonces la respuesta sería ésta: si construimos un cerebro humano y lo socializamos, entonces ya no estamos hablando de ordenadores sino de seres biológicos. En ese caso habríamos realizado una "duplicación" y no una "imitación" de un ser humano, lo cual nada tiene que ver con la tesis de la I.A. Y además, no veo por ningún lado dónde está la gracia de construir cerebros; hay métodos más placenteros y reconfortantes de hacer seres humanos.

Para concluir: Creo que la filosofía no debe *ex abundantia cordis* alimentar las fantasías de los hombres; de eso ya se ocupan, y lo hacen excelentemente, la literatura y el cine. La filosofía, por el contrario, debe conformarse con aclarar nuestros pensamientos. Y si se atreve a construir hipótesis, lo debe hacer dentro de lo que es empíricamente posible y no dentro de lo que no es lógicamente imposible. Por ello, y teniendo en cuenta las argumentaciones indicadas, encuentro desacertadas las tesis que defienden tanto la idea de que los ordenadores pueden llegar a tener mente, como aquella otra de que la mente humana sea esencialmente un programa informático. En este sentido alguien podría pensar que estas conclusiones son arrogantes o antropocéntricas, algo así como si yo no estuviese dispuesto, en el fondo, a compartir con otros seres -verbigracia, los ordenadores- el privilegio de poseer una mente. La cuestión, sin embargo, es muy distinta: no creo que poseer una mente sea ningún privilegio. Los seres humanos poseemos mentes de la misma manera que podríamos no poseerlas. Se trata de un hecho del que no caben valoraciones morales: es nuestro logro evolutivo natural y social, la respuesta que nuestra especie ha encontrado a las exigencias que natural y socialmente se le han planteado.

[26] P. Ziff, "El sentir de los robots" (1959), en A. R. Anderson (ed.), *op. cit., págs.* 151 y ss.