

J. H. MARTÍN

¿GÖDEL *VERSUS* EL MECANICISMO?:
UN EXAMEN DE LA TESIS DE LUCAS

§1

Conforme nos aproximamos al fin del milenio aumenta el número de matemáticos, ingenieros, neurofisiólogos, lingüistas y psicólogos que confían en que el diseño y posterior construcción de sistemas mecánicos, básicamente computadoras digitales, deberá brindar a la postre la clave definitiva para una explicación naturalista de las funciones comúnmente asociadas a la mente humana. La inteligencia artificial, un vocablo especialmente acuñado al efecto para agrupar esas investigaciones, ha terminado por obtener carta de ciudadanía entre personas de cultura media y un nuevo campo de investigación, cuyo objeto es el estudio de los procesos internos de la mente mediante la simulación del comportamiento inteligente a través de artefactos (o, en su defecto, de modelos) computacionales, se desarrolla aceleradamente como trasfondo de ese fenómeno. El reconocimiento de patrones, el aprendizaje a partir de nueva información, la adaptación a circunstancias novedosas y la decisión ante cursos de acción alternativos para la solución de determinados problemas constituyen, *inter alios*, tan sólo una muestra de las operaciones sobre las cuales parece haber arrojado luz la construcción y el análisis del comportamiento de mecanismos computacionales especialmente diseñados para semejantes propósitos. En tales investigaciones confluyen de forma novedosa, y por lo demás provechosa, la lógica matemática, la ingeniería cibernética, la neurofisiología y la psicología de los procesos cognitivos.

Los filósofos no han podido sustraerse, afortunadamente, al impacto de una innovación que ha llegado a ser incluso calificada por los más entusiastas como "la última revolución en filosofía"¹. Ello en modo alguno significa, obviamente, que entre los filósofos no se hallan levantado también las voces escépticas: bastará recordar, simplemente a título de muestra, los

¹ Cfr., e.g., Sloman, A., *The Computer Revolution in Philosophy. Philosophy, Science and Models of Mind*, Sussex, The Harvester Press, 1979, pp. 3 ss.

trabajos de H. Dreyfus y J. R. Searle². Tales discrepancias, por lo demás, a nadie deberían extrañar, pues las investigaciones en inteligencia artificial son en parte responsables del replanteo de cuestiones filosóficas de vieja prosapia, como el problema relativo al esclarecimiento de las relaciones existentes entre los eventos de orden mental y los fenómenos físicos o neurofisiológicos que con ellos se correlacionan, cambio que parece haber terminado por desplazar la cuestión mente/cuerpo de los predios de un materialismo inspirado en el reduccionismo neurofisiológico, el fisicalismo o el conductismo prevaleciente en la psicología hasta la década de los cincuenta hasta el dominio de la inteligencia artificial, terreno en el que se ha venido consolidando desde entonces una terminología funcionalista que amalgama elementos procedentes del mecanicismo y del materialismo con otros de índole claramente intencional.

Las posturas escépticas respecto a la asimilación de funciones de tipo mental a programas computacionales, así como la analogía subyacente entre computadoras y cerebros, no siempre encuentran su inspiración en prejuicios de naturaleza romántica. Dicho escepticismo, al contrario, en ocasiones afina sus raíces en la imprecisión del lenguaje comúnmente asociado a la interpretación de las investigaciones con simuladores, apuntando, *verbi gratia*, a la discutible adscripción a las máquinas de atributos u operaciones que empleados en una descripción de la conducta humana ordinariamente no presentan mayores complicaciones, a la no menos controvertida noción de "simulación" o en determinados resultados metamatemáticos que atañen a las limitaciones formales que se encuentran a la base de los sistemas simbólicos que integran el *software* de los mecanismos computacionales.

Entre las objeciones que proceden de la interpretación de resultados matemáticos bien entabados la más conocida, si no la más ampliamente debatida, es la defendida por el filósofo oxoniense J. R. Lucas a lo largo de varias publicaciones, entre las cuales la más conocida seguramente sea el artículo intitulado "Mind, Machines and Gödel" (1961), donde pretende probar que el teorema de Gödel "... proves that mechanism is false, that is, that minds cannot be explained as machines"³. El mismo Gödel, dicho sea incidentalmente, a juzgar por una conferencia ampliamente reseñada por Wang y dictada en Providence en 1951, no parece haber sido del todo ajeno

² Cfr. Dreyfus, H. L., *What Computers Can't Do. A Critique of Artificial Intelligence*. New York: Harper Row, 1972. De J. R. Searle pueden consultarse *Mind, Brain and Science*, Cambridge, Mass., Harvard University Press, 1984 y "Is the Brain a Computer Program?", *Scientific American* (Jan., 1990).

³ Lucas, J. R., "Minds, Machines and Gödel", reimpr. en A. R. Anderson, ed., *Mind and Machines*, N.J., Englewood Cliffs, Prentice Hall 1965, p. 43. Este artículo apareció, por vez primera, en *Philosophy* XXXVI (1961).

a consideraciones por el estilo⁴.

En general, las reacciones frente a este tipo de argumentos cubren un amplio espectro de posiciones, las cuales van desde la completa recusación del argumento en cuestión adoptada por Putnam, Benacerraf, Good y Hofstadter *inter alios*, pasando por el tibio reconocimiento que a la sazón le dispensara A. M. Turing antes de desestimarlos como verdadero obstáculo para el planteo de la construcción de programas inteligentes, hasta su completa aceptación por Nagel, Newman y Penrose o con ciertas reservas por parte de Wang y Webb⁵.

El propósito de esta exposición es simplemente llamar la atención sobre las complicaciones que lleva aparejada la pretendida pertinencia de los teoremas de limitación de Gödel en conexión con las restricciones adjudicadas, por esa vía de razonamiento, a la analogía entre cerebros/mentes y máquinas/programas y al tratamiento de la mente como un programa computacional y, de pasada, examinar argumentos como el presentado por Lucas, tomando en consideración el alcance de la demostración de Paris y Harrington a finales de los setenta, de la indecidibilidad de una extensión conservativa del sistema de Peano para la aritmética elemental del teorema de Ramsey sobre particiones infinitas⁶.

§2

El reconocimiento de que los modelos computacionales de la mente

- 4 Cfr., en el sentido aludido, Wang, H., *From Mathematics to Philosophy*, Londres, Routledge & Kegan Paul, 1974, pp. 324-326. También puede verse, del mismo autor, *Reflections on Kurt Gödel*, Cambridge, Mass., M.I.T. Press, 1987. El trabajo de Gödel citado por Wang aparece, bajo el título de "Some Basic Theorems on the Foundations of Mathematics and Their Implications", en Feferman, S. *et al.*, eds, *Kurt Gödel. Collected Works. Vol. III*, Oxford, Oxford University Press, 1995.
- 5 Cfr., a título de muestra, Putnam, H. "[Review] Nagel, Ernest & J. R. Newman, *Gödel's Proof*. New York: New York University Press. 1958", *Philosophy of Science* XXVII (1960); Benacerraf, P., "God, the Devil and Gödel", *The Monist* 51 (1967); Good, I. J. "Human and Machine Logic", *British Journal for the Philosophy of Science* XVIII (1967); Hofstadter, D., *Gödel, Escher, Bach. An Eternal Golden Braid*, New York, Vintage Books, 1979; Turing, A. M., "Computing Machinery and Intelligence", *Mind* LIX (1950); Nagel, E. & Newman, J. R., *Gödel's proof*. Londres, Routledge & Kegan Paul, 1958; Penrose, R. *The Emperor's New Mind. Concerning Computers, Minds and the Laws of Physics*, Oxford, Oxford University Press, 1989; Wang, H., *op. cit.*; Webb, J. C., "Metamathematics and the Philosophy of Mind", *Philosophy of Science* XXXV (1968).
- 6 Paris, J. & Harrington, L., "A Mathematical Incompleteness in Peano Arithmetic", en J. Barwise, ed., *Handbook of Mathematical Logic*, Amsterdam, North Holland Publ., 1977.

comportan limitaciones basadas en resultados de naturaleza formal, especialmente en logros como los obtenidos por Gödel, Church, Kleene y Turing, es prácticamente un lugar común entre los investigadores que laboran en inteligencia artificial. Es así como A. M. Turing, en un artículo ya clásico: "Computing Machinery and Intelligence" (1950), muestra como incluso una *máquina de Turing*, vale decir: una suerte de máquina universal dirigida mediante un conjunto de reglas recursivamente aplicables, capaces de identificar, ordenar y disponer los símbolos que corresponden a sus *input* y no sujeta a ninguna limitación material, confronta limitaciones en el sentido de que existen problemas relevantes para el *test de Turing* que ella no estará en capacidad de resolver o que resolverá incorrectamente. El cerebro humano no está sometido, según se sostiene habitualmente, a limitaciones por el estilo. *Ergo*: no es posible lógicamente (o, si así se lo prefiere, matemáticamente) construir un modelo computacional que represente ajustadamente al cerebro humano ni tampoco diseñar un programa de información que permita simular nuestras capacidades mentales⁷.

La clase de limitación en la que pensaba Turing, y la cual fuera agrupada por él bajo el rótulo de objeciones matemáticas al juego de la simulación, no era precisamente el tipo de limitación descubierto por Gödel. Esta última clase de limitaciones, acorde con él, no podría plantearse sin el establecimiento de un puente entre los resultados de Gödel y la constitución de una máquina que participe en el juego de la imitación. A decir verdad el requisito requerido por Turing es mucho más fuerte: "if Gödel's theorem is to be used we need in addition to have some mean of describing logical systems in terms of machines, and machines in terms of logical systems"⁸. La respuesta que ofrece Turing a la cuestión en ciernes es, pienso yo, inadecuada: según él, podemos construir máquinas cada vez más potentes, de modo que resulte absurda la pretensión de que el cerebro humano supere en inteligencia a todas la máquinas⁹. Una objeción matemática no se supera, indudablemente, construyendo máquinas: ello sería como pretender cancelar una paradoja sintáctica exhibiendo contraejemplos.

En todo caso, los programas computacionales operan sobre la base de sistemas formales y es razonable pensar que, precisamente por eso, las limitaciones de estos últimos sean heredados por los primeros. Este es el camino emprendido por E. Nagel y J. R. Newman en *Gödel's Proof* (1958), quienes, a raíz de la aceptación de la incompletitud de cualquier formalización consistente de la teoría elemental de los números concluyen que el resultado de Gödel permite sostener con firmeza que la estructura de la mente humana es mucho más compleja y sutil que cualquier máquina hasta el presente

7 Turing, A. M., *op. cit.*, pp. 15-16.

8 *Ibidem*.

9 *Ibid*, p. 16.

imaginada¹⁰.

Las críticas al argumento esbozado no se han hecho esperar. Según Putnam, por ejemplo, la base de este argumento es falaz, pues descansa en una incorrecta formulación de los resultados logrados por Gödel, pues lo que este demostró no es la incompletitud absoluta de la aritmética elemental (*i.e.*: un enunciado del tipo 'G es indemostrable'), sino su incompletitud relativa (*i.e.*: un enunciado del tipo "Cons(S) \supset G es indemostrable"). El punto radica en que no es posible demostrar "Cons(S)" sin el empleo de métodos transfinitos y, por ende, más potentes que los involucrados en la teoría elemental de los números, salvo que, por supuesto, el lenguaje formal en cuestión adoleciera de alguna inconsistencia, en cuyo caso el problema carecería de interés teórico¹¹.

Scriven, por su parte, señala un aspecto metamatemático involucrado en el argumento que no carece de importancia: si bien es cierto que cualesquiera que sean los axiomas y las reglas de inferencia proporcionados a una computadora siempre habría verdades que nunca alcanzaría mediante esos axiomas y reglas, no es cierto que al programar una computadora le brindamos una idea adecuada de la verdad matemática, por lo que la argumentación de Nagel y Newman, en resumidas cuentas, da por sentado la corrección del formalismo. En su opinión: "The Gödel's theorem is no more an obstacle to a computer than to ourselves. One can only say that mathematics would have been easier if the formalists had been right, and it would in that case be comparatively easy to construct a mechanical mathematician. They weren't and isn't. But just as we can recognize the truth of improbable formula by comparing what it say what we know to be the case, so can a computer do the same"¹².

No estoy seguro de que tales argumentos, aunque tienen el mérito de llamar la atención sobre aspectos involucrados en la argumentación de Nagel y Newman, resulten completamente contundentes. En primer lugar, pese a que el señalamiento de Putnam es correcto en sus detalles, dar por establecido la incompletitud de la aritmética elemental es algo completamente aceptable y jugar intelectualmente con el hecho de que no podemos arribar a una garantía absoluta de su consistencia también lo es: actualmente nadie, hasta donde mi conocimiento alcanza, tiene una idea clara de lo que podría ser una demostración absoluta de la incompletitud de la teoría elemental de los números. Cabe preguntarse, en segundo lugar, si el planteo del problema supone la aceptación del formalismo sugerida por Scriven. No está nada claro que así sea: el teorema de Gödel nada afirma sobre la verdad

¹⁰ Nagel, E. & Newman, J. R., *op cit.*, p. 10.

¹¹ Putnam, H., "Minds and Machines", en S. Hook, de., *Dimensions of Mind: A Symposium*, New York, Collier Books, 1961, p. 142.

¹² Scriven, Michael, "The Compleat Robot: A Guide to Androidology", en S. Hook, de., *op. cit.*, p. 125.

matemática, de hecho ni siquiera está del todo claro que constituya el acta de defunción del programa originario de Hilbert¹³, aunque si de la versión canónica del formalismo, sino sobre una clase de enunciados verdaderos no derivables en los sistemas de la aritmética elemental siempre y cuando este sistema sea consistente. De los teoremas de Gödel se desprende, eso sí, que verdad matemática y derivabilidad no coinciden extensionalmente. Sin embargo, en descargo de Scriven, los teoremas de Gödel tampoco afirman nada a propósito de la complejidad y posibles sutilezas de la mente humana en comparación a las máquinas.

Entre los argumentos instrumentados en la dirección señalada quizás el de mayor alcance filosófico y el más elaborado, o al menos el más conocido, sea el presentado por J. R. Lucas. Los resultados de Gödel, acorde con este autor, son relevantes para la refutación del mecanicismo y proporcionan un criterio para el trazado de una distinción nítida y tajante entre hombres y máquinas¹⁴. Necesario es reconocer que, al contrario de otros filósofos que parecen encontrar entre las consecuencias lógicas de los resultados de Gödel sobre la incompletitud de la aritmética elemental ciertas tesis filosóficas, haciéndose de esa forma culpables de un *non sequitur* al intentar extraer consecuencias materiales de resultados puramente formales, Lucas, pese a que en ocasiones es un tanto descuidado en este sentido, presenta su argumento como una suerte de argumento dialéctico, no matemático. y más precisamente como un esquema de refutación del mecanicismo¹⁵. Por 'mecanicismo' entiende Lucas aspectos muy dispares y cuya reunión bajo una sola categoría resulta, por decir lo menos, altamente controvertido: cualquier tentativa de simulación de un ser humano mediante un programa computacional que utilice cualquier lenguaje de programación (e.g.: BASIC), la representación de seres humanos mediante máquinas de Turing, cualquier descripción física que, en conjunción con la descripción de determinados parámetros ambientales, permita la descripción, de acuerdo con leyes naturales, de la conducta humana (e.g.: el cálculo de determinada operación aritmética). A veces, en lugar de mecanicismo, habla de determinismo, expresión que eventualmente nos remite a problemas éticos, sobre todo en *The Freedom of the Will*¹⁶. Según este punto de vista es esencial a la simulación de la conducta humana por computadoras una reducción de

13 La pertinencia del segundo teorema de Gödel para el programa original de Hilbert aparece discutida en detalle en Defleisen, M., "On Interpreting Gödel's Second Theorem", *Journal of Philosophical Logic* 8 (1979), *Hilbert's Program*, Dordrecht, D. Reidel Publ., 1986 y "On an Alleged Refutation of Hilbert's Program Using Gödel's First Incompleteness Theorem", *Journal of Philosophical Logic* 19 (1990).

14 Cfr. Lucas, J. R., *op. cit.*, pp. 56-57.

15 *Ibid*, pp. 46 ss.

16 Cfr. Lucas, J. R., *The Freedom of the Will*, Oxford, Clarendon Press, 1971

esta índole, por ejemplo, mediante la representación del pensamiento humano mediante un sistema formal completamente especificado.

A partir de aquí el argumento de Lucas procede resaltando que las limitaciones impuestas por la naturaleza de los sistemas formales involucrados en sistemas tan simples como los equivalentes a la aritmética elemental constituyen un límite infranqueable para cualquier programa computacional, pero no para un ser humano. El argumento de Lucas echa mano de una serie de términos bastante ambiguos y poco claros: 'asumir', 'conocer', 'aseverar', 'inferir', 'razón', relacionándose esta falta de precisión, sin duda, con el difícil tránsito de un sistema formal a las cuestiones conceptualmente más complejas relativas al mecanicismo, la simulación y tópicos afines. Empero, en aras de la discusión, puede concederse que el sentido general de su argumentación es mal que bien comprensible y que estas imprecisiones tienen que hacer con las reglas de correspondencia que le permiten endosar las limitaciones formales de los sistemas matemáticos a los límites de las computadoras que aspiren participar en el juego de la simulación.

Más complicado resulta aceptar las consecuencias de orden moral que el mismo Lucas extrae del hecho reseñado: "Since the time of Newton, the bogey of mechanist determinism has obsessed philosophers. If we were to be scientific, it seems that we must look of human being as determined automata, and not as autonomous moral agents... But now, though many arguments against human freedom still remain, the argument from mechanism, perhaps the most compelling argument of them all, has lost his power... now we can see that no mechanical model will be completely adequate, nor any explanations in purely mechanist terms. We can produce model and explanations... There is not arbitrary bound to scientific inquiry: but no scientific inquiry can ever exhaust the infinite variety of the human mind"¹⁷.

Uno no puede menos que concluir que en este salto de la insuficiencias de la concepción mecanicista de la mente al reino de la libertad en la explicación de la conducta humana se ha operado, sin mayor justificación, una deslizamiento que supone un cambio de lenguaje. Sea como fuere, este paso no hace más que obscurecer el principal punto en discusión, ya de por sí bastante obscuro: si es posible una explicación de las funciones de la mente humana bajo el enfoque computacional.

Cabe destacar, regresando a la cuestión principal, que la explicación de las funciones mentales puede plantearse, de forma lícita y consistente, mediante una serie jerárquica de sistemas computacionales, y no mediante un sistema particular, trascendiendo así las limitaciones de cualquier sistema individual. Este tendría que ser el planteamiento de las analogías en juego, aunque de este modo tampoco escapamos a las limitaciones establecidas por Gödel. Tampoco es del todo claro, necesario es reconocerlo, que el pensamiento humano se encuentre restringido a patrones puramente deductivos, pero en su estado actual la llamada lógica inductiva y los sistemas análogos

17 "Minds, Machines and Gödel", p. 59.

como las lógicas no monotónicas están muy lejos de un desarrollo formal aceptable.

§3

Siguiendo esta línea que apunta a una serie de sistemas, y no a un sistema en particular, probablemente susceptible de una ordenación jerárquica, podemos calibrar la significación del resultado de Paris y Harrington para el tema que nos ocupa.

Desde la publicación de los resultados de Gödel diversos matemáticos se abocaron a la búsqueda rigurosa de ejemplos de enunciados matemáticos indecidibles en la aritmética elemental y a demostrar que la inexistencia de demostraciones para enunciados como la conjetura de Golbach y el último teorema de Fermat se debía al hecho de que se trataba de enunciados indecidibles. Los intentos en relación a estos resultados, hasta el momento, habían fallado, pero hacia finales de la década de los setenta Paris y Harrington presentaron la primera demostración de un enunciado matemático indecidible en una extensión conservativa de la aritmética elemental de Peano, *i.e.*, una extensión predicativa de PA que permita la introducción de conjuntos en las fórmulas del sistema. Se trata esencialmente del mismo sistema construido por Peano, salvo que nos permite hablar directamente sobre los conjuntos definidos por los predicados aritméticos. El resultado mencionado corresponde a una versión del teorema de Ramsey sobre particiones infinitas en PA (de paso: el teorema de Ramsey sobre particiones finitas es demostrable en PA):

Teorema de Ramsey: Dado cualquier conjunto A, cualquier n y una partición X de $[A]^n$, existe un subconjunto infinito B de A tal que $[B]^n$ está contenido en una parte de la partición.

En símbolos:

$$[A]^n = (Y_1 \cup \dots \cup \dots Y_K) \supset [B]^n \subseteq Y_i \text{ para algún } i, 1 \leq i \leq K.$$

La demostración procede, dicho sea incidentalmente, por pasos canónicos y no utiliza el procedimiento de codificación de Gödel: simplemente se demuestra que $TR \rightarrow \text{Cons}(PA + E)$ y que $(PA + E) \rightarrow TR$.¹⁸

¿Qué podemos decir de este resultado que se relacione con la objeción de Lucas?. Nada hay en esta demostración que justifique un rechazo del mecanicismo, en cualquiera de las diversas acepciones empleadas por Lucas, ni tampoco nada místico ni misterioso que tenga algo que hacer "con la infinita variedad de la mente humana". Una máquina debidamente programada, por ejemplo, podría producir la demostración. Por ende, no es nada claro hasta qué punto los teoremas de Gödel constituyen una piedra de tranca al programa de explicación de la mente humana *via* simulación

18 Paris & Harrington, *op. cit.*, pp. 1136 ss.

mediante programas computacionales. El resto, que las máquinas no son completamente idénticas a los hombres, ya lo sabíamos con anterioridad a Gödel y no representa un problema de naturaleza matemática. Las hipótesis basadas en la analogías mente/programa computacional y cerebro/computadora son sólo conjeturas que sirven como guía a las investigaciones sobre la inteligencia, ficciones útiles para nuestro comercio con los hechos, como todo lo que nos ofrece la investigación científica.

J. H. MARTÍN

Universidad Central de Venezuela