RdL Revista de Libros

Las sombras de la mente

ROGER PENROSE

480 págs.

Crítica, Barcelona, 1996

Acerca de un mal uso del teorema de Gödel en la especulación sobre la mente

Hilary Putnam
1 marzo, 1997

Los dos últimos libros de Roger Penrose, *La nueva mente del emperador* y *Sombras de la mente*, han estado destinados a defender una nueva tesis filosófica sobre el problema de la mente. La tesis de Penrose podría resumirse en tres puntos básicos: 1) El teorema de Gödel implica la naturaleza no computacional de la mente humana. 2) Es necesaria una nueva física para explicar esta naturaleza no computacional de la mente. 3) Esta nueva física debe derivarse de la correcta teoría de la

cuantización de la gravedad. De las tres tesis sólo la primera es presentada en *Sombras de la mente* como una tesis probada matemáticamente, manteniendo las otras dos dentro de un tono más conjetural. Hilary Putnam, profesor de lógica matemática y filosofía de la universidad de Harvard, presenta en este artículo, escrito especialmente para *Revista de Libros,* una crítica convincente a las pretensiones de Penrose sobre la posibilidad de derivar rigurosamente, a partir del teorema de Gödel, la naturaleza no computacional de la mente. Este artículo se inscribe en una polémica, entre Penrose y Putnam, iniciada hace algunos meses en el *New York Review of Books,* y continuada en el prestigioso boletín de la *American Mathematical Society,* polémica similar a la suscitada por el filósofo de Oxford John Lucas, en el seno de la filosofía de la lógica hace algunos años, con argumentos similares a los de Penrose.

En 1961, el filósofo oxoniense John Lucas afirmó que podía probar que en nuestras mentes se producen «procesos no computacionales» (procesos que, en principio, no pueden ser realizados por un computador digital, aun en el caso de que supongamos ilimitada su memoria). Su prueba se basaba en un famoso teorema de la lógica matemática, «el teorema de Gödel». Concluía que nuestras mentes no pueden ser idénticas a nuestros cerebros o a cualquier sistema material (ya que, según Lucas, los sistemas materiales no pueden realizar procesos no computacionales), y por lo tanto que nuestras mentes son inmateriales. Ahora, Roger Penrose utiliza un argumento similar pero más elaborado para sostener que necesitamos realizar cambios fundamentales en la manera en que vemos tanto la mente como el mundo físico. Afirma también Penrose que el teorema de Gödel muestra que en nuestra mente deben producirse procesos no computacionales; pero en lugar de postular un alma inmaterial en la que tendrían lugar aquéllos, concluye que deben ser procesos físicos en el cerebro, y que nuestra *física* debe cambiar para dar cuenta de ellos. (Él es un experto en física matemática, y propone algunos de los cambios que considera necesarios, pero no muestra que sus propuestas explicarían realmente cómo funciona el cerebro cuando hace matemáticas.) El teorema de Gödel hacía referencia a lo que se puede probar en sistemas formales, pero es sabido hace tiempo que se puede reformular en términos de programas de computador: por ejemplo diciendo que si un programa para hacer matemáticas es consistente, entonces una de las cosas que ese programa no puede probar es *que* es consistente. Así pues, como observa Penrose, se sigue inmediatamente que ningún programa del que puedan *probar* los matemáticos que es consistente puede simular *toda* nuestra competencia matemática. Sin embargo, Penrose es consciente de que este hecho en sí mismo no excluye la posibilidad de que un computador del que *no pudiéramos* saber que es consistente, pudiera simular serlo. Parece, por ejemplo, posible que pueda probarse que son consistentes cada una de las reglas en que confía implícitamente un matemático humano, y que el computador genere todas y sólo esas reglas, pero el programa que el computador utilice sea tan complicado que nos resulte imposible *probar* que eso es lo que hace. Un programa que simulara el cerebro de un matemático idealizado podría muy bien consistir en cientos de miles (o millones, o billones) de líneas de código. ¡Imagínese como un volumen del tamaño de la guía telefónica de Madrid! Podríamos ser incapaces de entenderlo, o decir si su resultado consiste en pruebas matemáticas correctas. Para salvar este escollo, Penrose argumenta que no es creíble que un programa del que *no pudiéramos probar que es consistente* pudiera generar todas las matemáticas que somos capaces de hacer. Si Penrose está en lo cierto, dado que sabemos (por el argumento gödeliano) que ningún programa del que pudiéramos probar que es consistente podría hacerlo, podemos concluir que ningún programa puede ser responsable de nuestra competencia matemática

humana. Pero, ¿por qué piensa Penrose que no es creíble?

Penrose sostiene que si el programa fuera tal que no pudiéramos conocer su consistencia, sería «esencialmente dudoso». Sería extraño, afirma, que un programa para el que no lográramos ver ninguna justificación siempre condujera «de alguna manera milagrosa» a resultados que *pudiéramos* justificar. Pero esto tiene visos de plausibilidad solamente si el programa en cuestión es suficientemente simple como para que lo entendamos enteramente, y de hecho Penrose limita su discusión en este punto a programas que son «suficientemente simples como para apreciarlos de manera perfectamente consciente».

A partir de este momento Penrose abandona toda pretensión de hacer lo que originariamente afirmaba que iba a hacer, es decir aportar «una prueba clara y simple» de que la mente humana no puede ser simulada por un computador, y ofrece una serie de argumentos que no son en absoluto matemáticos, sino filosóficos. Desde luego, nada hay de malo en el argumento filosófico en sí mismo; pero presentar lo que en el mejor de los casos es un argumento filosófico discutible como si fuera un resultado matemático, que toda persona competente en las matemáticas relevantes debería aceptar, ofende a cualquiera que conozca la diferencia entre las matemáticas y la filosofía. Por supuesto, el uso por parte de Penrose de expresiones como «sería extraño que» y «de alguna manera milagrosa» descubre obviamente que no cuenta con una prueba matemática.

Son numerosos los argumentos filosóficos que Penrose lanza para llenar el vacío en su razonamiento matemático, y no están planteados de forma bien definida, pero el centro de su argumento parece ser el siguiente: supongamos que el cerebro de un matemático idealmente competente puede ser representado por un programa del tipo que he descrito (o por un sistema formal equivalente), un programa que genera todas y sólo las pruebas correctas y convincentes. El programa mismo, sin embargo, no sería susceptible de justificación formal, como hemos visto. Podemos, de hecho, asumir que el programa es demasiado largo para aprehenderlo conscientemente –tan largo como la guía telefónica de Madrid, como he dicho más arriba–. En tal caso, pregunta Penrose, ¿cómo podría la evolución haber producido este programa? ¡No puede ver ninguna forma, y utiliza este argumento de ignorancia para mostrar que podemos archivar la cuestión!

Pero la superioridad evolutiva en cuestión no es tan difícil de ver. (Y el problema acerca de que es difícil ver cómo un programa semejante podría desarrollarse en partes es una objeción estándar a toda explicación evolutiva –que se ha demostrado que es posible vencer en un caso tras otro, aunque los detalles, por supuesto, no pueden predecirse de antemano–.) En primer lugar, la evolución no debería haber mantenido seres inteligentes que tuvieran posibilidades de sobrevivir con esquemas de razonamiento que conducen a contradicciones en la práctica cotidiana. ¡Si siempre estuviéramos incurriendo en contradicción en nuestro razonar diario, estaríamos tan confundidos que no habríamos estado por aquí el tiempo suficiente como para transmitir nuestros genes! Por supuesto, hay partes de las matemáticas en las que han surgido contradicciones: por ejemplo en la teoría de los números transfinitos de Cantor era posible probar a la vez que existía y que no existía el más grande «número ordinal» transfinito. Lo que hicieron los matemáticos fue añadir a la teoría de conjuntos de Cantor ciertas restricciones (más o menos *ad hoc*), tales como la «teoría de los tipos» de Russell, para eliminar la paradoja (en lo que se nos alcanza). En resumidas cuentas, el razonamiento matemático de cada día es consistente –no habría surgido si no lo fuera– y cuando la matemática *sesuda* contiene

contradicciones simplemente «la amañamos». La consistencia de las matemáticas (en la medida en que es consistente) no es un milagro.

Creo que puedo anticipar las respuestas de Penrose a este argumento. Dirá, supongo, que la evolución podría explicar por qué desarrollamos un programa que es *consistente* al menos en sus partes elementales (como por ejemplo la teoría de los números enteros), pero que hubiéramos desarrollado un estilo de razonar que es *correcto,* es decir, que *se corresponde con la verdad platónica sobre los números,* sería un *milagro.* La alternativa que él sugiere es que en lugar de un programa enorme que se corresponde milagrosamente con la verdad platónica acerca de los números, tenemos algo muy diferente: una capacidad puramente física de *realizar operaciones no computacionales en nuestros cerebros.*

Nótese que este argumento de Penrose apelará *solamente* a aquellos que estén dispuestos a aceptar su visión realista metafísica de las matemáticas *y* su visión puramente materialista de nuestros cerebros (una peculiar combinación de visiones). Lo que yo diría es que su argumento ignora el hecho de que la interpretación correcta de nuestros conceptos matemáticos se sostiene sobre el papel que desempeñan en nuestra vida (como Ludwig Wittgenstein apuntó). La evolución podría desde luego habernos programado con un «sistema formal» diferente; pero imaginar que hemos sido investidos por la evolución de un sistema radicalmente diferente es lo mismo que imaginar que hemos sido provistos con una disposición para desarrollar *conceptos* diferentes. No es como si los significados de nuestras palabras estuvieran fijados *previamente* a lo que hacemos con ellos, y la evolución tuviera la tarea de proporcionarnos un programa para usar las palabras *con esos significados fijados de antemano;* la evolución simplemente tiene que darnos un programa que nos permita salir adelante en nuestras vidas. Si lo hacemos, nuestros conceptos matemáticos admitirán *alguna* interpretación bajo la cual lo que decimos es correcto. (Incidentalmente, esta última afirmación puede ser demostrada formalmente usando otro teorema de Gödel, el «teorema de *Completitud* de Gödel».)

Hay otras muchas objeciones filosóficas que podrían esgrimirse contra la línea argumental de Penrose en su totalidad. ¿Es realmente tan clara la noción de simular la actividad de un matemático? Acaso la pregunta acerca de si es posible construir una máquina que se comporte como se comporta un matemático humano *típico* es una pregunta empírica con sentido, pero un matemático humano típico *comete errores.* El resultado de un matemático real contiene inconsistencias (especialmente si imaginamos que él o ella van a verse forzados a probar teoremas incesantemente, como exige la aplicación del teorema de Gödel); así la cuestión de probar si la totalidad de su resultado es consistente ni siquiera surge. A esto replica Penrose que el matemático puede cometer errores, pero él o ella los corrige tras reflexionar. Esto es verdad, pero para simular matemáticos que a veces cambian de opinión sobre lo que han probado necesitaríamos un programa que también pudiera cambiar de opinión; existen programas así, pero no son del tipo al que se aplica el teorema de Gödel.

En las *Investigaciones filosóficas,* su obra maestra, Wittgenstein enfatizaba la importancia de distinguir entre lo que puede hacer una máquina real (o una personal real) y lo que puede hacer una máquina *idealizada* (o una persona idealizada). Sorprendentemente, Penrose escribe: «No me preocupa qué argumentos detallados podría ser capaz de seguir un matemático *en la práctica*». Por tanto, admite que está hablando sobre un matemático *idealizado* y no sobre uno real. Sería una gran proeza descubrir que cierto programa es el que sigue el cerebro de un matemático real; pero sería

muy distinto descubrir que un programa es el que seguiría el cerebro de un matemático *idealizado.* Confundir estas cuestiones es perder de vista la *normatividad* de la pregunta: ¿cómo es la matemática *ideal*? Penrose teme que si decimos que nuestro resultado matemático idealizado no es el resultado de un proceso físico que podemos describir, entonces nos veremos forzados a concluir que hay algo sobre nosotros («la conciencia») que es «científicamente inexplicable», pero ésta no es una preocupación razonable. La descripción de una *práctica normativamente ideal* en esta área o en cualquier otra difícilmente constituye un problema para la física.

(Traducción de M. Carmen González Marín)