

PLATONISMO (MATEMÁTICO):
diferentes tipos, cómo Roger Penrose lo entiende y lo usa
como argumento en contra de la Inteligencia Artificial *fuerte*

Daniel Heredia González. Universidad de Sevilla

Resumen: Roger Penrose es para muchos uno de los pensadores interdisciplinares por excelencia en la actualidad. Su crítica a la postura de la Inteligencia Artificial *fuerte* constituye un bastión tenido en cuenta por la mayor parte de la comunidad dedicada al debate acerca de la IA. En este texto se recogerá uno de los rasgos más filosóficos del pensamiento del autor inglés: el platonismo. En el desarrollo de este trabajo veremos en qué medida Penrose hace una defensa de esta concepción con respecto a las matemáticas y cómo ello establece un punto a favor de su postura en el debate de la IA.

Palabras clave: computación, comprensión, conciencia, matemáticas, platonismo

Platonism (mathematical): different types, how Roger Penrose understands it and uses it as an argument against strong Artificial Intelligence.

Abstract: Roger Penrose is, to many people, one of the interdisciplinary thinkers *par excellence* today. His critic on the *strong* Artificial Intelligence is one of the most important in the scope of the debate about AI. This text contains one of the most philosophical features of his thought: Platonism. In this paper we will see to what extent Penrose makes a defense of this conception —regarding mathematics— and how to use it in favor of his position in the AI debate.

Keywords: computation, understanding, consciousness, mathematics, Platonism

Recibido: 1 de noviembre de 2019. **Aprobado:** 1 de junio de 2020.

Tú ves entonces, mi amigo, que este estudio ha de resultarnos realmente forzoso, puesto que parece obligar al alma a servirse de la inteligencia misma para alcanzar la verdad misma
(Platón sobre las matemáticas en *La República*, 1992: 355).

1. Introducción

La influencia de las ideas de Platón en el pensamiento occidental es innegable. Tal es su alcance y riqueza que una y otra vez el filósofo griego

vuelve a aparecer en escena. Es cierto que hoy en día no es tarea fácil encontrar una defensa del platonismo en todas las facetas en las que este está presente. Pero muchos encuentran aspectos en dicha filosofía que les son útiles para el desarrollo de distintos saberes. Un ejemplo de ello es la aceptación del platonismo dentro del ámbito matemático.

De sobra es sabido lo importante que son para Platón las matemáticas. Pero al hablar de «platonismo» en este artículo no se estará haciendo una referencia exclusiva a lo expuesto por el filósofo griego. De hecho, hoy en día, en el ámbito matemático, puede prescindirse de la referencia directa a Platón, históricamente hablando.

El platonismo que será tratado es el defendido por el físicomatemático inglés Roger Penrose. Cuando Penrose hace referencia al platonismo alude de manera directa al mundo de las ideas de Platón, pero sus ideas guardan ciertos detalles que le hacen perteneciente a un platonismo particular. El propósito de este artículo es analizar y situar el platonismo del autor de *La nueva mente del emperador*. Para ello, me he servido del texto que introduce por primera vez el concepto de platonismo en el contexto matemático, que no es otro que “Sobre el platonismo en matemáticas” de Paul Bernays.

Pero mejor que dar algunos detalles acerca de lo que tratará el presente texto, es exponer el esquema del mismo para tener una visión panorámica.

En primer lugar, veremos en qué contexto construye Penrose su esquema intelectual. Dicho contexto es el debate acerca de si es o no posible una inteligencia artificial. En este primer apartado veremos i) la aclaración de algunos conceptos fundamentales de dicho debate, ii) su postura en este (que será contraria a aquellos que defienden que sí es posible tal tipo de inteligencia) y iii) cómo entender de un modo adecuado el pensamiento matemático nos permite dar argumentos en contra de la posible inteligencia artificial.

En segundo lugar, nos encontraremos con el análisis del platonismo de Penrose. En un principio podremos observar la simpatía que siente Penrose con respecto a la idea originaria de Platón del mundo matemático. Pero poco a poco veremos que su pensamiento (por otra parte, no podría ser de

otro modo) tiene matices que lo diferencian de la posición del filósofo griego. Más tarde se introducirá el trabajo de Ferreirós (que sigue al de Bernays) acerca de los distintos platonismos, que nos servirá de patrón para ver a qué tipo de platonismo puede pertenecer el defendido por Penrose.

Y, por último, en las conclusiones ofreceré una respuesta general acerca del platonismo penroseano.

2. Contexto: debate de la Inteligencia Artificial y la postura penroseana

El platonismo de Penrose es un tema cuando menos interesante. El pensador inglés no dedica mucho tiempo a desarrollar a qué tipo de platonismo se considera perteneciente¹. Es más, cada mención a dicho tema no constituye una declaración precisa, aunque sí firme. El platonismo en sí mismo no ocupa un lugar central, en cuanto a extensión, en las explicaciones del autor inglés. Otra de las partes que lo hace interesante es que si no se tiene en cuenta este matiz en el pensamiento de Penrose es muy difícil comprender de forma adecuada sus planteamientos.

El escenario en el que se desarrolla el pensamiento más divulgado² de Penrose es en el debate entre los defensores de una Inteligencia Artificial y sus detractores, siendo nuestro autor perteneciente al segundo grupo. La postura de Penrose a este respecto puede ser considerada peculiar, ya que no responde a ninguna concreta de las que normalmente se sitúan al frente de la Inteligencia Artificial *fuerte* (que es contra la que se posiciona nuestro autor). El mismo Penrose se encarga en *Las sombras de la mente* de hacer una distinción entre los distintos bloques de pensamiento en relación a este debate, los cuales divide en cuatro categorías: A, B, C y D (Penrose, 2012: 26-31). Si bien los diferentes puntos de vista están explicados en las páginas

1 Para un análisis sobre los distintos tipos de platonismos véase Ferreirós (1999). La razón por la que menciono este texto en particular es porque será citado a lo largo de este trabajo y tendrá una relevancia manifiesta en el apartado 2, tal y como lo he expuesto en la introducción.

2 En referencia a sus obras *La nueva mente del emperador* y *Las sombras de la mente*, de las cuales será la segunda en la que me centraré en mayor medida.

citadas es conveniente definir brevemente a qué posturas se refiere Penrose cuando nombra a cada una de ellas. El punto de vista A es el conocido como IA *fuerte*, el B como IA *débil*, el C, que no tiene una definición concreta y el D corresponde al misticismo (Penrose, 2012: 26).

Nuestro autor se incluye dentro del punto de vista C. Esta perspectiva defiende que para una computadora es imposible llegar a tener las capacidades mentales propias del ser humano. Penrose se refiere más concretamente a la consciencia humana. La ciencia sería el campo de estudio que puede dar con la demostración de su postura. Entonces, parece que todo está solucionado. Obviamente no. Dicha demostración sólo podría tener lugar si se realiza una reforma en la ciencia actual, ya que es evidente que esta no da respuestas satisfactorias (al menos que le satisfagan a nuestro autor). Por tanto, tenemos que la reforma es una condición necesaria para que podamos tener constancia de la imposibilidad de pensamiento «humano» por parte de las máquinas (Penrose, 2012: 29-30).

Otro rasgo a tener en cuenta de C es que en él se parte de un cierto fisicalismo³, de tal forma que se entiende que la mente emerge como una «manifestación del comportamiento de ciertos objetos físicos» (Penrose, 2012: 30). Penrose basará parte de este tipo de afirmaciones en sus estudios acerca del cerebro humano, los cuales no se verán desarrollados en este presente trabajo, ya que son prescindibles para el asunto que aquí nos ocupa.

Una postura entendida de la forma en la que hasta ahora ha sido expuesta (en este caso la de Penrose, es decir, la C) inevitablemente se encuentra agrietada por casi todos sus pilares. Para evitar de alguna forma que el edificio se derrumbe veremos algunos de los conceptos que maneja Penrose para poder divisar de un modo más adecuado su punto de vista.

3 Penrose menciona la posible dificultad que puede entrañar declarar «fisicalista» al punto de vista C y por ello aclara que no utiliza este término con el fin de distinguirlo del resto de puntos de vista, sino porque desde la perspectiva de C la importancia de los objetos físicos para la emanación de la mente es crucial, mientras que, para otros puntos de vista, aunque puedan ser entendidos dentro del fisicalismo, no entienden del mismo modo este aspecto (Penrose, 2012: 31).

En el debate que estamos situados es imprescindible aclarar ciertos conceptos fundamentales para cimentar una base firme para el conjunto de argumentos tanto de uno como de otro bando. «Conocimiento», «comprensión», «consciencia» e «inteligencia» son cuatro de los más importantes a tener en cuenta en la disputa sostenida entre partidarios y detractores de la IA. Penrose da cuenta de ello. Nuestro autor tiene la opinión de que las tentativas de decretar definiciones completas o definitivas son un ejercicio poco recomendable, o al menos no son provechosas. Pero, por otra parte, encuentra necesario puntualizar a qué términos se está refiriendo él en concreto para, de esta forma, no dar lugar a malentendidos (Penrose, 2012: 52). En sentido estricto Penrose no se dedica a volver a definir tales conceptos, sino más bien los reconoce en su forma más general (con la intención de no entrar en una nueva polémica), para matizar las diferentes relaciones entre ellos, tal y como puede verse en las diferentes citas:

[...] Por ejemplo, mi propio uso del término «comprensión» implica ciertamente que una auténtica posesión de esta cualidad requeriría que esté presente algún elemento de *conocimiento*. Sin ningún conocimiento lo que trata cierto argumento no puede haber seguramente ninguna comprensión auténtica de dicho argumento (Penrose, 2012:52).

También me parecería irreprochable usar la palabra «inteligencia» sólo cuando puede haber alguna comprensión implicada. [...] En mi opinión, «inteligencia» sin comprensión es un nombre equivocado. [...] Así pues, mi uso del término «inteligencia» (especialmente cuando va acompañado de la palabra «auténtica») implicaría la presencia de algún conocimiento real (Penrose, 2012: 53).

En cuanto a «consciencia» nuestro autor siente la necesidad de acentuar que existen dos aspectos que es aconsejable tener en cuenta. Estos son el aspecto *pasivo* y el *activo*. Para entenderlos de un modo más claro, ya que no siempre es fácil, nuestro autor se apoya en ejemplos. A través de dichos ejemplos Penrose consigue establecer un esquema que ayuda a concebir de

un modo más claro las relaciones entre los conceptos. Los ejemplos son los siguientes:

[...] La percepción del color rojo, por una parte, es algo que ciertamente requiere consciencia pasiva, como lo es la sensación de dolor o la apreciación de una melodía. La consciencia activa está involucrada en la acción deliberada de salir de la cama, como lo está en la decisión voluntaria de dejar de hacer alguna actividad enérgica. El traer a la mente un recuerdo del pasado hace intervenir tanto el aspecto activo como el pasivo de la consciencia. La consciencia, activa y pasiva, estaría también normalmente involucrada en la formulación de un plan de acción futuro, y parece ciertamente que hay una necesidad de algún tipo de consciencia en el tipo de actividad mental que normalmente se englobaría en la palabra «comprensión». Además, podemos ser conscientes (pasivamente), hasta un cierto grado, incluso cuando estamos dormidos, siempre que tengamos algún sueño (e incluso el aspecto activo de la consciencia puede a veces empezar a jugar un papel en el momento en que nos despertamos) (Penrose, 2012: 55).

Tenemos entonces que *detrás* de la inteligencia debe haber comprensión y que *detrás* de la comprensión debe haber conocimiento⁴. La consciencia por su parte (tanto la activa como la pasiva, aunque existen casos en los que algunos matices no lo dejan del todo claro) parece estar relacionada con todo lo demás, siendo el eslabón principal de esta cadena de conceptos. Penrose sabe de la dificultad de entender de este modo el concepto de consciencia. Pero ello no será motivo para que nuestro autor no tome parte en el debate de la IA (siendo contrario a ella).

El modo de actuación de la máquina consiste en un procedimiento propio, y este es el de la computación. Penrose, siguiendo con su postura de no querer enmarañar la cuestión principal, entiende «computación» de tal forma que esta responde a la acción de algoritmos, que son sutiles instrucciones (aunque no en todos los casos) que siguen los dispositivos para llevar

4 Penrose lo explica en términos de requerir, con el siguiente esquema: a) «inteligencia» *requiere* «comprensión» y b) «comprensión» *requiere* «conocimiento» (Penrose, 2012: 54).

a cabo diferentes tareas. A lo largo de todo su estudio acerca de este tema (tanto en *La nueva mente del emperador* como en *Las sombras de la mente*) nuestro autor utiliza «computación» y «algoritmo» de manera indiferenciada, por lo que en este texto haré el mismo uso.

Habiendo quedado claro con qué tipo de conceptos expone Penrose sus argumentos, pasaremos a ver los procedimientos por los cuales nuestro autor los edifica.

2. La Inteligencia Artificial y las matemáticas

Concretamente en matemáticas, las computadoras son muy eficaces. Pero en lo que se refiere a «pensar» las matemáticas, el modo de proceder de las máquinas muestra limitaciones. Sin entrar en detalles de los problemas matemáticos que señala Penrose, sí que mencionaré algunos de ellos de forma panorámica para entender su posición.

Penrose trae a colación dos problemas que son solubles para el ser humano mientras que para las máquinas se presentan como irrealizables desde el quehacer propio de los algoritmos. Estos problemas son los pertenecientes a las *ecuaciones diofánticas* y el *problema de la teselación*⁵.

En primer lugar, tenemos el problema de las ecuaciones diofánticas, propuesto por el matemático alemán David Hilbert en 1900. Este problema en particular pertenece a una serie (toda ella propuesta por Hilbert) que hasta la fecha no tenía solución. El que nos presenta Penrose corresponde al décimo de esta serie y trata acerca de «encontrar un procedimiento computacional para decidir, dado un sistema de *ecuaciones diofánticas*⁶, si

⁵ Para una explicación más detallada véase Penrose, 2012: 43-48.

⁶ «Son ecuaciones polinómicas, con un número cualquiera de variables, en las que todos los coeficientes y todas las soluciones deben ser números enteros» (Penrose, 2012: 43). Este tipo de ecuaciones pueden tener o no tener solución y el problema de Hilbert consistía, como hemos visto, en buscar un algoritmo que estableciera qué ecuaciones tienen solución y cuáles no.

Como comentario adicional, cabe clarificar que los números enteros son aquellos que comprenden a 0, -1, 1, -2, 2, etc. Estos números son con los que trabaja exclusivamente los algoritmos.

las ecuaciones tienen una solución común» (Penrose, 2012: 43). La respuesta a este problema llegó en 1970 a través de los estudios del matemático ruso Matiyasévich (no de forma completa), quien concretó que no existe ese procedimiento computacional que pudiera dar respuesta a las ecuaciones diofánticas (Penrose, 2012: 44). Esta respuesta nos dice que el modo de obrar de las máquinas es inútil para encontrar una solución. Lo importante para el argumento en contra de la Inteligencia Artificial es tener presente que si una máquina no puede hacer frente a este problema es por falta de comprensión. Las máquinas no tienen esa relación especial que tiene el ser humano con las matemáticas. Un ser humano con herramientas matemáticas limitadas puede llegar a dar con la solución de un problema como el citado, porque puede ir *más allá* de las herramientas que posee. Pero si una máquina tiene la misma limitación que dicho ser humano, esta estará abocada a guardar silencio (con respecto a dar solución al problema) para el resto de los tiempos.

En cuanto al denominado *problema de la teselación*, las máquinas encuentran en el fondo las mismas dificultades que las que tenían que afrontar con respecto a las ecuaciones diofánticas. El *problema de la teselación* es un problema geométrico que enuncia la siguiente cuestión: teniendo un plano que contiene distintas formas poligonales, ¿se puede determinar si estas formas podrán teselar (cubrir, sin amontonamiento de las formas, sino de manera solapada) todo el plano⁷? Este es un problema que tiene solución. La respuesta a la cuestión es afirmativa, pero contiene un inconveniente para las máquinas: que su respuesta es alcanzada sólo a través de números reales. Esto supone una condición que dificulta, o más bien imposibilita, que las computadoras puedan hacer frente a este problema en particular. ¿En qué sentido? Ya vimos más arriba que las máquinas responden a algoritmos. Los algoritmos no tienen la capacidad de manejar números reales. Esta condición supone que resolver el problema de la teselación queda fuera del alcance de las máquinas. No tener la capacidad de poder

7 Debemos entender dicho plano en términos de geometría euclidiana.

trabajar con el sistema numérico necesario para ello (como hemos visto, el de los números reales) impide a las máquinas obtener dicha aptitud que sí es manifiesta en el ser humano. Al fin y al cabo, la diferencia fundamental entre humanos y máquinas reside en la capacidad de «pensar» las matemáticas que tienen los primeros, mientras que los segundos sólo (podría entrecomillarse) las «manejan».

Es obvio que Penrose está diciendo que en el pensamiento matemático muy posiblemente se encuentre la clave de la particularidad de la mente y la consciencia humana. Pero también se encarga de aclarar que, si lo percibe de ese modo, en mayor medida, es porque él mismo es matemático (Penrose, 2012: 433). De todos modos, ello no impide que esté convencido de que los útiles resultados que nos dan las matemáticas sólo están al alcance de estas. Es decir, que no quiere establecer un estatus, pero deja claro cómo sería este si lo hubiese.

El ser humano tiene una relación tal con las matemáticas que por mucho que avance la ciencia tecnológica jamás logrará alcanzar la sutileza con la que nuestra especie *piensa* las matemáticas. Que el modo de usar las matemáticas por parte de las máquinas pueda llegar a asemejarse (y mucho, hasta poder superarnos) no tiene la importancia que quieren darle los defensores de la IA. La *comprensión* de las matemáticas no es algo a lo que estas puedan aspirar. La idea que sostiene esta hipótesis penroseana es aquella que motiva el estudio de este texto. Ello es el platonismo que sostiene nuestro autor, el cual intentaré clasificar y desarrollar en el siguiente apartado.

4. El platonismo de Penrose

El platonismo de Penrose no es un rasgo de su pensamiento al que dedique una parte específica. Pero esta concepción tiene un peso significativo a la hora de querer llegar a entender de un modo oportuno su postura con respecto al debate acerca de la posibilidad de una Inteligencia Artificial. Lo que haré a continuación será mostrar aquellos textos en los que Penrose

habla explícitamente de su platonismo, para luego definir a qué tipo de platonismo se adhiere.

Veamos algunos pasajes de distintas obras de nuestro autor para comprobar que el asunto le persigue y tiene una importancia patente. Dichos pasajes pertenecen a obras tan dispares (en cuanto a hilo conductor de las mismas) como *Camino a la realidad*, *Lo pequeño, lo grande y la mente humana*, y a la obra que he estado manejando en la mayor parte de este estudio, *Las sombras de la mente*:

Los enunciados matemáticos que pueden pertenecer al mundo de Platón son precisamente aquellos que son objetivamente verdaderos. De hecho, yo consideraría que la objetividad matemática es realmente el objeto del platonismo matemático. Decir que una afirmación matemática tiene una existencia platónica es sencillamente decir que es verdadera en un sentido objetivo. Un comentario similar es aplicable a las *nociones* matemáticas —tales como el concepto del número 7, por ejemplo, o la regla para la multiplicación de números enteros, o la idea de que cierto conjunto contiene infinitos elementos— todas las cuales tienen una existencia platónica porque son nociones objetivas. En mi opinión, la existencia platónica es simplemente una cuestión de objetividad y, en consecuencia, no debería verse como algo “místico” o “acientífico”, pese que así la consideran algunos (Penrose, 2006: 58).

[...] Uno puede muy bien adoptar el punto de vista de que el *mundo platónico* contiene otras ideas, tales como la *Bondad* y la *Belleza*, pero aquí solo me interesaré por los conceptos platónicos de las matemáticas. Para algunas personas resulta difícil concebir que este mundo tenga una existencia independiente. Preferirán considerar los conceptos matemáticos como meras idealizaciones de nuestro mundo físico y, desde esta perspectiva, el mundo matemático se concebiría como algo que emerge del mundo de los objetos físicos.

Pero no es así como yo concibo las matemáticas, ni creo que la mayoría de los matemáticos y los físicos matemáticos tengan esa idea del mundo. Lo conciben de un modo bastante diferente, como una estructura gobernada de manera precisa y de acuerdo con leyes matemáticas intemporales. Por eso encuentran más

Platonismo (matemático) en Roger Penrose

apropiado considerar el mundo físico como algo que emerge del *intemporal* mundo de las matemáticas (Penrose, et. al, 2008: 14-15).

[...] El mundo de Platón es un mundo ideal de formas perfectas, distinto del mundo físico, pero en cuyos términos debe entenderse este mundo físico. También está más allá de nuestras imperfectas construcciones mentales; pese a todo, nuestras mentes tienen algún acceso directo a este reino platónico a través de un «conocimiento» de las formas matemáticas y nuestra capacidad para razonar sobre ellas. Encontraremos que, aunque nuestras percepciones platónicas puedan ser ayudadas a veces por la computación, no se limitan a la computación (Penrose, 2012: 66).

En las citas hemos podido ver que Penrose intenta esclarecer qué entiende por platonismo y con qué tipo de platonismo se identifica el suyo propio. Para nuestro autor, las matemáticas tienen una verdad objetiva, inamovible e indiscutible. Normalmente este tipo de planteamientos lleva a ser partidario de un platonismo al modo del filósofo griego, en el sentido de concebir la existencia de un mundo matemático aparte del mundo físico, estando el segundo subordinado al primero. ¿Es Penrose partidario de esta relación entre ambos mundos? En sus textos no se encuentran atisbos de negar esta condición de manera categórica, aunque sí que destaca algún matiz que otro que lo diferencia del modo de entender el mundo platónico dualista.

Uno de los aspectos más significativos que Penrose se encarga de poner encima de la mesa es saber entender de modo correcto este mundo platónico (de las matemáticas, que es al que se refiere en todo momento), esto es, evitando que se conciba de forma «mística» o «acientífica» (2008: 58). Esta puntualización asevera la dificultad que supone mantener una postura, como la que está sosteniendo Penrose, sin entrar en polémicas que pueden llegar a desacreditarle por falta de rigor científico.

Penrose entiende que las matemáticas tienen una verdad intemporal, que existen por sí mismas. Todo lo que podamos conocer de estas ya está previamente establecido en un mundo aparte. Y a este mundo aparte tiene

acceso el ser humano, pudiéndose distinguir, por ello, una característica especial en nuestra especie. Si hago mención al concepto de «mundo» es, precisamente, porque Penrose sostiene una teoría de tres mundos, al modo de Popper (aunque con sus pertinentes diferencias). Centrarnos en esta teoría no nos ayudará en exceso a comprender el platonismo de Penrose, y por ello es mejor proseguir sin entrar en detalles hasta más adelante.

Para tratar el asunto de los tipos de platonismos de un modo más adecuado he decidido tomar como guía el trabajo de José Ferreirós “Matemáticas y platonismo(s)” (1999). En el artículo de Ferreirós se tiene muy en cuenta el artículo del matemático Paul Bernays “Sobre el platonismo en matemáticas”, en el cual se acuña el término que nos ocupa. Obviamente no sería la primera vez que se utilizara el concepto platonismo. Sin embargo, sí que lo sería en el contexto del ámbito matemático.

En dicho artículo⁸ Bernays presenta el concepto de platonismo como respuesta a aquellos que proclaman una crisis fundacional en el campo de las matemáticas, a lo que el matemático suizo responde de manera negativa. Su defensa tiene como base que las ideas de Dedekind, Poincaré y Hilbert (por ejemplo) han sido sistemáticamente desarrolladas con gran éxito (1996: 258). Ello se debe en cierta medida a que el platonismo reina⁹ en las matemáticas (1996: 261). Pero para que la base del argumento sea firme y pueda seguir siéndolo¹⁰, Bernays considera que el platonismo que se debe seguir es un platonismo alejado del extremo (1996: 261).

Para saber a qué se refiere con platonismo extremo es necesario tener clara la clasificación de los distintos platonismos. Para ello seguiremos de cerca el texto de Ferreirós.

8 Aquí he manejado la obra de Putnam y Benacerraf *Philosophy of mathematics* (1996), la cual recoge (junto a muchos otros) el artículo de Bernays.

9 Entendámoslo en el contexto de Bernays, que es al que se refiere.

10 Expresado de tal modo puede parecer que Bernays no concede a los contrarios del platonismo ni un ápice de razón. Pero esto no es así. En el artículo se mantiene precisamente la idea de que cuantos más puntos de vista se tengan en cuenta más positivo será para el desarrollo de las matemáticas. En la página 264 se puede ver que concede a los intuicionistas parte de razón en lo que defienden.

La primera diferenciación a tener en cuenta, porque ello permite una esquematización general, es el platonismo *interno* (o *matemático*) y el platonismo *externo* (o *filosófico*).

Hablar de platonismo *interno* es hacerlo del modo de proceder de todas las matemáticas, ya que se entiende como el componente ideal de este campo. Este tipo de platonismo abarca todas las matemáticas, incluso las corrientes que niegan cualquier tipo de platonismo, como el constructivismo radical. Esta postura constituye uno de los polos de los grados de platonismo *interno*, siendo el platonismo extremo el opuesto (Ferreirós, 1999: 454). El planteamiento penroseano en principio no casa enteramente con ninguno de los dos tipos mencionados, aunque si hubiese que elegir entre uno de ellos obviamente tiene sentido colocarlo en el segundo grupo. Atendiendo al modo de entender el platonismo extremo, esto es, aquel que concibe un mundo ideal en el que toda la matemática está presente, no sería descabellado pensar que Penrose lo aceptara. Pero siendo justos sería mejor conocer el platonismo *externo* para, de ese modo, encontrar una afinidad más próxima (si es que la hay).

Por su parte, el platonismo *filosófico* o *externo* es aquel en el que los matemáticos defienden estar *desvelando* y trayendo al mundo físico los principios de este saber. A pesar de que esta actitud implica concebir la independencia a las matemáticas, no todas las variantes del platonismo (al menos reconocidas como tal) sostienen esa misma idea. Nos encontramos con una rama que mantiene que las matemáticas son abstracciones de nuestra experiencia (las figuras geométricas son idealizadas a partir de las formas que vemos en el mundo físico, etc.), no otorgando un mundo aparte, aunque sí una naturaleza diferente, en este caso abstracta¹¹. Para una observación más esquemática tenemos la definición de Ferreirós:

1. Platonismo *interno* o propiamente *matemático*: es característico de las teorías de la matemática abstracta o moderna, donde se hace referencia

¹¹ En el paréntesis anterior hago la aclaración de «reconocidas como tal» porque, como se puede ver en Ferreirós (1999: 458), esta rama suele considerarse platonismo cuando en el fondo está más cerca del aristotelismo.

a elementos cuya existencia se postula y se considera dada (se podría hablar de *existencia ideal*).

2. Platonismo *externo*, ontológico, o propiamente *filosófico* (una de las posibles interpretaciones filosóficas de la matemática, en particular de la característica antes señalada de la matemática abstracta): consiste en la afirmación de que los objetos matemáticos gozan de una existencia real, análoga en algún sentido (aunque diferente) a la existencia de los objetos físicos. (Ferreirós, 1999: 448).

Una vez aclarado cómo quedaría la clasificación pasamos a ver distintos tipos de platonismos.

Existe un platonismo que ha tenido una relevancia innegable en el ámbito de la filosofía de las matemáticas, y ese es el profesado por el matemático Georg Cantor. Su platonismo consistía en reconocer una dualidad de mundos o realidades, de las cuales una sería *inmanente* y la otra *transiente* (Ferreirós, 1999: 458), siendo la primera aquella que atañe a las matemáticas y la segunda de la que se tiene que hacer cargo la ciencia y la metafísica. Tenemos entonces que, por una parte, está lo que concierne a las matemáticas y que no ofrece dudas sobre su existencia¹²; y por la otra donde tampoco se ofrecen dudas sobre su existencia (ni su relación con las de la otra realidad), pero sí que supone un problema de gran calibre, ya que se tratan conceptos de una abstracción que difícilmente podemos determinar en el mundo físico (en el caso que se debate en el artículo trabajado se trata de los conjuntos transfinitos). Este platonismo se acerca mucho a un platonismo extremo, casi sin poder llegar a distinguirse de este de un modo claro. Pero existe una diferencia fundamental, esto es, que en el platonismo extremo existe la posibilidad de no querer ceder a otros campos el tratamiento de la realidad transiente. Mientras que para Cantor esta concepción es así necesariamente, ya que ello libra al quehacer de los matemáticos de problemas tales como las paradojas de conjuntos ordinales o de los alefs (Ferreirós, 1999: 459).

12 Entendida «existencia» como *real*, en su sentido más general.

Kurt Gödel también se mostró partidario de un platonismo en concreto. El motivo por el que también resaltamos el punto de vista de este eminente lógico matemático es, aparte de intentar localizar el pensamiento de Penrose, porque el trabajo de Gödel es muy importante para la construcción de los argumentos del pensador inglés en contra de la IA. Precisamente una de las aportaciones más trascendentales en el campo de las matemáticas por parte de Gödel, el teorema de incompletitud, constituye uno de los pilares esenciales del esquema penroseano¹³. Pero por ahora pasemos a ver el análisis el platonismo del lógico matemático.

La postura de Gödel se distingue de lo hasta ahora visto en que otorga «evidencia» (o «realidad») a una parte de la matemática y no a la totalidad de ella. Los números naturales y sus leyes, por ejemplo, serían privilegiados en esta concepción (Ferreirós, 1999: 460), dejando lo que de ellos surge

13 [...] muchos filósofos y defensores de la IA son de la razonable opinión de que, aunque el teorema de Gödel es indudablemente importante en su contexto original de la lógica matemática, puede tener consecuencias muy limitadas, en el mejor de los casos, para la IA o para la filosofía de la mente. Después de todo, muy poco de la actividad mental humana se dirige a cuestiones relacionadas con el contexto original de Gödel: los fundamentos axiomáticos de las matemáticas. Mi respuesta, por el contrario, es que una gran parte de la actividad mental humana supone la aplicación de la consciencia y la comprensión humana. Mi uso del argumento de Gödel consiste en demostrar que la comprensión humana no puede ser una actividad algorítmica (Penrose, 2012: 66-67). El modo en el que Penrose utiliza el teorema de Gödel como argumento en contra de la IA es interesante y lúcida. A grandes rasgos, el teorema de Gödel nos dice que para un sistema formal es imposible dar cuenta de sí mismo, ya que en su misma naturaleza no reside esta capacidad. Esta sentencia no es fruto de una opinión personal por parte de Gödel, sino que constituye una de las aportaciones más importantes al campo de la lógica y las matemáticas. Pero entrar en los detalles de la explicación de Gödel podría desviarnos en exceso. Penrose utiliza el argumento de Gödel del siguiente modo: la forma que tienen las máquinas de pensar y actuar es a través de algoritmos; los algoritmos son en última instancia sistemas formales; los sistemas formales no pueden dar cuenta sobre sí mismos (porque al querer hacerlo con un ejercicio de reflexión tan sólo conseguirían caer en un círculo vicioso); por tanto, las máquinas no pueden reflexionar, lo que lleva necesariamente a entender que las máquinas no piensan, al menos al modo en el que lo hacemos los seres humanos. De este modo (simplificado) es cómo utiliza Penrose el teorema de Gödel como argumento en contra de la IA. Para ver con mayor detalle los argumentos con los que Penrose legitima esta utilización véase la Primera parte de *Las sombras de la mente*.

(hipótesis, axiomas, etc.) en el plano de la convención. De todos modos, Gödel no considera que esa convención sea plena (Ferreirós, 1999: 460), llegando incluso a admitir posteriormente un platonismo más acorde a lo anteriormente expuesto (es decir, al extremo, aunque sin llegar a él). Sin embargo, es de destacar el primer matiz porque ello confiere una agudeza que permite entender el platonismo de Gödel de una forma singular. Esa singularidad está relacionada con el papel que tiene el pensar humano dentro de las matemáticas. Que Gödel se atreva a conferir una parte de las matemáticas a la convención¹⁴ hace manifiesto que encuentra en el quehacer humano un *algo especial* que lo relaciona a las matemáticas y a su desarrollo.

Ahora la pregunta pertinente es, ¿en cuál de los platonismos expuestos encaja mejor el de Penrose? Según lo visto arriba, el platonismo extremo no se aleja demasiado de las ideas de nuestro autor, mientras que los demás mostraban estar distantes en lugar de ser mínimamente acordes en sus diferentes matices (porque como platonismos, todos guardan una relación básica). Aun teniendo esto presente, lo cierto es que la postura de Penrose también guarda su particularidad.

La originalidad del planteamiento de nuestro autor reside y puede localizarse en el contexto en el que expone sus ideas, pero también guarda un peso significativo la concepción de realidad que se maneja con respecto a este asunto. Penrose llega a exponer un esquema en el que habla de tres mundos, que guarda cierta similitud (como vimos más arriba) con los de Popper¹⁵.

De forma breve, debemos saber que estos mundos son el físico, el mental y el matemático-platónico, siendo fácilmente identificable qué tipo de entes se encuentra en cada uno de ellos. Los tres mundos guardan tres misterios, que son: 1) el misterio que enlaza el mundo físico con una pequeña

14 Recordemos que entiende convención de un modo singular y no en su sentido estricto.

15 Como él mismo llega a reconocer, indicando también que su modelo no es idéntico al de Popper (Penrose, 2012: 433).

porción del mundo matemático-platónico, que funciona como su fundamento, 2) el misterio que enlaza el mundo mental con una pequeña parte de las estructuras físicas que funcionan como su sustrato físico y 3) el misterio que enlaza el mundo mental con una pequeña fracción del mundo matemático-platónico¹⁶.

Estos mundos están conectados unos con otros haciendo la realidad de cada uno de ellos más amplia. Nuestro autor entiende que esta disposición de los mundos puede resultar poco adecuada para un debate en el que se pretende poner las cosas en claro, en lugar de verse empantanado en otra discusión derivada de ella misma¹⁷. No obstante, esta primera intención no impide que sigan surgiendo debates, ya que su esquema es amplio, invitando a la discusión.

Rubén Herce (2014) encuentra un aspecto en la postura de Penrose, a propósito de esto que estamos viendo, que es interesante comentar. Penrose en su exposición de las ideas de los intuicionistas señala que no comparte con estos la dependencia de las matemáticas de la mente humana. Los entes matemáticos tienen existencia propia (como aclara en su teoría de los tres mundos). Pero también es cierto que reconoce un grado de construcción. Nuestro autor se apresura en puntualizar que esta construcción no se asemeja a la propia del intuicionismo, sino que este constructivismo tendría que ver con la matemática clásica. Sin embargo, como señala Herce, la relación entre el constructivismo (que defiende) y el platonismo matemático (que también defiende) no queda del todo clara (Herce, 2014: 63).

¹⁶ Estos tres misterios están citados de Herce (2016: 10).

¹⁷ Eduardo Badía Serra da cuenta de ello en un artículo (2008) acerca del pensamiento de Penrose, en el que dice: «Esta estructura de relación y dependencia entre los tres mundos plantea un problema fundamental, el de aclarar la relación que se establece entre ellos, y tres grandes misterios: El primero es el de la relación que existe entre las matemáticas y la física. El segundo es el de la relación entre el mundo físico y el mundo de la mentalidad. Y el tercer misterio: El de la relación entre el mundo mental y el de las matemáticas, esto es, ¿qué subyace a nuestra capacidad de acceder a la verdad matemática?». Apostillando que el mismo Penrose admite estos tres misterios y que además añade creencias propias (se refiere a ellas como prejuicios), advirtiendo de la complejidad que plantea el asunto (Badía Serra, 2008: 59).

Daniel Heredia

A partir de ahí, aquello que se quiera decir sobre esto pertenecería al plano de la conjetura. Por mi parte comparto la opinión de Rubén Herce, quien dice a propósito de esto lo siguiente:

[...] Según mi parecer la construcción matemática quedaría dentro del mundo mental: después de que la mente ha alcanzado los objetos matemáticos mediante la intuición, cabría la construcción (Herce, 2014: 63-64).

Dejando a un lado tal asunto, y siguiendo la intención de Penrose de no intentar complicar el debate, me detendré en un aspecto que puede arrojar luz sobre el modo de entender el platonismo por parte de nuestro autor, y ello es el desarrollo del conjunto de Mandelbrot.

Tal y como lo expone Penrose, el conjunto de Mandelbrot «tiene una estructura extraordinariamente elaborada, pero no se debe a ningún diseño humano. Lo realmente notable es que esta estructura está definida por una regla matemática particularmente simple» (2008: 58). En esta descripción existen dos puntualizaciones que es conveniente tener en cuenta: a) la forma elaborada que tiene el conjunto a pesar de provenir de una regla matemática de una simplicidad concreta y b) no es producto de un diseño humano. Con la primera de las características tenemos algo no muy diferente de otros problemas matemáticos. Es más, se podría decir que en la mayoría de los casos en los que entramos en cuestiones con un nivel de abstracción determinado, las matemáticas se desarrollan de este mismo modo (de formas simples a más complejas). Con respecto a la segunda, Penrose está exhibiendo su platonismo, pero de manera tal que con ello no podríamos detectar la particularidad de su pensamiento. Por ello pasemos a lo que dice concretamente acerca de este conjunto.

El conjunto de Mandelbrot se desarrolla a partir de una regla sencilla y a partir de ella se extiende alcanzando una «variedad infinita y complicación ilimitada». Tenemos entonces que una vez el conjunto empieza a desarrollar su potencial queda patente si no la imposibilidad de que este sea una

construcción humana, sí la sospecha de que ello sea posible de tal manera¹⁸. Para Penrose los procesos formales —que son contra los que nuestro autor dirige su crítica—, por muy completos y ricos que puedan llegar a ser, no tienen la capacidad de alcanzar el grado de derivación¹⁹ del conjunto de Mandelbrot en sus resultados. Lo que buscan los procesos formales es, precisamente, no extender sus pasos para evitar, así, caer en argumentos que los llevarían necesariamente a error. Las matemáticas que *manejan* las máquinas deben ser simples, mientras que las matemáticas que *pensamos* los humanos pueden llegar a ser muy complejas.

Cabe decir que, a pesar de que queda de manifiesto que nuestro autor otorga una independencia a las matemáticas, no se encuentra ningún indicio de que se niegue la capacidad del ser humano para tomar contacto con ellas (es más, es evidente que ese contacto existe y ipor eso hacemos matemáticas!). Se podría incluso decir que Penrose concede un mínimo de crédito al saber construir por parte de nuestra especie, siempre teniendo en cuenta que esa construcción no viene de la nada o simplemente del genio del ser humano, sino de esa existencia por sí misma que tienen las matemáticas.

5. Conclusiones

El platonismo de Penrose se asemeja al extremo y al de Gödel en su etapa posterior y se distingue de estos por su contexto (el debate de la IA). Su modo de entender las matemáticas ha traído no pocas críticas en su contra. Casi todas coinciden en que esta manera de defender tal punto corre el riesgo de entrar en un terreno demasiado metafísico con el que a pocas conclusiones en claro puede llegarse.

¹⁸ Penrose por su parte no lo sospecha, sino que está convencido de que no es una construcción.

¹⁹ Entendamos aquí a «derivar» en su sentido de «expandirse» y no a la derivación matemática.

Por ejemplo, el filósofo Mark Steiner, cree que el mismo Penrose no entiende de forma adecuada el platonismo. En su contribución a la obra editada por Herbert Breger, Emily Grosholz, *El crecimiento del conocimiento matemático* (2000), este filósofo defiende que Penrose, al poner el foco en el pensamiento matemático, hace una distinción entre los conceptos que Platón no hace en el planteamiento de su filosofía:

En realidad, dudo que Platón sea una buena fuente histórica para la visión de Penrose. Platón no hizo distinción entre los conceptos matemáticos y otros, lo que implicaría que los conceptos matemáticos tienen más realidad que otros. En todo caso, lo contrario era cierto: los conceptos matemáticos, teniendo un pie tanto en el mundo de los sentidos como en el mundo del intelecto, eran metafísicamente inferiores a aquellos conceptos que se aplican solo al mundo inteligible (en particular, «La idea del Bien»). Sin embargo, por la misma razón, Platón sostenía que los conceptos matemáticos eran una buena entrada al mundo inteligible, una cuña entrante en la metafísica (Steiner, 2000: 134).

De hecho, Steiner piensa que aquello que expone Penrose se corresponde de un modo más directo con la filosofía de Descartes que a la de Platón (Steiner, 2000: 134), ya que Descartes da prioridad a las matemáticas (concretamente a sus conceptos)²⁰. Este es el modo en el que Steiner concreta esa mayor similitud del pensamiento penroseano con el cartesiano en detrimento del platónico.

A pesar de que dentro de su análisis Steiner tiene razón, ello no impide que dicho análisis esté en un error. De acuerdo que Penrose pone el énfasis de su defensa en el pensamiento matemático. Pero a través de una lectura detenida no es posible ver que el matemático inglés quiera llevar a cabo una

20 Uno de los muchos ejemplos en los cuales Descartes expresa la idea del *estatus* de las matemáticas puede encontrarse en su *Conversación con Burman* (concretamente en la parte en la que habla sobre el *Discurso del Método*). En dicha *Conversación* dice: La matemática nos acostumbra a reconocer la verdad, porque en ella se hallan razonamientos correctos que no encontrarás en ninguna otra parte. Por lo tanto, quien haya acostumbrado su ingenio a los razonamientos matemáticos, también será apto para investigar otras verdades, porque la razón es en todas partes una y la misma (Descartes, 2011: 457).

distinción entre los conceptos matemáticos y los demás. Es más, Penrose sabe de este problema y no duda en dejar claro que no pretende entrar en un debate de tal naturaleza. Si nuestro autor se centra en las matemáticas no es porque crea en una superioridad plena, sino, más bien, porque este es su ámbito de estudio, es decir, aquel en el que se siente cómodo y en el que le resulta más fácil poder ver esa conexión con el platonismo. Así lo expresa casi al final de su obra *Las sombras de la mente*:

El propio Platón había insistido en que también debía atribuirse una realidad al concepto ideal de «lo bueno», y «lo bello» (cf. §8.3), igual que debe hacerse con los conceptos matemáticos. Personalmente, no desdeño una posibilidad semejante, pero no ha tenido una significación importante en mis reflexiones. Cuestiones de ética, moralidad y estética no han tenido una función relevante en mis exposiciones presentes, pero esta no es razón para desecharlas como si no fueran, de raíz, tan «reales» como las que he estado abordando. Evidentemente son cuestiones importantes independientes para considerar aquí, pero no han constituido mi interés particular en este libro (Penrose, 2012: 439).

Es decir, Penrose está apelando a ese refrán español que dice «zapatero, a tus zapatos». Intentar abordar otros conceptos supondría complicar una tarea que ya por sí misma es tremendamente complicada.

En mi opinión, la defensa del platonismo que hace nuestro autor cobra una mayor dificultad cuando intenta expresarlo en términos de realidad. Cuando Penrose apela a la idea de realidad como un concepto aclaratorio, lo que consigue es, precisamente, complicar más el asunto. Su teoría de los tres mundos divide la realidad, aunque su pretensión con ella es que se entienda a esta como un todo²¹. Esto, como digo, no deja de ser conflictivo.

21 Este punto de vista presenta un problema filosófico muy importante, este es, el debate entre monistas, dualistas y pluralistas. Muchos, como Javier Monserrat (Valls, 2012: 113-115), no dudan y declaran a Penrose como perteneciente al monismo. El motivo principal de esta sentencia reside en que, al final, el matemático inglés entiende (como hemos podido ver en el texto) la realidad como un todo, lo cual le sitúa dentro de esta corriente. En mi opinión, la postura de Penrose es más compleja. Si tuviera que hacer el ejercicio de tener

En esta parte soy de la misma opinión, de nuevo, que Rubén Herce, quien resalta que a pesar de lo complejo que resulta el esquema una vez se ha expuesto, Penrose lo deja todo en el aire:

Pero este esquema (el de los tres mundos) requiere una clarificación: es demasiado oscuro; no parece apto para entender la realidad como un todo; y de esta manera la única realidad es *causa sui*. Este último problema manifiesta una dificultad obvia, sin embargo, Penrose no vierte más luz en este esquema, pero insiste en decir que este esquema sólo intenta expresar sus *prejuicios* (Herce, 2016: 11).

¿Es un error el modo en el que Penrose plantea y afronta el problema? No seré yo quien cometa la imprudencia de categorizarlo como error. Primero, porque tal y como se plantea aceptar el platonismo que defiende es legítimo. Y, en segundo lugar, filosóficamente su modo de afrontarlo es indudablemente rico, ya que no pretende obtener una respuesta, sino, más bien, invita a seguir preguntándose sobre ello.

Lo último que me gustaría añadir sería que aquello que entiendo es que Penrose cree en ese platonismo propio, pero es consciente de que una defensa de ello lo lleva a dar unas explicaciones que por el momento no puede dar. Un reflejo de este callejón en el que se encuentra son las palabras con las que acaba *Las sombras de la mente*:

[...] Me he referido a tres mundos y los misterios que los relacionan mutuamente. Sin duda no existen realmente tres mundos sino uno, cuya naturaleza verdadera ni siquiera vislumbramos en el presente (Penrose, 2012: 443).

Bibliografía empleada

Arana, J. (2012), *Los sótanos del universo: La determinación natural y sus mecanismos ocultos*, Madrid, Biblioteca Nueva.

que clasificarlo diría que la postura que más le convence es la dualista. Pero, por otro lado, debido al embrollo que supone defender una postura como esa, prefiere ser cauto y se decanta por un monismo (que en todo caso no dejaría de ser uno muy débil). Para un estudio aclarador sobre el debate entre monistas, dualistas y pluralistas, véase Juan Arana (2015).

Platonismo (matemático) en Roger Penrose

Arana, J. (2015), *La conciencia inexplicada: Ensayo sobre los límites de la comprensión naturalista de la mente*, Madrid, Biblioteca Nueva.

Badía Serra, E. (2008), “Roger Penrose: Una aproximación elemental a su filosofía de la ciencia”, *Teoría y Praxis*, 13, pp. 53-80.

Bernays, P. (1996), “Sur le platonisme dans les mathematiques”, versión inglesa en P. Benacerraf, H. Putnam, *Philosophy of mathematics: Selected readings*, New York, Cambridge University Press, pp. 258-295.

Descartes, R. (2011), *Obras completas*, trad. por Ana Gómez Rabal, Madrid, Gredos.

Ferreirós, J. (1999), “Matemáticas y platonismo(s)”, *La Gaceta de la Real Sociedad Española de Matemáticas*, 2, pp. 446-473.

Herce Fernández, R. (2014), *De la física a la mente: El proyecto filosófico de Roger Penrose*, Madrid, Biblioteca Nueva.

Herce Fernández, R. (2016), “Penrose on what scientists know”, *Found Sci*, 21: 679.

Penrose, R. (1996), *La nueva mente del emperador*, trad. por José Javier García Sanz, México D.F., Fondo de Cultura Económica.

Penrose, R. (2006), *El camino a la realidad*, trad. por Javier García Sanz, Barcelona, Random House Mondadori, S.A. (Debate).

Penrose, R. (2008), (con A. Shimony, N. Cartwright y S. Hawking), *Lo grande, lo pequeño y la mente humana*, trad. por José Javier García Sanz, Madrid, Cambridge University Press, pp. 13-78.

Penrose, R. (2012), *Las sombras de la mente*, trad. por José Javier García Sanz, Barcelona, Crítica.

Penrose, R. (2014), (con S. Hameroff), “Consciousness in the universe: A review of the «Orch OR» theory”, *Physics of Life Reviews*, 11, 39-78.

Platón (1992), *República*, trad. por Conrado Eggers Lan, Madrid: Gredos, pp. 295-377.

Monserat, J., en F. Rodríguez Valls [Ed.] (2012), *La inteligencia en la naturaleza: Del relojero ciego al ajuste fino del universo*, Madrid: Biblioteca Nueva, pp. 101-118.

Steiner, M. (2000), “Penrose and Platonism”, en Emily Grosholz, Herbert Breger, *The Growth of Mathematical Knowledge*, Springer-Science+Business Media, B.Y, Volume 289, pp. 133-142.

Daniel Heredia González
Dani_hergon@gmail.com