
Universidad de Navarra
Facultad Eclesiástica de Filosofía

Rubén HERCE FERNÁNDEZ

La propuesta heurística
de Roger Penrose
En busca de las bases de la consciencia

Extracto de la Tesis Doctoral presentada en la
Facultad Eclesiástica de Filosofía de la Universidad de Navarra

Pamplona
2013

Ad normam Statutorum Facultatis Philosophiae Universitatis Navarrensis,
perlegimus et adprobavimus

Pampilonae, die 10 mensis decembris anni 2012

Dr. Ioannes ARANA
Dr. Paulus COBREROS

Dr. Henricus MOROS

Coram tribunali, die 21 mensis decembris anni 2012, hanc
dissertationem ad Lauream Candidatus palam defendit

Secretarius Facultatis
Sr. D. Eduardus FLANDES

Cuadernos Doctorales de la Facultad Eclesiástica de Filosofía

Vol. XXIII, n. 3

Presentación

Resumen: Este trabajo es una síntesis y un análisis crítico de los ensayos de reflexión epistemológica y ontológica publicados por Roger Penrose. Desde su filosofía, se exponen y se analizan los argumentos físico-matemáticos por los que sugiere la necesidad de un elemento no computacional que, dentro de una futura teoría de la gravedad cuántica, pudiera explicar la asimetría temporal del universo o el fenómeno de la consciencia. Una conclusión es que, si bien se observan aspectos no computacionales tanto en la actividad matemática como en el problema de medida de la mecánica cuántica, los argumentos de Penrose, y en especial el *nuevo argumento de Penrose*, no concluyen. Aun así, constituyen una propuesta heurística digna de consideración.

Palabras clave: gravedad cuántica, nuevo argumento de Penrose, consciencia

Abstract: This paper is an overview and critical analysis of epistemological and ontological essays published by Roger Penrose. Based on his philosophy, this work presents and analyzes the physical and mathematical arguments by which Penrose suggests the presence of a non-computational element. This element, within a future theory of quantum gravity, would give an explanation of universe's time asymmetry or the phenomenon of consciousness. One of the conclusions is that, while non-computational aspects are observed in both mathematical activity and measurement problem of quantum mechanics, the arguments of Penrose, and especially *Penrose's new argument*, are not sound. Still, they constitute a heuristic proposal worthy of consideration.

Key words: quantum gravity, Penrose's new argument, Consciousness

En el presente trabajo he pretendido realizar una síntesis y un análisis crítico del pensamiento de Roger Penrose. Con este objetivo me he centrado en los ensayos de reflexión epistemológica y ontológica que ha publicado en los últimos años. En ellos se recogen argumentos y puntos de vista matemáticos y físicos, en los que no falta una visión filosófica de la realidad. Además, su reflexión se ha sintetizado en una nueva propuesta heurística para la comprensión de la realidad física e incluso de la consciencia. A lo largo de esta publicación se irán mostrando algunas de sus sugerencias así como su visión de la relación entre las matemáticas y la física, o de estas con la consciencia y la libertad. Cada uno de estos puntos requeriría un trabajo *a se*, pero he preferido no atomizar el pensamiento del autor, sino analizarlo en su conjunto.

Este enfoque podría tener un punto débil si se busca un análisis exhaustivo; sin embargo, al tratarse del estudio de una propuesta heurística alternativa al paradigma vigente, he visto necesario acercarme a la obra ensayística de Roger Penrose en su totalidad.

* * *

A finales del siglo XIX la búsqueda de las leyes básicas de la naturaleza parecía casi finalizada. Los físicos presentaban un escenario ordenado y claro donde ensamblaban bien todos los elementos de la física conocida. Solo un par de *oscuras nubes* en el horizonte, como las llamó Lord Kelvin, hacían presagiar la *tormenta* que se avecinaba. Con el transcurrir del tiempo dichas *nubes* dieron lugar a las teorías de la relatividad y a la mecánica cuántica, modificando el concepto de universo que manejaban los físicos y presentando una «nueva física» comandada por esas dos grandes teorías. La física clásica se seguiría empleando como una adecuada aproximación a los objetos físicos cotidianos pero, a partir de ese momento, se abrirían nuevas líneas de investigación para explorar de nuevo el universo.

Durante las siguientes décadas se confirmó la asombrosa precisión de esas teorías y, con los conocimientos adquiridos, se desarrollaron infinidad de nuevos objetos de uso cotidiano. Sin embargo, ambas teorías todavía se resisten a ser comprendidas en su sentido último. De nuevo ante una física sólidamente afianzada aparecen algunas *nubes* en el horizonte que estimulan a buscar no solo nuevas teorías que funcionen sino una visión más profunda de la realidad. Se trata de anomalías que, por su relación con la comprensión global de la realidad, encuentran un reflejo en el ámbito filosófico, aunque en sentido estricto pertenezcan al ámbito científico.

Así por ejemplo, en la física clásica existían anomalías científicas en *dos pequeñas nubes oscuras*: el resultado negativo de la *experiencia de Michelson-Morley* y la *catástrofe del ultravioleta de Rayleigh Jeans*. Pero, a la vez, el mismo concepto de universo, como una Gran Máquina determinista, no engranaba bien con algunos de los argumentos filosóficos mejor trabados y con algunas de las experiencias más comunes, como el libre albedrío.

De modo análogo, algunas anomalías de las teorías físicas actuales tienen un reflejo en la comprensión filosófica de la realidad. Así sucede, por ejemplo, con el indeterminismo, el platonismo o la existencia de la libertad. La íntima conexión entre ciencia y filosofía también se aprecia en los

científicos que no se conforman con profundizar en el dominio técnico de la naturaleza, sino que se aventuran más allá de los esquemas científicos ortodoxos para explorar nuevos caminos en la búsqueda de la verdad última. Responden así al anhelo humano de conocer cómo son las cosas y no solo cómo funcionan. En este contexto filosófico de comprensiones globales es donde se sitúan tanto el presente trabajo como parte de la obra del físico-matemático inglés Roger Penrose.

La principal contribución científica de Penrose se sitúa en las nuevas perspectivas y técnicas geométricas que en los años sesenta impulsaron la investigación sobre la teoría de la relatividad. Aun así, su aportación no se reduce solo a esa célebre dimensión de su faceta profesional, sino que se le puede considerar un filósofo natural, en el sentido más clásico de la expresión. Penrose ha sabido relacionarse con una amplia variedad de temas físicos, matemáticos y filosóficos, desde la mecánica cuántica hasta la libertad. No obstante, algunas de sus contribuciones más estimulantes y originales son controvertidas y, en ocasiones, están fuera de la corriente principal de pensamiento¹. Por eso, no compartiré bastantes de las posturas de Penrose, ni pretenderé recoger todas las críticas que se le hacen. Me centraré en la búsqueda de los elementos más nucleares de la filosofía que subyace en sus planteamientos.

Personalmente siempre me ha atraído el conocimiento práctico de las cosas, como ingeniero y científico, y he procurado desarrollar un interés por conocer la verdad, como sacerdote y filósofo. Sin embargo, no puedo separar en mí ambos aspectos y, aunque unos sean más científicos y otros más vitales, todos están unidos en un conocimiento racional que libremente se confronta con la realidad para reforzarse o debilitarse. Me parece que este dinamismo de la razón se da tanto en la ciencia como en la filosofía o en la fe: todas estas dimensiones personales encuentran un punto de unión en su racionalidad. Es la persona humana con su racionalidad libre (científica y moral, teórica y práctica) la que busca la verdad objetiva y subjetiva. Desde mi punto de vista cada uno de estos binomios se relaciona inclusivamente con los otros dos, de tal modo que, por ejemplo, se puede hablar de una ciencia práctica subjetiva, de una moral teórica subjetiva o de una ciencia teórica objetiva. No pretendo ahora argumentar el porqué, pero me parece relevante señalar la importancia

¹ Valenti 2002, 131.

clave de esta racionalidad libre, unitaria y polifacética, porque sin ella resultaría difícil entender el presente trabajo.

Por otro lado, mi interés por las relaciones entre ciencia y fe, cultura y filosofía me ha llevado primero a realizar una tesis de licencia sobre la relación entre religión y cultura en Christopher Dawson y, después, a buscar un argumento para la tesis doctoral entre autores abiertos a las relaciones entre ciencia y fe. En este ámbito comencé leyendo algunas de las obras de Mariano Artigas y de Juan Arana para, posteriormente, orientarme hacia otros autores como John Polkinghorne, Paul Davies, Michael Heller, Douglas Hofstadter o Michael Ruse.

Durante estas lecturas y a través de algunas conversaciones con físicos, filósofos y teólogos *tropecé* con la obra de Roger Penrose. Desde el primer instante me atrajo el estilo de sus libros, donde las motivaciones y los argumentos partían de la ciencia y se desarrollaban con interés por conocer la verdad. Además, la obra de Penrose era la más físico-matemática de todas las que había leído y ese ir a las raíces sin dejar de lado la visión de conjunto me atrajo especialmente. Por último, y a pesar de su reconocido fisicismo, observaba en sus obras una apertura ante la filosofía, el sentido común y la libertad humana. Por lo tanto, se puede decir que fueron la actitud y el enfoque de los ensayos científicos de Roger Penrose los que me movieron a profundizar en su obra.

Esta motivación suponía también que no me podía centrar solo en un aspecto de su obra, sino que tenía que buscar la visión de conjunto. Sería necesario dejar de lado la valoración concreta de muchas de sus tesis para llegar a lo más nuclear. Ahí es donde comprendí la motivación heurística de toda su obra ensayística. Penrose pretendía sugerir, desde su punto de vista y con su experiencia científica, cuáles podían ser los caminos más viables hacia una nueva física que permitiese una comprensión más completa de la realidad, en la que cupiesen aspectos comunes de experiencia humana como la libertad. Mi tarea, por tanto, sería analizar su enfoque no tanto en sus motivaciones físico-matemáticas, cuya crítica queda en manos de los físicos y matemáticos, como en su comprensión de totalidad. Detrás de la obra de Penrose, había una comprensión filosófica de la realidad que sería el objeto de mi estudio.

Este enfoque requeriría describir las tesis del autor sin un excesivo y constante aparato crítico, para centrarme en el estudio del entrelazamiento de sus tesis –en Penrose todo conecta con todo– hasta alcanzar una comprensión global de su pensamiento. Se trataba de describir suficientemente bien

la amplia base de la pirámide para llegar a la cúspide. Lejos de pretender que cada apartado se pudiese leer como un artículo autónomo, no sería hasta las conclusiones donde el lector encontraría las valoraciones más jugosas y las críticas más sustanciales.

Con el presente trabajo no pretendo resolver problemas filosóficos de gran calado, como pueda ser el platonismo matemático o el indeterminismo cuántico, sino sacar a la luz los pros y los contras, las virtudes y los defectos de la aproximación de Penrose a algunos de esos problemas. Con este objetivo he intentado hacer sus razonamientos teóricos más accesibles al pensamiento filosófico a la vez que he valorado la profundidad e implicaciones de sus propuestas y he sugerido cambios de perspectiva donde sus fundamentos filosóficos me parecían más débiles.

Al afrontar temas que se mueven entre la física, las matemáticas y la filosofía, con frecuencia resultará que el lector más familiarizado con alguna de estas áreas encuentre facilidad de lectura o incluso una excesiva simplicidad en las afirmaciones sostenidas. Por otro lado y a la vez, es probable que le resulten arduos o carentes de suficiente explicación aquellos temas con los que se encuentre menos familiarizado. Soy consciente de estas posibles críticas, que asumo con gusto, ya que mi esfuerzo ha consistido más en una integración sistemática que en un análisis exhaustivo de cada tema. En esta línea he de agradecer las sugerencias que he recibido de filósofos, físicos o lógico-matemáticos para precisar el contenido de mis afirmaciones. A la vez, deseo recalcar que la zona intermedia donde se mueve este trabajo es de especial dificultad, así como de esencial utilidad para abrir horizontes de comprensión y para establecer puentes de comunicación. Siendo consciente tanto del encuadre como de la ambición del presente trabajo, procedo a desglosar el contenido de este resumen de mi tesis doctoral.

* * *

Para situar sucintamente al autor en su contexto personal y profesional, conviene señalar que desde sus primeros años destaca su interés por la visión de conjunto en los temas relacionados con la física, las matemáticas y la consciencia. El descubrimiento del *determinismo no local* en su estudio de los objetos imposibles, durante esta primera etapa, jugará un papel tan fundamental que influirá incluso en su modo de entender la libertad. Su determinismo se diferenciará del *determinismo local* de Einstein y se opon-

drá a una lectura filosófica del indeterminismo cuántico. Será significativa la claridad con que Penrose explicará que el principio de indeterminación de Heisenberg determina con unas probabilidades precisas dónde se puede encontrar una partícula.

En el primer apartado se mostrarán algunos presupuestos fundamentales de su esquema de pensamiento como su enfoque científico abierto a la filosofía y al sentido común, su platonismo matemático o su visión de la realidad como tres mundos entrelazados. En este primer momento, he preferido no criticar con excesivo detenimiento la pobreza filosófica de su esquema para evitar una prevención en contra de lo que se diga después. A mi parecer, una vez estudiados cada uno de esos mundos (físico, matemático y mental) en los sucesivos apartados, se podrá hacer una crítica desde dentro del sistema sin incidir demasiado en que los esquemas sostenidos por Penrose adolecen de una clara justificación filosófica.

Tras esta aproximación, en el segundo apartado se mostrará el acercamiento matemático de Penrose a las cuestiones sobre la consciencia. Se verá cómo, apoyándose en los teoremas de incompletitud de Gödel, Penrose critica a quienes piensan que los ordenadores pueden llegar a ser conscientes de un modo esencialmente similar a como lo son los hombres. El elemento central de su crítica consistirá en la afirmación de que en la consciencia humana se tiene que dar algún proceso no-algorítmico con su consiguiente substrato físico. Esta crítica realizada desde las matemáticas vendrá a engrosar las que se pueden hacer desde la filosofía o desde otros niveles de conocimiento ya que, a mi parecer, no tiene el calado que pretende justificar Penrose. Para comprender su justo alcance será necesario aclarar dos puntos que se refieren a las asunciones matemáticas que hace nuestro autor. Con este objetivo, me detendré en considerar si es necesario asumir el platonismo matemático como la única postura filosófica válida y si los presupuestos de su crítica tienen una fundamentación adecuada. Uno de los aspectos más complicados de este apartado ha sido discernir el alcance de las críticas hechas a Penrose, ya que el mismo autor se defiende de ellas en su libro *Shadows of the mind*.

Por último, en el tercer apartado, recogeré los elementos centrales expuestos hasta entonces para aproximarme a las sugerencias que hace Penrose en la búsqueda de un nuevo paradigma físico. En este caso explicaré cómo la necesidad de un elemento no-algorítmico, detectada por Penrose, podría estar en la base de algunos enigmas físicos como la *paradoja de la medida*. Llegados a este punto, expondré la teoría –sostenida por Penrose y, en su realización con-

creta, compartida por pocos— según la cual nuestra consciencia puede tener un reflejo en la mecánica cuántica. En este punto me abstendré de criticar una postura que me parece excesivamente simplificadora para dejar que la ciencia siga aportando los datos que ayuden a comprender mejor lo que puede estar sucediendo. Por último, mostraré la virtualidad de un esquema determinista en el que nuestro autor quiere dar cabida también a la libertad.

A lo largo del presente trabajo he intentado ceñirme al planteamiento de Penrose. Pero también, para alcanzar una comprensión más adecuada de lo que este autor expresa de distintos modos y en diversos lugares, he visto necesario confrontarlo con otras teorías físicas o encuadrarlo en algunos marcos filosóficos. Así ha sucedido, por ejemplo, al explicar su método científico. Se trata de un método que en muchas de sus facetas no está explicitado y, sin embargo, asume presupuestos o modos de razonar no suficientemente justificados que son relevantes para matizar el alcance de sus conclusiones. Aun así, he intentado hacer estas aclaraciones con respeto, sin forzar las afirmaciones más allá de lo que pueden decir y sin intentar situarlas en la teoría de ningún autor en concreto.

Espero que el presente trabajo contribuya a suscitar nuevas perspectivas y nuevos intereses sobre las relaciones entre ciencia y filosofía, física y libertad, matemáticas y consciencia, etc., al igual que la obra de Penrose los ha suscitado en mi persona.

Índice de la tesis

INTRODUCCIÓN	7
--------------	---

Capítulo I

Aproximación al contexto personal y profesional

1. Los primeros años de su vida	24
1.1. Retrato de familia	24
1.2. La herencia de un padre	27
1.3. Pasión por las matemáticas	32
2. Despliegue de su producción científica	35
2.1. Unos años intensos	35
2.2. Maduración de una vida	43
2.3. Reconocimiento a su trabajo	47
2.4. Balance general	52

Capítulo II

Líneas básicas de su filosofía

1. El método científico	60
1.1. La posibilidad del conocimiento	64
1.2. El punto de partida	66
1.3. El trabajo científico	69
1.4. Sentido común y apertura a la filosofía	76
1.5. Límites del método científico	86
2. La importancia de las matemáticas	88
2.1. El formalismo frente al realismo	90
2.2. La intuición directa como método de descubrimiento	97
3. El dinamismo de la realidad	102
3.1. Paralelismo entre matemáticas y realidad física	103
3.2. Selección natural	105
3.3. Principio antrópico débil	108

Capítulo III

Bases físicas del universo

1. Teorías de la relatividad especial y general	120
1.1. La relatividad galileana	120
1.2. El principio de equivalencia	121
1.3. La velocidad de la luz	124
1.4. Los conos de luz	125
1.5. Paradoja de los gemelos	127
1.6. La relatividad especial	130
1.7. La relatividad general	132
1.8. Singularidades en el espaciotiempo	134
2. Teoría estándar de la física cuántica	136
2.1. Dualidad onda-corpúsculo	138
2.2. Heisenberg y Schrödinger	141
2.3. El proceso de medida	142
2.4. Función de onda y entrelazamiento cuántico	145
2.5. Efectos EPR	148
2.6. El papel de la relatividad	152
2.7. Las interacciones débil y fuerte	154
2.8. Soluciones de tipo infinito	160
3. Termodinámica y asimetría temporal	163
3.1. Entropía	164
3.2. El big bang	168
3.3. Los agujeros negros	169
3.4. La baja entropía de los seres vivos	174
3.5. El modelo estándar de cosmología	176

Capítulo IV

Necesidad de una nueva física para entender la mente

1. El estatuto de las matemáticas	187
1.1. Tipos de realismo matemático	191
1.2. Una alternativa al realismo de Penrose	196
2. Relación entre computación y consciencia	200
2.1. Cuatro perspectivas: A B C D	201
2.2. El argumento de John Searle	204
2.3. El argumento de David Chalmers	206
2.4. El argumento «científico»	208
3. Procedimientos de computación	210
3.1. Top-down y bottom-up	212
3.2. Caos	214
3.3. Aleatoriedad	216

ÍNDICE DE LA TESIS

4.	No-computabilidad en el pensamiento matemático	218
4.1.	Gödel y el programa de Hilbert	219
4.2.	Turing y el problema de decisión	221
4.3.	El nuevo argumento de Penrose	223
4.4.	Alcance del argumento	227
4.5.	Conclusiones	231
4.6.	Necesidad de un elemento no-algorítmico	238

Capítulo V

Una aproximación heurística para un nuevo paradigma científico

1.	Hacia la gravitación cuántica	247
1.1.	Ecuaciones dinámicas y condiciones de contorno	250
1.1.1.	En los agujeros negros	250
1.1.2.	En la singularidad inicial	254
1.2.	Determinismo y probabilismo	259
1.2.1.	Formalismo matemático subyacente	259
1.2.2.	Ontología de la mecánica cuántica	263
1.3.	Gravedad, reducción del estado cuántico y asimetría temporal	268
1.4.	Elementos no algorítmicos	275
2.	El fenómeno de la consciencia	280
3.	Hacia la base física de la consciencia	288
4.	Dificultades de la propuesta	293

CONCLUSIONES	297
--------------	-----

BIBLIOGRAFÍA	319	
1.	Fuentes	322
1.1.	Libros	322
1.2.	Artículos	323
1.3.	Otros autores	328
2.	Bibliografía secundaria	331

Bibliografía de la tesis*

1. FUENTES

1.1. Libros

- PENROSE, R. (1994a): *Shadows of the mind: a search for the missing science of consciousness*, Oxford, Oxford University Press.
- (1999a): *Lo grande, lo pequeño y la mente humana*, Madrid, Cambridge University Press. Traducción realizada por Javier García Sanz de *The large, the small and the human mind*, Cambridge, Cambridge University Press, 1997.
- (1999b): *La nueva mente del emperador*, Barcelona, Grijalbo Mondadori. Traducción realizada por Javier García Sanz de *The emperor's new mind: concerning computers, minds, and the laws of physics*, Oxford, Oxford University Press, 1989.
- (2006): *El camino a la realidad: una guía completa de las leyes del universo*, Barcelona, Debate. Traducción realizada por Javier García Sanz de *The Road to Reality: A Complete Guide to the Laws of the Universe*, London, Jonathan Cape, 2004.
- (2010a): *Ciclos del tiempo: una extraordinaria nueva visión del universo*, Barcelona, Debate. Traducción realizada por Javier García Sanz de *Cycles of Time: An Extraordinary New View of the Universe*, London, Bodley Head, 2010.
- (2010b): *Roger Penrose: Collected Works*, Oxford, Oxford University Press. Recopilación en seis volúmenes de los artículos publicados por Roger Penrose hasta 2003.

* La bibliografía de este trabajo se ha dividido en dos partes. La primera parte hace referencia a las fuentes y la segunda a la bibliografía secundaria. En el apartado de fuentes se contiene tanto la obra ensayística de Roger Penrose, separada en libros y artículos, como algunas entrevistas o artículos directamente relacionados con su obra que han sido realizados por otros autores. Los artículos de Penrose no incluyen la totalidad de sus publicaciones, sino solo aquellas que se han citado o que tiene especial relevancia filosófica. Por otro lado, en la bibliografía secundaria tampoco se incluye la totalidad de las obras consultadas, sino solo las que se han citado y las que son más significativas para la comprensión global de los argumentos.

HAWKING, S. W. y PENROSE, R. (2010): *The Nature of Space and Time*, Princeton, Princeton University Press. Edición actualizada del debate entre ambos autores que tuvo lugar en la universidad de Cambridge, en 1994, y que se publicó por primera vez en 1996.

1.2. Artículos

- PENROSE, R. (1955): «A generalized inverse for matrices», *Proc. Camb. Phil. Soc.* 51: 406-413.
- (1957): *Tensor methods in algebraic geometry*, Cambridge.
- (1959): «The apparent shape of a relativistically moving sphere», *Proc. Camb. Phil. Soc.* 55: 137-139.
- (1974): «The role of aesthetics in pure and applied mathematical research», *Bull. IMA* 10: 266-271.
- (1979): «Singularities and time-asymmetry», en Hawking, S. W., Israel, W. (eds.), *General relativity an Einstein centenary survey*, Cambridge, Cambridge University Press: 581-638.
- (1987a): «Newton, quantum theory and reality», en Hawking, S. W., Israel, W. (eds.), *Three hundred years of gravitation*, Cambridge, Cambridge university Press: 17-49.
- (1987b): «Quantum physics and conscious thought», en Hiley, B., Peat, F. D. (eds.), *Quantum implications: Essays in honour of David Bohm*, London, Routledge: 105-120.
- (1988a): «Big bangs, black holes and ‘Time’s Arrow’», en *The nature of time*, Oxford, Blackwell: 36-62.
- (1988b): «Fundamental Asymmetry in physics laws. The mathematical heritage of Hermann Weyl (1987)», *Proc. Sympos. Pure Math.* 48: 317-328.
- (1988c): «On the physics and mathematics of thought», en *The universal Turing machine: A half-century survey*, New York, Oxford University Press: 491-522.
- (1989): «Minds, machines and mathematics», en Blakemore, C., Greenfield, S. (eds.), *Mindwaves*, Oxford, Blackwell: 259-276.
- (1990a): «Précis of the Emperor’s New Mind: Concerning Computers, Minds and the Laws of Physics», *Behavioral and Brain Sciences* 13 (4): 643-655.
- (1990b): «Author’s response – The nonalgorithmic mind», *Behavioral and Brain Sciences* 13 (4): 692-705.
- (1991): «¿Para qué sirve la mente?», *Revista de Occidente* (119): 175-182.
- (1992a): «The modern physicist’s view of nature», en Torrance, J. (ed.), *The concept of nature: The Herbert Spencer Lectures*, Oxford, Clarendon Press: 117-166.
- (1992b): «Setting the scene: the claim and the issues», en Broadbent, D. (ed.), *The simulation of human intelligence (Wolfson College Lectures)*, Oxford, Blackwell: 1-32.

- (1994b): «Mathematical intelligence», en *What is intelligence?*, Cambridge, Cambridge university Press: 107-136.
- (1995): «Must mathematical Physics be reductionist?», en Cornwell, J. (ed.), *Nature's imagination*, Oxford, Oxford University Press: 12-26.
- (1996a): «La conciencia incluye ingredientes no computables», en Brockman, J. (ed.), *La tercera cultura: más allá de la revolución científica*, Barcelona, Tusquets: 224-241.
- (1996b): «Space-time and cosmology», en Peterson, G. B. (ed.), *The Tanner lectures on human values*. Vol. 17, Salt Lake City, University of Utah Press: 347-391.
- (1997a): «The need for a non-computational extension of quantum action in the brain», en Arhem, P. (ed.), *Matter matters? On the material basis of the cognitive activity of mind*, Berlin, Springer: 11-27.
- (1997b): «On understanding understanding», *International Studies in the Philosophy of Science* 11 (1): 7-20.
- (1998a): «Can a computer understand?», en Rose, S. (ed.), *From brains to consciousness?*, London, Allen Lane: 154-179.
- (1998b): «Lionel Penrose: colleague and father», en Povey, S., Press, M. (eds.), *Penrose: Pionner in Human Genetics (Report on a symposium held to celebrate the centenary of the birth of Lionel Penrose, held 12th and 13th march 1998)*, London, Center for Human Genetics at University College: 4-8.
- (1998c): «Foreword», en Stachel, J. (ed.), *Einstein's miraculous year: five papers that changed the face of physics*, Princeton, Princeton University Press: vii-xiv.
- (1999c): «Por qué se necesita una nueva física para comprender la mente», en *La biología del futuro: ¿qué es la vida? cincuenta años después*, 1st ed. Tusquets editores: 161-180.
- (2000a): «Mathematical Physics of the 20th and 21st Centuries», en Arnold, V., Atiyah, M., Lax, P. D. y Mazur, B. (eds.), *Mathematics: Frontiers and perspectives*, Providence, American Mathematical Society: 219-234.
- (2000b): «Wavefunction collapse as a real gravitational effect», en Fokas, A., Kibble, T. W. B., Grigouriou, A. y Zegarlinski, B. (eds.), *Mathematical physics*, London, Imperial College Press: 266-282.
- (2001): «Consciousness, the Brain, and Spacetime Geometry: An addendum. Some new developments in the Orch OR model for consciousness», en Marijuan, P. C. (ed.), *Cajal and consciousness: Scientific approaches to consciousness on the centennial of Ramón y Cajal's textura*. Vol. 929, New York, Annals of the NY Academy of Sciences: 105-110.
- (2002a): «Consciousness, computation, and the Chinese room», en Preston, J., Bishop, M. (eds.), *Views into the Chinese room: New essays on Searle and artificial intelligence*, Oxford, Oxford University Press: 226-249.
- (2002b): «The rediscovery of gravity. The Einstein equation of General Relativity», en Farmelo, G. (ed.), *It must be beautiful*, London, Granta: 180-212.

- (2003a): «The problem of spacetime singularities: implications for quantum gravity?», en Gibbons, G. W., Shellard, E. P. S. y Rankin, S. J. (eds.), *The future of theoretical physics and cosmology*, Cambridge, Cambridge University Press: 51-73.
- (2003b): «The remarkable role of mathematics in the physical universe», en Wells, R. O. (ed.), *The founding of International University Bremen: Perspectives for the twenty-first century*, Bremen, International University Bremen: 98-109.
- PENROSE, R. y PENROSE, L. S. (1957): «A self-reproducing analogue», *Nature* 179: 1183.
- (1958a): «Puzzles for Christmas», *New Scientist* (December): 1580-1581, 1597-1598.
- (1958b): «Impossible objects: A special type of visual illusion», *British journal of psychology* 49: 31-33.
- PENROSE, R., WHITEHEAD, J. H. C. y ZEEMAN, E. C. (1961): «Imbedding of manifolds in Euclidean space», *Annals of Math.* 73 (2): 613-623.

1.3. Otros autores

- ALDISS, B. W. y PENROSE, R. (1999): *White Mars or the mind set free: a 21st century utopia*, New York, St. Martin's Press.
- ALFIERI, E. (2007): «Roger Penrose: “creo en un universo de ciclos sucesivos”», *Revista de occidente* 139: 118-131.
- ALONSO, E. (2001): «Mentalismo, mecanicismo: el nuevo argumento de Penrose», *Revista de filosofía* 26: 139-164.
- BÉJAR GALLEGO, M. (2008): «Physics, Consciouness and Transcendence: The Physics of Roger Penrose and David Bohm as Regards a Scientific Explanation of the Human Mind Open to Reality», *Pensamiento* 64 (242): 715-739.
- EASTWOOD, M. y PENROSE, R. (2000): «Drawing with Complex Numbers», *Mathematical intelligencer* 22 (4): 8-13.
- FEFERMAN, S. (1995): «Penrose's Gödelian argument», *Psyche* 2: 249-256.
- GARCÍA PRADA, O. (2000): «Interview with Sir Roger Penrose. Part 1», *EMS* (December): 17-21.
- (2001): «Interview with Sir Roger Penrose. Part 2», *EMS* (March): 12-17.
- GHERAB-MARTÍN, K. (2011): «La relación entre determinismo y libertad en la filosofía natural de Roger Penrose», en Rodríguez Valls, F., Diosdado, C. y Arana, J. (eds.), *Asalto a lo mental: neurociencias, consciencia y libertad*, Madrid, Biblioteca nueva: 81-88.
- GURZADYAN, V. G. y Penrose, R. (2010): «Concentric circles in WMAP data may provide evidence of violent pre-Big-Bang activity», <<http://arxiv.org/abs/1011.3706>> (submitted on 16/11/2010).
- (2011): «CCC-predicted low-variance circles in CMB sky and LCDM», <<http://arxiv.org/abs/1104.5675>> (submitted on 29/4/2011).
- HAMEROFF, S. y PENROSE, R. (1995): «What 'gaps'? Reply to Grush and Churchland», *J. Consc. Stud.* 2: 98-111.

- (1996a): «Conscious events as orchestrated space-time selections», *J. Consc. Stud.* 3: 36-53.
- (1996b): «Orchestrated Reduction of quantum coherence in brain microtubules: a model for consciousness?», en Hameroff, S. R., Kaszniak, A. W. y Scott, A. C. (eds.), *Toward a science of consciousness: The first Tucson discussions and debates*, Cambridge, MIT Press: 507-540.
- HAWKING, S. W. y Penrose, R. (1993): *Cuestiones cuánticas y cosmológicas*, Madrid, Alianza.
- HUGGETT, S. A. (1998): *The geometric universe: science, geometry, and the work of Roger Penrose*, Oxford, Oxford University Press.
- KRUGLINSKI, S. (2009): «The discover Interview: Roger Penrose», *Discover* 30 (8): 54-57.
- KUMAR, M. (2010): «Cycles of Time: An Extraordinary New View of the Universe by Roger Penrose – review», *The Guardian*, Saturday, 16 october 2010.
- LINDSTRÖM, P. (2001): «Penrose's New Argument», *Journal of Philosophical Logic* 30 (3): 241-250.
- (2006): «Remarks on Penrose's 'New Argument'», *Journal of Philosophical Logic* 35 (3): 231-237.
- MALDONADO, C. E. (1998): «El programa de Roger Penrose contra el dualismo», *Universitas Philosophica* 31: 31-54.
- MCKEMMISH, L. (2009): «Penrose-Hameroff orchestrated objective-reduction proposal for human consciousness is not biologically feasible», *Phys.Rev.E* 80 (2), <<http://link.aps.org/doi/10.1103/PhysRevE.80.021912>>.
- MONSERRAT, J. (1999): «Penrose y la mente computacional», *Pensamiento: Revista de investigación e Información filosófica* 55 (212): 177-216.
- PASCUAL, R. (1991): «Los sentimientos de un ordenador: sobre «The Emperor's New Mind», de Roger Penrose», *Saber leer* 45: 8-9.
- (1995): «Entender la inteligencia: sobre «Shadows of the mind», de Roger Penrose», *Saber leer* 87: 10-11.
- ROWE, D. E. (2008): «Remembering an Era: Roger Penrose's paper on "Gravitational collapse: the role of general relativity"», *Mathematical intelligencer* 30 (1): 27-36.
- SÁNCHEZ CAÑIZARES, J. (2011): «Reseñas. Cycles of time», *Anuario filosófico* 44 (2): 416-418.
- SÁNCHEZ RON, J. M. (1991): «El mundo de Roger Penrose», *Claves de razón práctica* 17: 58-63.
- (1993): «Roger Penrose», *Claves de razón práctica* 29: 48-55.
- SHAPIRO, S. (2003): «Mechanism, Truth, and Penrose's New Argument», *Journal of Philosophical Logic* 32 (1): 19-42.
- VALENTINI, A. (2002): «Book reviews. The Geometric Universe: Science, Geometry, and the Work of Roger Penrose», *Studies in history and philosophy of modern physics* 33 (1): 131-135.

2. BIBLIOGRAFÍA SECUNDARIA

- AERTSEN, J. (2003): *La filosofía medieval y los trascendentales: un estudio sobre Tomás de Aquino*, Pamplona, Eunsa.
- ALBRIGHT, J. R. (2000): «Cosmology: what one needs to know», *Zygon* 35 (1): 173-180.
- ARANA, J. (1988): «Naturaleza y razón: ciencia natural y filosofía de la naturaleza», *Thémata* 5: 9-31.
- (2000): «¿Es idealista la Interpretación de Copenhague de la mecánica Cuántica?», *Ontology Studies* 1: 185-202.
- (2001): *Materia, universo, vida*, Madrid, Tecnos.
- ARTIGAS, M. (1999a): *Filosofía de la ciencia experimental: la objetividad y la verdad en las ciencias*, 3ª ed., Pamplona, Eunsa.
- (1999b): *La mente del universo*, Pamplona, Eunsa.
- (2002): «Finalità», en Tanzella-Nitti, G., Strumia, A. (eds.), *Dizionario Interdisciplinare di Scienza e Fede*, Roma, Città Nuova: 652-664.
- (2008): *Filosofía de la naturaleza*, 5 ed., Pamplona, Eunsa.
- BARKER-PLUMMER, D. (2012): «Turing Machines», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2012 ed.
- BARROW, J. D. y TIPLER, F. J. (1989): *The anthropic cosmological principle*, Oxford, Oxford University Press.
- BENACERRAF, P. y PUTNAM, H. (1964): *Philosophy of mathematics: selected readings*, Oxford, Blackwell.
- BERG, J. M. (1998): «Lionel Sharples Penrose (1898-1972): aspects of the man and his works, with particular to his undertakings in the fields of intellectual disability and mental disorder», *Journal of intellectual disability research* 42 (2): 104-111.
- BERTO, F. (2009): *There's something about Gödel!: the complete guide to the incompleteness theorem*, Malden Massachusetts, Wiley-Blackwell.
- BISHOP, R. (2009): «Chaos», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2009 ed.
- CABA, A. (2000): «Aspectos metodológicos del naturalismo matemático. La aproximación conjuntista de Maddy», *Contrastes* 4: 5-23.
- ČAPEK, M. (1973): *El impacto filosófico de la física contemporánea*, Madrid, Tecnos.
- CHALMERS, D. J. (1997): *The conscious mind: in search of a fundamental theory*, Oxford, Oxford University Press.
- COLE, D. (2009): «The Chinese Room Argument», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2009 ed.
- CORCÓ JUVIÑA, J. (1995): *Novedades en el universo: la cosmovisión emergentista de Karl R. Popper*, Pamplona, Eunsa.
- CRUZ, J. (1992): «Filosofar hoy. Entrevista con Leonardo Polo», *Anuario Filosófico* 25: 27-51.

- DAVIES, P. C. W. (1992): *The mind of god: science and the search for ultimate meaning*, London, Simon & Schuster.
- DEL PONTE, M. (2006): *Realismo y entidades abstractas. Los problemas del conocimiento en matemáticas*, La Laguna, Universidad de La Laguna.
- DIÉGUEZ LUCENA, A. (2005): *Filosofía de la ciencia*, Madrid, Biblioteca nueva.
- EAGLE, A. (2012): «Chance versus Randomness», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2012 ed.
- FEYNMAN, R. P. (1985): *QED: the strange theory of light and matter*, Princeton, Princeton University Press.
- GAIFMAN, H. (2000): «What Gödel's Incompleteness Result Does and Does Not Show», *The journal of philosophy* 97 (8): 462-470.
- GHIRARDI, G. C., GRASSI, R. y RIMINI, A. (1990): «Continuous-spontaneous-reduction model involving gravity», *Phys.Rev.A* 42 (3): 1057-1064.
- GHIRARDI, G. C., RIMINI, A. y WEBER, T. (1986): «Unified dynamics for microscopic and macroscopic systems», *Phys.Rev.D* 34 (2): 470-491.
- GÖDEL, K. (1947): «What is Cantor's Continuum Problem?», *The American Mathematical Monthly* 54 (9): 515-525.
- HARRIS, H. (1974): «Lionel Sharples Penrose (1898-1972)», *Journal of medical genetics* 11 (1): 1-24.
- HARTVIG, P. y KJELSBORG, E. (2009): «Penrose's Law revisited: The relationship between mental institution beds, prison population and crime rate», *Nordic journal of psychiatry* 63 (1): 51-56.
- HELLER, M. (2003): *Creative tension: essays on science and religion*, Philadelphia, Templeton Foundation Press.
- HOD, S. (2008): «Return of the quantum cosmic censor», *Physics letters B* 668 (4): 346-349.
- HOFSTADTER, D. R. (2008): *Yo soy un extraño bucle*, Barcelona, Tusquets.
- HORSTEN, L. (2008): «Philosophy of Mathematics», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2008 ed.
- IEMHOFF, R. (2012): «Intuitionism in the Philosophy of Mathematics», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2012 ed.
- LAPIEDRA, R. (2008): *Las carencias de la realidad: la conciencia, el universo y la mecánica cuántica*, Barcelona, Tusquets.
- LAXOVA, R. (1998): «Lionel Sharples Penrose, 1898-1972: A personal memoir in celebration of the centenary of his birth», *Genetics* 150 (4): 1333-1340.
- LINDNER, I. (2008): «A Special Case of Penrose's Limit Theorem When Abstention is Allowed», *Theory and Decision* 64 (4): 495-518.
- LINNEBO, Ø. (2011): «Platonism in the Philosophy of Mathematics», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2011 ed.
- LIVI, A. (1995): *Crítica del sentido común*, Alcalá, Rialp.
- LOMBO, J. Á. y RUSSO, F. (2005): *Antropología filosófica: una introduzione*, Roma, Edusc.

- MADDY, P. (2007): *Second philosophy: a naturalistic method.*, Oxford, Oxford University Press.
- (2011): *Defending the axioms: on the philosophical foundations of set theory*, Oxford, Oxford University Press.
- McMULLIN, E. (1998): «Contingencia evolutiva y finalidad del cosmos», *Scripta theologica* 30 (1): 227-251.
- MILLER, L. (2001): *Roland Penrose, Lee Miller: the surrealist and the photographer*, Edinburgh, National Galleries of Scotland.
- MONSERRAT, J. (2005): «John Polkinghorne, ciencia y religión desde la física teórica», *Pensamiento: Revista de investigación e información filosófica* 61 (231): 363-393.
- MOORE, M. (2007): «Naturalism, Truth and Beauty in Mathematics», *Philosophia mathematica* 15 (3): 141-165.
- MOSTERÍN, J. y TORRETTI, R. (2010): *Diccionario de lógica y filosofía de la ciencia*, Madrid, Alianza.
- OPPY, G. y DOWE, D. (2011): «The Turing Test», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2011 ed.
- ORTIZ DE LANDÁZURI, C. (2010): «El infinito cuántico y relativista: Hacia una revisión de Plank y Einstein después de Bell (A través de Heisenberg, De Fineti, Ulam y Neumann, Chandrasekhar, Hawking y Penrose)», *Ontology Studies* 10: 101-114.
- PASEAU, A. (2010): «Naturalism in the Philosophy of Mathematics», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2010 ed.
- POLKINGHORNE, J. C. (2005): *Quarks, chaos & Christianity: questions to science and religion*, New York, Crossroad Pub. Co.
- (2007): *Quantum physics and theology: an unexpected kinship*, New Haven, Yale University Press.
- RIAZA, E. y SOLS, F. (2010): *La historia del comienzo: Georges Lemaître, padre del big bang*, Madrid, Encuentro.
- ROLAND, J. (2007): «Maddy and Mathematics: Naturalism or Not», *British Journal for the Philosophy of Science* 58 (3): 423-450.
- ROSENBLUM, B. y KUTTNER, F. (2010): *El enigma cuántico: encuentros entre la física y la conciencia*, Barcelona, Tusquets.
- RUSE, M. (2010): *Science and spirituality: making room for faith in the age of science*, Cambridge, Cambridge University Press.
- RUSSELL, B. (2009): *ABC of relativity*, London, Routledge.
- SCHRÖDINGER, E. (2000): *What is life?*, London, Folio Society.
- (2001): *La nueva mecánica ondulatoria y otros escritos*, Madrid, Biblioteca Nueva.
- SEARLE, J. R. (1980): «Minds, brains, and programs», *Behavioral and Brain Sciences* 3 (3): 417-457.
- SHAPIRO, S. (2000): *Thinking about mathematics: the Philosophy of Mathematics*, Oxford, Oxford University Press.

- SILVA, I. (2009): *Divine action in nature. Thomas Aquinas and the contemporary debate*, Tesis doctoral, Universidad de Oxford.
- SMITH, M. (1999): *Lionel Sharples Penrose: a biography*, London, M. Smith.
- SMITH, P. (2007): *An introduction to Gödel's Theorems*, Cambridge, Cambridge University Press.
- SMULLYAN, R. M. (1992): *Gödel's incompleteness theorems*, New York, Oxford University Press.
- STEINHARDT, P. (1996): «New perspectives on forbidden symmetries, quasicrystals, and Penrose's tilings», *Proceedings of the National Academy of Sciences* 93 (25): 14267-14270.
- THORNE, K. S. (1995): *Agujeros negros y tiempo curvo: el escandaloso legado de Einstein*, Barcelona, Grijalbo Mondadori.
- VARGAS, C. (2009): «Implicaciones filosóficas del abandono del concepto de “masa relativística”», *Rev. Filosofía Univ. Costa Rica* 47 (122): 81-86.
- WATT, D. (1998a): «Lionel Penrose, F.R.S. (1898-1972) and Eugenics: Part One», *Notes and Records of the Royal Society* 52 (1): 137-151.
- (1998b): «Lionel Penrose, F.R.S. (1898-1972) and Eugenics. Part Two», *Notes and Records of the Royal Society* 52 (2): 339-354.
- WEINBERG, S. (1999): *Los tres primeros minutos del universo: una concepción moderna del origen del universo*, Madrid, Alianza.
- WEIR, A. (2011): «Formalism in the Philosophy of Mathematics», en Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2011 ed.
- WITTEN, E. (2004): «Perturbative gauge theory as a string theory in twistor space», *Commun.Math.Phys.* 252: 189-258.

La propuesta heurística de Roger Penrose

En busca de las bases de la consciencia

I. LÍNEAS BÁSICAS DE SU FILOSOFÍA

La capacidad de admiración y la adaptación flexible a los nuevos descubrimientos forman parte del estilo de pensar de Roger Penrose. Con honestidad cambia de opinión cuando lo considera acorde con las nuevas contribuciones de la ciencia y así lo reconoce tanto al inicio de su trayectoria como en los años sucesivos. De este modo, admite que ha cambiado su visión sobre el origen del universo desde que al principio quedase fascinado por la belleza matemática de la teoría del estado estacionario. Lo mismo le ha sucedido al considerar el alcance de la actividad cuántica en el cerebro humano: su primer enfoque expresado en *La nueva mente del emperador* ha cambiado ante los nuevos descubrimientos de la actividad cerebral.

Junto a esta flexibilidad, Roger Penrose también reconoce el límite de sus ideas y frecuentemente las presenta como un acercamiento heurístico a una solución que quizá no se encuentre por ese camino. De este modo intenta aportar ideas que guían su pensamiento, a la vez que reconoce sus límites¹.

Esta flexibilidad, unida a la capacidad de admiración y a la transparencia, resulta necesaria para valorar los resultados aportados por la ciencia, y hace que, por contraste, resalten las ideas de fondo que permanecen bajo el enfoque concreto de cada científico. En el caso de Penrose estas ideas de fondo se refieren a planteamientos científicos pero, sobre todo, a elementos básicos de su modo de pensar y de hacer ciencia. Así, por ejemplo, defiende que la aproximación científica a la realidad genera un conocimiento verdadero cada vez

¹ Penrose 1999b, 24, aunque es un tema recurrente al inicio de sus libros.

más estable; o que las matemáticas *están ahí*; o también que la ciencia permite conocer la realidad. Se inclina así hacia posturas como las de Einstein, Schrödinger o Dirac, mientras se aleja de las de Bohr o Hawking. La estabilidad de estas ideas de fondo no solo remite a la objetividad del método científico en general, sino también a la interpretación subjetiva que cada científico hace del método y de los resultados. Es decir, las ideas de fondo remiten a una filosofía, entendida de modo coloquial, como aproximación personal (subjetiva) a la totalidad de la realidad (objetiva)².

En los próximos apartados se intentará exponer cuáles son los elementos fundamentales de la filosofía de nuestro autor, tal y como los explica. Ocasionalmente se insertaran aclaraciones personales para sacar a la luz ciertos esquemas y conexiones implícitas. Si bien Penrose hace pocas elaboraciones filosóficas explícitas, no por eso deja de ser un científico que traspasa las fronteras de la ciencia en busca del sentido último de las cosas³; y que, por lo tanto, se comporta como un auténtico filósofo.

1. *El método científico*

En el presente apartado se expondrán los elementos más significativos de la comprensión que tiene Roger Penrose del método científico. Dejaré de lado algunas características importantes del método, que no son objeto de atención por parte del físico inglés, para centrarme en sus observaciones. Desde este enfoque me aproximaré primero al conocimiento como condición de posibilidad de la actividad científica para posteriormente describir el inicio, el desarrollo, la apertura y los límites de dicha actividad.

Los que defienden la validez del método científico se apoyan frecuentemente en sus innegables logros. Penrose comparte esa apreciación y concluye –al igual que otros autores, como Artigas– que esos logros afirman un mínimo de realismo y de causalidad *ahí fuera*⁴. Ese realismo y esa causalidad, además de fundamentar la actividad científica y su método, también son condición de posibilidad de otros modos de conocer como la filosofía o el sentido

² Así emplea Penrose el término filosofía.

³ Para un estudio más detenido de esta idea se puede leer Artigas 1999a.

⁴ Artigas 1999a, 29-36 para una fundamentación más detallada del realismo y la causalidad que subyacen bajo la ciencia experimental, también en el caso de la mecánica cuántica.

común. Esos modos alternativos de conocer la realidad serían necesarios para alcanzar una visión más completa del mundo⁵, aunque el método más fiable seguiría siendo el científico.

En este momento me gustaría aclarar que si el conocimiento fuera eminentemente pragmático, entonces se podría afirmar –como sostiene Penrose– la primacía de la ciencia experimental sobre otros métodos, porque da resultados comprobables y permite dominar la realidad. Algo que no es tan obvio en la filosofía. Sin embargo, si se parte de que la realidad *está ahí* entonces es más fácil defender la validez de otros modos de aproximación que permiten contemplar o admirar la realidad sin necesidad de *intervenir*. En este caso no se defiende la eficacia controladora de una construcción teórica sino una riqueza del conocimiento que no se auto-limita a lo experimental, ni a los resultados. En última instancia supone un predominio del conocimiento en sí sobre la utilidad práctica.

La postura de Penrose pretende ser intermedia. A la vez que defiende la prioridad del método científico sobre el resto de métodos por su utilidad, no reduce la utilidad a mero pragmatismo sino que resalta su dimensión teórica desveladora de la realidad. Además, por esta utilidad teórica, el método científico no se debería detener ante preguntas difíciles, sino que debería buscar y encontrar respuestas incluso a preguntas fundamentales sobre el origen del universo, el origen de la consciencia o el origen de la vida. Para Penrose las únicas preguntas a las que la ciencia experimental nunca podrá aportar respuesta alguna serán las preguntas sobre el bien y el mal, porque son exclusivas de una aproximación moral. Lo que no queda claro es si este conocimiento moral es relativo o puede ser equivalente al *científico*.

Desde mi punto de vista y si se me permite una comparación con el fútbol, Penrose acentúa demasiado dos aspectos. En primer lugar, a la hora de buscar respuestas, sobrecarga al método científico con *responsabilidades* que le exceden y que son propias del conocimiento en su conjunto. Y en segundo lugar, ensalza demasiado los talentos *individuales* del método científico respecto a otros modos de conocer. Personalmente me inclino por valorar más el conjunto del conocimiento frente a la individualidad del método científico. Este método sería *un jugador más dentro del equipo*, quizá el que tiene más talento

⁵ En especial el de un acercamiento moral que no excluye ni compete con la aproximación científica. Pero veremos más adelante dónde sitúa Penrose la moral.

para marcar, para conseguir resultados, pero en el conjunto del conocimiento todos los métodos *juegan* y cada uno desempeña su papel.

Cada uno de los distintos métodos, según Penrose, aportaría unas respuestas que serían complementarias y, por lo tanto, con una mayor variedad de métodos se podría obtener un conocimiento más completo. Sin embargo, en algunos casos, esas contribuciones podrían dificultar la investigación en lugar de contribuir a ella. Así por ejemplo, se podría desistir de formular en clave científica algunas preguntas difíciles que encuentran respuesta satisfactoria en otros tipos de conocimiento⁶. Esto, para Penrose, sería un error.

Un ejemplo de este tipo de error sería que la fe en un Dios creador limitase la investigación científica sobre el origen del universo⁷. Esta observación me parece acertada, pero también se puede hacer una lectura en sentido contrario: conocer los entresijos físicos del origen del universo no conlleva afirmar que Dios no ha creado el universo. Un planteamiento *aut-aut* entre explicación científica y explicación divina de la realidad implicaría que se tiene una comprensión muy pobre de Dios: como una causa más entre las causas y no como la Causa Primera de todo lo creado⁸.

En resumen, desde mi punto de vista convendría matizar la neta prioridad que Penrose da al método científico e integrarlo mejor con el resto de métodos de conocimiento. A la vez considero que su enfoque tiene varios aspectos positivos. Uno que deseo resaltar ahora es que la confianza que nuestro

⁶ Una postura parecida se puede encontrar en Ruse 2010, aunque este autor se refiere solo a dos tipos de aproximaciones, una científica y otra religiosa, ambas válidas y complementarias, que no dialogan entre ellas. También en Polkinghorne 2005 y 2007 se distingue entre respuestas teológicas y cuestiones científicas. El autor pone en paralelo estos dos modos de conocimiento partiendo de un Dios que ha escrito dos libros el de la Escritura y el de la Naturaleza. A la vez, Polkinghorne se muestra a favor de cierto *God-of-the-gaps* que puede aportar información, no energía, en los procesos caóticos de la naturaleza.

En el caso de Penrose, me parece que se limita a afirmar el valor de la aproximación científica y a defenderla de precipitadas soluciones filosófico-religiosas, sin que necesariamente caiga en un dualismo entre razonamiento científico y razonamiento filosófico-religioso, ni niegue el valor de este último, ya que hay preguntas propias de cada aproximación (preguntas científicas o preguntas morales) y preguntas comunes que pueden tener respuesta con los dos acercamientos, como las preguntas fundamentales.

⁷ En Penrose 2006, 1025 se presenta el falso dilema de considerar si el *big bang* es un acto divino o si hay que buscar alguna teoría científica matemática para explicarlo.

⁸ Habría mucho que aclarar sobre este punto. Para una profundización con terminología científica de este argumento sobre la causalidad divina, se puede leer la tesis doctoral Silva 2009, que está en proceso de publicación.

autor deposita en el método científico remite a una confianza en el conjunto del conocimiento humano.

1.1. La posibilidad del conocimiento

Pensar, según Penrose, ha sido siempre una prerrogativa humana que nos ha permitido trascender nuestras limitaciones físicas. Un ejemplo de esta trascendencia es la creación de herramientas y máquinas que han obtenido logros difícilmente alcanzables para el ser humano. Eso hace que ante la aparente superioridad puntual de las máquinas nuestro orgullo no resulte herido, ni nuestra hegemonía parezca ser amenazada. Sino que el progreso fomenta nuestro orgullo como seres humanos, al ver lo que somos capaces de hacer con las máquinas. Es una manifestación más de nuestro dominio, fruto de nuestra capacidad de pensar.

Esa capacidad se fundamenta en que el hombre es un ser consciente capaz de comprender, algo que según Penrose estaría totalmente ausente en los ordenadores. Quizá las máquinas puedan tener cierto *dominio* porque son capaces de manejar más datos que un ser humano, pero es un *dominio* sin comprensión, es un *dominio* no-humano, es un *conocimiento* inconsciente.

Por lo tanto, para intentar saber cómo es el conocimiento humano, será fundamental profundizar en qué es estar consciente, qué es ser un ser consciente, y cómo es posible que se dé la consciencia en los seres humanos. Estos interrogantes constituyen un verdadero reto sobre el que la física tiene algo que decir. Penrose llegará a afirmar –sin entrar de momento en consideraciones más profundas– que las máquinas no pueden ser conscientes porque están hechas con una física distinta a la de los seres humanos. Lo veremos más adelante.

Nos encontramos por tanto ante una realidad que está *abí fuera* y que puede ser conocida porque en ella existe una causalidad que se puede desentrañar conscientemente a través de la elaboración de juicios⁹. Esta afirmación es coherente con el realismo metafísico y gnoseológico de Penrose, a la vez que remite al progreso en el conocimiento de lo real. Dicho conocimiento siempre puede alcanzar un *mayor contenido de realidad*. Lo que, en el caso de

⁹ Un aspecto capital de las argumentaciones de Penrose es la relación entre consciencia y formación de juicios, como el mismo explica y reconoce en Penrose 1999b, 36

la ciencia experimental, permitirá un mayor dominio, contendrá parte de la realidad ya conocida y abrirá nuevas áreas de investigación. Un ejemplo clásico de este progreso lo constituyen las teorías de la relatividad de Einstein en relación con la mecánica newtoniana. En este caso, el conocimiento de la realidad ha sido aclarado, perfeccionado y profundizado, a la vez que se han abierto nuevos campos de investigación. Por tanto, hay más realidad conocida tanto extensiva como intensivamente¹⁰.

1.2. El punto de partida

Una vez resaltada esta capacidad de comprender la realidad, que va más allá de la capacidad funcional o pragmática de hacer teorías y obtener resultados correctos, la pregunta que se plantea es: cuál es el punto de partida del científico. Para Penrose ese punto de partida es doble. Está constituido tanto por las teorías consolidadas que han sido aceptadas por la comunidad científica¹¹ como por los resultados que aportan los nuevos experimentos.

A la vez, ese doble punto de partida no es una base inamovible, sino que tiene la solidez de las placas tectónicas: las teorías recibidas se revisan mediante la elaboración de nuevos experimentos y los datos experimentales están sujetos a reinterpretación. Hay un continuo flujo de teorías, experimentos e interpretación, donde el ser humano juega el papel fundamental. En ese acceso a la realidad, el ser humano formula las teorías, prepara los experimentos, interpreta los datos y juzga la oportunidad de qué es lo que hay que cambiar: la teoría, los experimentos o la interpretación. De modo que las teorías y los experimentos no constituyen solo el punto de partida, sino también un punto de continuo retorno mediante la interpretación y el juicio. La revisión de una teoría dependerá del juicio científico sobre lo fundamentales que sean los datos aportados por los experimentos.

Con este enfoque, es lógico que Penrose formule sus teorías asociándolas a experimentos. Así, en sus publicaciones más recientes sobre mecánica cuántica, sobre el origen del universo o sobre las bases físicas de la consciencia, aparecen propuestas de experimentos que en algunos casos se podrán llevar a

¹⁰ En Artigas 1999b, al final del segundo capítulo, se desarrollan más ampliamente estas ideas al considerar la retroacción del progreso científico.

¹¹ Estas teorías, que tienen distintos grados de fiabilidad o solidez, Penrose las divide en tres tipos: soberbias, útiles y tentativas. Penrose 1999b, 199-203

cabo en unos pocos años, mientras que en otros habrá que esperar a que progrese la técnica¹². Por eso mismo, Penrose rechaza algunas teorías que no son experimentalmente comprobables, como las del multiverso, y califica otras como interesantes elucubraciones porque no postulan ningún experimento. Nuestro autor también tiene claro que no todos los datos experimentales fundamentan una teoría con la misma solidez y reconoce que mientras algunos de sus experimentos servirían para fundamentar o desacreditar una teoría, hay otros que solo darían indicios de viabilidad de la teoría¹³.

Hasta aquí podríamos calificar lo explicado, en terminología kuhniana, como el periodo de ciencia normal, en el que la investigación se mueve dentro de un paradigma consensuado que sirve de modelo y que se consolida o retoca con nuevos experimentos¹⁴. Pero además, nuestro autor señala la aparición de anomalías en el paradigma vigente. Anomalías que empiezan a ser de entidad porque se observa la proliferación de nuevas teorías y eso es un síntoma de debilidad del paradigma.

Esta postura de Penrose, señalando la crisis del paradigma actual y apuntando hacia un nuevo escenario de la física, no implica que defienda una revolución científica o un periodo de ciencia extraordinaria al estilo de Thomas Kuhn. De igual modo, tampoco responde al esquema popperiano de refutación del paradigma actual, ni supone la existencia de experimentos cruciales.

¹² En Penrose 2006, 1144-1145 habla del experimento FELIX para probar su teoría sobre la Reducción Objetiva del vector de estado en mecánica cuántica; en Penrose 2010, 221 apunta hacia un estudio más profundo de la radiación de fondo para apoyar con datos su teoría de la Cosmología Cíclica Conforme (CCC) sobre el origen del universo, en la que se sigue investigando: Gurzadyan y Penrose 2011 y Gurzadyan y Penrose 2010; y en Penrose 2006, 1381-1382 comenta de algunos experimentos que se podrían hacer en relación con su propuesta sobre las bases físicas de la consciencia, aunque de momento los resultados no parezcan muy alentadores por lo que se puede ver en McKemmish 2009.

¹³ Por ejemplo el experimento FELIX, respecto a la teoría de la reducción objetiva del vector de estado, sería una prueba sólida. Mientras que unos resultados positivos en los estudios de radiación de fondo, en relación con su teoría CCC, podrían manifestar la existencia de eones previos pero no serían suficiente para fundamentarla.

Para una exposición más detallada cfr. Penrose 2006, 1365-1369. Ahí se explica cómo se hace ciencia y cómo se refuta una teoría errónea. Además, se considera que existen teorías serias casi inmunes a la refutación o a la confirmación, como la que postula una constante $K=0$ en cosmología inflacionaria. Por último, Penrose también explica cómo se reelaboran las teorías ante la aparición de nuevos datos.

¹⁴ Para una profundización cfr. Diéguez Lucena 2005, en especial el capítulo 5. Aunque nuestro autor no se puede incluir en el esquema kuhniano, sí que se encuentran puntos de contacto y, por eso, me parece interesante usar su terminología de un modo flexible.

Más bien sugiere una mejora del paradigma mediante una relectura de los datos ya conocidos que integre los últimos avances de la ciencia física¹⁵.

Nuestro autor, por tanto, apunta a un cambio de enfoque, a reinterpretar los experimentos, a estudiar qué se ha podido pasar por alto, integrándolo todo conforme al método científico que tan buenos resultados ha dado. Para conseguir este cambio de enfoque Penrose resalta la importancia de los conocimientos heurísticos, del sentido común y de la intuición matemática del hombre de ciencia. Veremos poco a poco estos aspectos, pero antes me detendré en considerar qué entiende Penrose por *conocer la realidad*.

1.3. El trabajo científico

Uno de los argumentos transversales que ha aparecido en secciones anteriores es el realismo. Para algunos autores se trataría de una cuestión poco relevante¹⁶ mientras que para Penrose es esencial. Por eso, conviene que él mismo aclare qué entiende por realidad:

«Hablo de la realidad de los objetos físicos: esta mesa, este bolígrafo, la Tierra (...) [y] distingo tres mundos de la realidad. Por un lado, el de la realidad física; por otro, el de la experiencia mental, y por último, el mundo platónico de los absolutos matemáticos. Así que concibo tres tipos distintos de realidad»¹⁷.

Este esquema de tres mundos está inspirado en el mundo de las ideas de Platón, en el mundo mental de Berkeley y en el esquema de los tres mundos de Popper¹⁸. Sin embargo, se trata de un esquema propio que Penrose expone

¹⁵ En Penrose 2006, 1061 hace un resumen de cómo se debería dar ese cambio sustancial de teoría, sin meros retoques, a la vez que reconoce que sus indicaciones son retoques.

¹⁶ Por ejemplo Stephen Hawking afirma: «I don't demand that a theory correspond to reality because I don't know what it is. Reality is not a quality you can test with litmus paper. All I'm concerned with is that the theory should predict the results of measurements. Quantum theory does this very successfully». Hawking y Penrose 2010, 81

¹⁷ Alfieri 2007, 125. Más adelante, sobre la relación entre las matemáticas y la física, añade: «Creo que es un misterio. Pero considero que estos tres mundos de la realidad están interrelacionados. Sin embargo, para ser sincero, no entendemos el porqué. De alguna manera, parece que hay un acuerdo extraordinario entre la realidad física y las estructuras matemáticas».

¹⁸ A los mundos físico y mental de Berkeley, Popper añade un tercer mundo, el mundo de la cultura o «mundo III», que se enriquece a partir del mundo mental («mundo II») de las personas. El mundo de la cultura correspondería con el conocimiento objetivo (ciencia, historia, filosofía,

en tres libros sucesivos. Aun así, no se ciñe a una terminología muy clara e introduce ligeros cambios con una mayor profundización. Por lo tanto, me centraré en su explicación más reciente, la que se encuentra en *El camino a la realidad*.

Como disposición previa, antes de explicar los tres mundos, Penrose pide al lector que amplíe su concepto de «existencia real». En este concepto deberían caber no solo las cosas físicas sino también las estructuras matemáticas, porque parecen existir antes de ser descubiertas. Si bien no poseen una existencia espacio-temporal como los objetos físicos, según Penrose, son atemporales, estaban ahí y el hombre las descubrió.

Una vez admitido esto, en la realidad habría tres tipos de existencia diferentes: la matemático-platónica, la física y la mental. De este modo, la existencia de un concepto matemático se fundamentaría en su consistencia matemática; la existencia de los elementos físicos sería espacio-temporal e implicaría algún tipo de percepción sensible; y la existencia mental sería la propia de las ideas o pensamientos¹⁹. Estos tres tipos de existencia propias de cada uno de los tres mundos constituirían la única realidad. Por lo tanto, aunque los mundos estén separados entre sí, se interconectarían de algún modo. Y esas interconexiones constituyen para Penrose tres *misterios*²⁰ que hay que investigar:

1. El misterio que relaciona el mundo físico con una pequeña parte del mundo matemático-platónico que le sirve de fundamento.
2. El misterio que relaciona el mundo mental con una pequeña parte de las estructuras físicas, que le sirven de soporte físico.

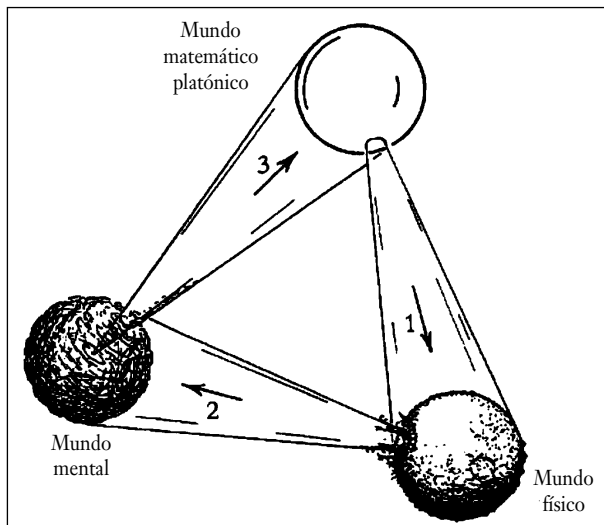
teología...), mientras que el mundo mental tendría que ver con el conocimiento subjetivo de las personas (percepciones, emociones, pensamientos, intenciones...). Penrose reemplaza este «mundo III» con un mundo matemático platónico, creando un esquema distinto, porque a su vez, ese mundo platónico fundamenta el mundo físico. Cfr. Penrose 1999a, 13-15, 79-82. También en Lombo y Russo 2005, 220-223, se puede encontrar una explicación más detallada del esquema popperiano.

¹⁹ En este contexto los términos de la existencia espacio-temporal y existencia mental no están bien definidos, mientras que el concepto de consistencia matemática es una noción sintáctica con un contenido preciso. La noción de consistencia matemática constituye uno de los elementos centrales de las pruebas matemáticas. Para una profundización se puede leer Shapiro 2000.

²⁰ Penrose 1994, 420: «These are deep issues and we are yet very far from explanations. I would argue that no clear answers will come forward unless the interrelating features of all these worlds are seen to come into play. No one of these issues will be resolved in isolation from the others. I have referred to three worlds [Platonic world, Physical world and Mental world] and the mysteries that relate them one to another. *No doubt, there are not really three worlds but one, the true nature of which we do not even glimpse at present*». El subrayado es mío.

3. Y finalmente, el misterio que relaciona el mundo matemático-platónico con una pequeña fracción de la actividad mental.

Estos misterios interconectan los distintos mundos de un modo preciso y unívoco de tal modo que: todo el mundo físico estaría gobernado por una pequeña parte de las matemáticas; todo el mundo mental estaría enraizado en una pequeña parte de la física; y todas las verdades matemáticas serían alcanzables por la razón, aunque no toda la actividad mental se emplee para hacer matemáticas. Además, cada mundo emergería del anterior, tal y como se muestra en el siguiente esquema.

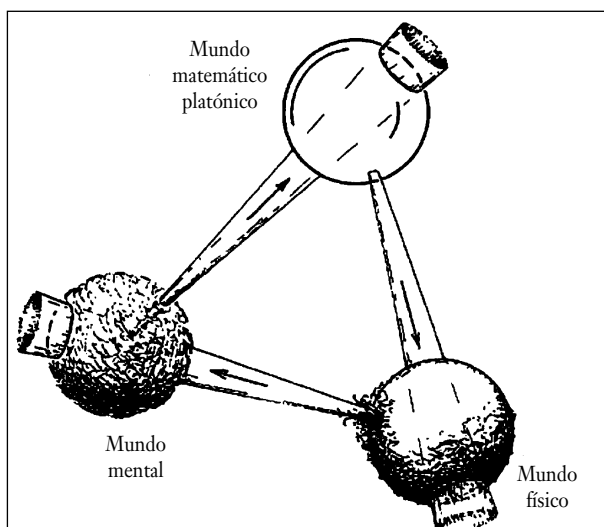


Hasta aquí las ideas expresadas por Penrose requerirían de aclaración. El esquema resulta demasiado oscuro y ni siquiera manifiesta una dificultad obvia: vista así, la realidad es *causa sui*. Sin embargo, Penrose no mostrará más claridad respecto a este esquema, sino que insistirá en que el esquema solo pretende expresar sus *prejuicios*²¹. Al final del presente trabajo volveré sobre

²¹ Al tratarse de un autor inglés es probable que el término «prejuicio» contenga una carga semántica relacionada con los *Idola* de Bacon. Francis Bacon en *Novum Organum* sostenía la necesidad de abandonar todos los prejuicios y actitudes preconcebidas, los *Idola*, para poder hacer ciencia. En el fondo de este planteamiento estaba su método experimental inductivo que mejoró notablemente las hipótesis científicas. Bacon pensaba que si se eliminaba toda noción

este punto y recalcaré que este esquema no es el más apto para comprender la realidad en su conjunto. Pero de momento seguiré describiendo su punto de vista, con la claridad que permita el autor en cada momento.

Ante las críticas recibidas por su esquema, Penrose admitió algunos cambios. Así aceptó la posibilidad de una acción física que no esté gobernada por las matemáticas, o de una actividad mental²² que no tenga como substrato las estructuras físicas, o de unos enunciados matemáticos verdaderos cuya verdad sea, en principio, inaccesible mediante la razón y la intuición. Esta postura vendría representada por el siguiente esquema, donde los mundos (esferas) no son abarcados por los *misterios* (conos con flechas).



preconcebida del mundo, se podía estudiar la realidad mediante observaciones controladas, y realizar generalizaciones cautelosas. Los científicos por tanto debían ser escépticos y sus hipótesis se debían validar por la observación y la experiencia sensible. Sin embargo, no parece que Penrose acepte las tesis de Bacon ya que en ese caso tendría que rechazar sus prejuicios en lugar de mostrarlos.

²² Mentalidad es el concepto usado en la traducción para referirse al término *mentality*. Sustancialmente son términos cuyo contenido semántico es el mismo en castellano que en inglés. Sin embargo, según el Diccionario de la Real Academia de la Lengua Española, la primera acepción en castellano se refiere a la «cultura y modo de pensar que caracteriza a una persona, a un pueblo, a una generación...». En este trabajo se usará el término siempre conforme a su segunda acepción: «capacidad, actividad mental», que es la que se corresponde con el significado que le quiere dar nuestro autor. Es el mundo de las percepciones conscientes.

Sin embargo, Penrose considera que estas modificaciones son superfluas ya que su primer esquema sería capaz de expresar bien los tres *misterios*: el misterio de cómo las leyes matemáticas pueden aplicarse al mundo físico con tanta precisión, belleza y sofisticación; el misterio de que en una materia física adecuadamente organizada, como el cerebro humano, se pueda evocar el conocimiento consciente; y el misterio de que las mentes humanas sean capaces de tener intuiciones matemáticas complejas o captar con sencillez los números enteros.

Profundizar en los tres mundos sin dejar de lado estos misterios es tarea del científico, ya que, no obstante el físico se dirija esencialmente al mundo físico, no puede obviar las relaciones con los otros dos mundos.

«El que suceda algo muy enigmático no significa que nunca seamos capaces de comprenderlo».

En este camino a la realidad, ambicioso y difícil, el científico debe preguntarse sobre el *qué* del objeto que investiga y no centrarse solo en *cómo* se comporta. Además, la pregunta por el *qué* le remitiría inmediatamente al *por qué*²³ y abriría las puertas de la filosofía. Sin embargo, como advierte Penrose, hasta ahora apenas se ha avanzado siquiera en el intento de desvelar el primero de los misterios: no se comprende bien la relación entre la realidad matemática y la realidad física.

1.4. Sentido común y apertura a la filosofía

Una vez afrontada la cuestión sobre qué es la realidad para Roger Penrose, recuperaré el hilo argumental. Vuelvo por tanto a la profundización del conocimiento como un *tender hacia un mayor contenido de realidad mediante diversos métodos y enfoques*. El científico desde su punto de partida busca un mayor contenido de realidad y en su tarea no se limita a lo experimental, sino que reclama una apertura a otros modos de conocer y en última instancia a las preguntas filosóficas. Por lo tanto, un *mayor contenido de realidad* reclama otros modos de aproximarse a la realidad, entre los que nuestro autor incluye

²³ Penrose 2006, 1375-1378. ¿Por qué cuanto más sondeamos la naturaleza, más nos dirigimos al mundo platónico de las ideas matemáticas?, se preguntará Penrose.

el sentido común, la moralidad o la filosofía. Antes de ver cada uno de ellos quizá sea útil una metáfora entre el modo de conocer y la luz.

Siguiendo esa metáfora, una diversidad en el modo de conocer equivale a una diversidad en las gamas del espectro de luz. Así, al igual que algunas bombillas emiten con más intensidad en unos tonos que en otros, un modo de conocer se centra más en algunos aspectos de la realidad que en otros. Esto hace que con una determinada iluminación o modo de conocer predomine una *tonalidad*.

Supongamos ahora que una aproximación científica a la realidad es equivalente a una iluminación artificial con bombillas incandescentes. En ese caso, si se aumenta la cantidad de iluminación sin cambiar el tipo de bombilla, la tonalidad se conservará. Lo que según la metáfora significa que con cuanta más intensidad se emplee el método científico mayor será el conocimiento de la realidad, pero siempre con una determinada *tonalidad*.

Otro modo de aumentar el conocimiento sería usar distintos métodos, lo que equivale a utilizar distintos tipos de bombillas (incandescentes, vapor de mercurio, fluorescentes...). Así se cubre un espectro de luz más amplio y se atenúa el predominio de una tonalidad. Si la metáfora se lleva todavía más lejos, lo que realmente interesaría es conocer la realidad con luz natural. O incluso, como el espectro de luz es mucho más amplio que el espectro visible, se podría completar con la ayuda de infrarrojos o de ultravioletas.

Como se puede apreciar, la metáfora del conocimiento como iluminación es bastante rica y se podría seguir explotando, pero no deseo sujetarme en exceso a ella. El último punto que me gustaría resaltar es que para Penrose la *primera luz* con que vemos la realidad es el sentido común²⁴. Un sentido común que siempre debe estar presente en toda actividad humana, también en la ciencia, porque para conocer la realidad es más esencial que la actividad científica. Ese sentido común lleva a fiarse de lo que recibimos, de que la realidad está ahí, de que no somos engañados cuando conocemos, de que nos podemos equivocar pero también nos podemos corregir... y esa confianza en nuestra

²⁴ Livi 1995, 261: «La *verdad* de las *certezas* del sentido común resulta *indudable*. Por el contrario las verdades alcanzadas por la ciencia –en este caso, la filosofía– son siempre más o menos, a tenor de los diferentes casos, *susceptibles de ser puestas en duda*. Por tanto, en la certeza, en cuanto tal, el sentido común goza de un primado que jamás se le podrá arrebatar». Para una profundización filosófica sobre el sentido común se puede consultar la obra de Antonio Livi.

percepción inicial de las cosas es esencial para hacer ciencia. No se parte de la duda cartesiana.

Por eso, a Penrose le gusta conectar el conocimiento científico con el sentido común y tener una imagen de lo que sucede. A la vez no le satisface lo excesivamente abstracto y le llama la atención que algunos físicos elaboren teorías que se alejan mucho de la percepción ordinaria. Si bien el sentido común llevaría a aceptar teorías científicas bien fundamentadas que inicialmente podían parecer *lejanas*, para Penrose, en ningún caso esas teorías podrían ser *extrañas*. La armonía de la actividad científica tendría una continuidad con el sentido común, con notas distintas pero no disonantes. Algo similar dirá Penrose del criterio estético, de la simplicidad y de la belleza de las teorías matemáticas. Se trata de criterios que no pueden suplantar al método científico, pero que están presentes en las teorías correctas. Las teorías correctas poseen cierta belleza intrínseca y no pueden ir contra cierto sentido común razonable.

Como contrapunto a esta postura, algunos autores defienden la necesidad de ir contra el sentido común para liberarse de clichés y prejuicios que dificulten hacer verdadera ciencia. Así por ejemplo Hofstadter afirma que el concepto de «yo» es en última instancia una convención social, una ilusión ficticia sobre la que nos hemos puesto de acuerdo y de la que nos resulta casi imposible liberarnos. En el fondo, el «yo» se refiere a una consciencia que es un extraño bucle, mera consecuencia de las leyes físicas, y que se asocia a la grandeza de sentimientos.

Este sería un ejemplo de una teoría *extraña*, alejada de la percepción ordinaria. Para Penrose, según el sentido común no solo percibo mi «yo» sino que también puedo percibir el «yo» de otras personas conscientes. Me podría equivocar al negar que alguien esté consciente, pero casi siempre acertaré al afirmar que alguien está consciente.

«Por lo tanto debe haber realmente algún modo de comportamiento que es característico de la consciencia (incluso aunque no sea *siempre* manifestado por la consciencia) y al que somos sensibles a través de nuestras intuiciones de sentido común».

También Rosenblum y Kuttner, apoyándose en las teorías actuales sobre mecánica cuántica, apelan a la necesidad de poner en duda tres intuiciones de sentido común para poder comprender lo que dice la física. Según ellos en la

mecánica cuántica sería la observación la que *crea* la realidad física observada. Por lo tanto, habría que dejar de suponer que dos objetos no pueden estar a la vez en el mismo sitio, lo que según ellos equivale a poner en duda el principio de no contradicción. Además, también habría que dejar de pensar que lo que sucede aquí no puede afectar *simultáneamente* en algún lugar muy lejano. Por último tampoco podríamos creer que exista un mundo real *abí fuera* con independencia de que lo contemplemos o no²⁵.

Sería discutible afirmar que algunas de estas posturas vayan realmente contra el sentido común y, además, no es el objetivo del presente trabajo. Pero quisiera recalcar la importancia de interpretar los datos experimentales a la luz del sentido común. Según Penrose, así se evitaría crear teorías como las de Rosenblum y Kuttner donde todo es posible salvo admitir que pueda haber un error en las teorías o en la interpretación que se hace de ellas. El sentido común constituye la primera luz con que vemos las cosas y conviene fiarse de él. Esa luz es esencialmente correcta y no se debe apagar nunca, aunque de vez en cuando, en la apreciación inicial de la realidad, pueda inducir a error.

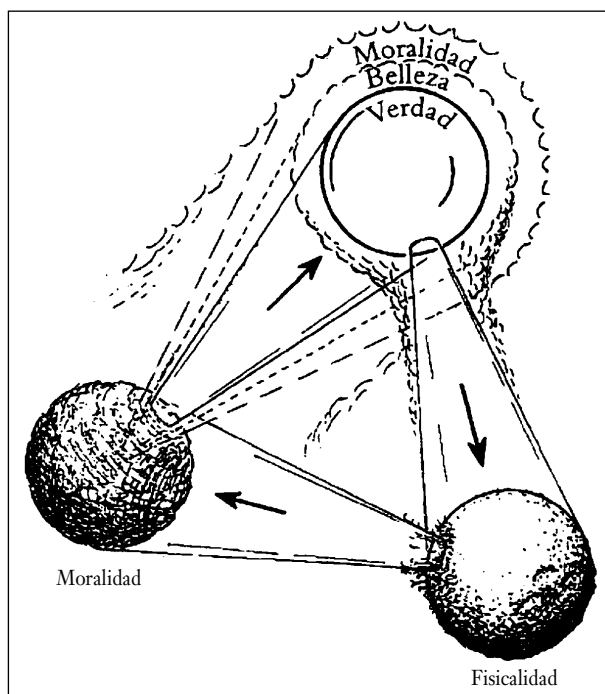
Esa confianza en el sentido común y el hecho de que la mecánica cuántica no tenga hasta ahora una formulación muy bella le llevan a afirmar que:

«La mecánica cuántica actual no tiene una ontología creíble, de modo que debe ser modificada para que la física del mundo tenga sentido».

Después de haber hablado del sentido común me centraré ahora en la moralidad, como modo válido de aproximarse a la realidad. Para ello me

²⁵ Rosenblum y Kuttner 2010, 18. Me parece que el concepto de sentido común que usan estos autores no es el mismo que el de Penrose. En Penrose el concepto es más clásico-natural, mientras que en estos autores hace referencia a lo científicamente aceptado. Se puede entender mejor a qué me refiero cuando en Rosenblum y Kuttner 2010, 48-52, se señalan cinco premisas newtonianas de sentido común que quedan comprometidas por la mecánica cuántica: *Determinismo* o paradoja del libre albedrío (inicialmente se evita la paradoja separando la consciencia-libertad de la materia y dejando esta última para la física pero, hoy en día, se pone en duda por la influencia del observador consciente); *Realidad física* (se pone en duda por la influencia del observador consciente en el experimento cuántico); *Separabilidad* (se pone en duda por la presencia de acciones instantáneas, sin la intervención de fuerzas físicas, que violan esta separabilidad); *Reduccionismo* (la mecánica cuántica sirve como apoyo de posturas antireduccionistas al señalar la importancia de la consciencia); *la física newtoniana es una explicación suficiente* (con la aparición de las teorías de la relatividad se observa que no lo es).

serviré del siguiente esquema, que es una ulterior modificación del esquema inicial de los tres mundos.



En este esquema, el mundo platónico de Penrose ya no está constituido solo por estructuras o conceptos matemáticos. Sigue situado encima de los otros dos mundos, para resaltar que es el más fundamental, pero ha sido enriquecido con los absolutos platónicos de belleza y moralidad. Entre las auras del mundo platónico, aparece la belleza que rodea a la verdad matemática y que se extiende a la relación de la verdad con el mundo físico²⁶. Refleja de este modo el *prejuicio* de Penrose según el cual las teorías matemáticas verdaderas que describen el mundo son también bellas.

²⁶ Considerando esta relación Penrose hablará de las reglas ocultas o influencias que hay entre la belleza o elegancia con las matemáticas, o incluso de milagros matemáticos, pero considerando que son una orientación para conocer la verdad, no una guía segura. También señalará que la belleza de las teorías científicas sólo la pueden captar los que las conocen bien y que la ciencia a partir de ciertos niveles tiene bastante de arte. Penrose 2006, 1389-1392.

Con el aura de la moralidad se pretenden reflejar también algunos de los *prejuicios* de Penrose en relación al conocimiento del bien y el mal. Una primera idea es que la moralidad no se asocia con el mundo físico, mientras que sí mantiene una relación con el mundo mental, al igual que la verdad y la belleza. A la vez se diferencia de estas en que la moralidad parece de algún modo abierta hacia algo más amplio que el mundo mental.

Dada la inclinación de Penrose a plasmar geoméricamente sus ideas, resulta bastante razonable que haya querido expresar sus *prejuicios* filosóficos en un esquema. Por eso, cada detalle de este esquema resultaría significativo: hacerlo de un modo o de otro conlleva unas consecuencias muy distintas sobre la subjetividad u objetividad de la moral, sobre la bondad del mundo físico... De todos modos no las analizaré con más detenimiento porque nuestro autor no se detiene en aclararlas con mucha precisión. Aun así me gustaría señalar su conexión con la doctrina clásica de los trascendentales, que ha sido elaborada durante muchos siglos y posee una indudable riqueza y precisión filosófica para describir la realidad²⁷.

Para terminar este apartado consideraré el concepto que Penrose tiene sobre la filosofía. Con frecuencia subraya que a él solo le competen las respuestas científicas y siempre evita cualquier tipo de alusión religiosa²⁸. Pero a medio camino entre ciencia y teología se encuentra la filosofía. Y con relación a este campo del saber sí que se encuentran algunas alusiones en nuestro autor, amén de que todo su planteamiento de fondo tiene un enfoque filosófico.

Ante una demanda filosófica sobre la respuesta que daría la física a la pregunta «por qué hay algo y no más bien la nada», nuestro autor responde:

«No lo sé. En un sentido matemático se puede contestar este interrogante hasta cierto punto. Porque un concepto matemático existe si es consistente. Así que si las reglas, las normas son consistentes y coherentes entre sí, entonces decimos que esa entidad existe. Pero claro, éste es el sentido matemático de la existencia. Se podría pensar que la existencia física es más o menos así, pero la verdad es que no estoy seguro. Necesitamos saber más acerca de qué es lo que constituye la existencia en el sentido físico del

²⁷ Para un estudio más profundo se puede consultar Aertsen 2003.

²⁸ Sobre todo para evitar que constituyan un atajo ante preguntas que la ciencia todavía no sabe responder y se evite así la formulación de la pregunta en categorías científicas. Un ejemplo en relación a si el universo ha sido creado por Dios se puede ver en Penrose 2006, 1011

término. Mi sospecha es que tendrá que ver con la percepción consciente, porque la pregunta sobre si existe un universo es posible porque soy consciente».

Por tanto nuestro autor no rechaza la pregunta, ni se encierra en un reduccionismo *fisicista*, sino que admite que no hay una respuesta clara. Algo parecido le sucederá en otros momentos cuando llegue a calificar como *obras de Dios* aquellas realidades en las que de su estructura sale mucho más de lo que se introdujo. A la vez, consciente de que solo quiere dar respuestas científicas, Penrose intenta no extralimitarse en sus contestaciones. Y cuando se le pregunta sobre la contingencia o necesidad de la existencia del universo responde que esa pregunta le lleva a los límites de la física y que no pasa nada por desembocar en la filosofía pero prefiere no dar una respuesta²⁹.

En ese difícil equilibrio por no decir más de lo que como científico puede decir, Penrose se formula varias preguntas sobre el sentido de la vida y manifiesta así la necesaria apertura a la actitud filosófica:

«Me parece claro que estas meditaciones y murmuraciones a las que nos entregamos cuando (quizá temporalmente) nos hacemos filósofos (...) son el ‘equipaje necesario’ que deben llevar los seres que son conscientes (...). Es cuando vemos que otros se comportan con esta extraña conducta filosófica cuando nos quedamos *convencidos* de que estamos tratando con individuos, distintos de uno mismo, que también tienen mentes».

En resumen, podría calificar el método científico de nuestro autor como «integrador». Por un lado, porque es un método que no se detiene ante los problemas. Por otro, porque subraya una continuidad histórica del conocimiento, donde no hay rupturas con el pasado aunque pueda haber saltos. En tercer lugar porque hay cierta conexión misteriosa entre los tres mundos que según Penrose constituyen la realidad. Y por último, porque defiende una integración gnoseológica con el sentido común como precursor y orientador

²⁹ Transcribo alguna de las respuestas de la entrevista que se encuentra en Alfieri 2007, 131:

– *Esto llevaría a otra pregunta: ¿es contingente el universo o necesariamente tenía que existir?*

– Con sus preguntas me está llevando a los límites de la física. Creo que voy a empezar a utilizar un comodín.

– *Quizás desembocamos en la filosofía.*

– Bueno, vale, no pasa nada. Lo que no sé es si contestaré a ciertas preguntas. En este caso voy a optar por pasar, como se permite hacer a los concursantes en algunos programas de televisión.

y con la filosofía como continuador y catalizador de la ciencia hacia respuestas que esta no puede dar.

Sin embargo, echo en falta una integración explícita con el comportamiento moral o con la religión. Esto se puede deber a que la filosofía está en contacto con la ciencia experimental, mientras que la religión y el comportamiento moral estarían *al otro lado del puente* tendido por la filosofía entre la ciencia y la teología. El hecho de que Penrose sitúe la moral sin aparente conexión con la actividad científica me parece que constituye la deficiencia más significativa de su esquema. Con esta afirmación no pretendo extraer conclusiones que él no extrae, ya que a mi parecer constituye más un olvido, fruto de la aparente poca conexión de la física con dilemas morales, que un deseo de alejar la moral de la actividad científica. Aun así me parece necesario subrayar la importancia del carácter ético de los valores científicos ya que constituyen una fructífera orientación de la actividad científica y no un límite, como tantas veces se pretenden mostrar³⁰.

1.5. Límites del método científico

Quizá la primera limitación consciente con la que un científico se encuentra en su trabajo es la precisión. Se trata de un límite importante ya que el poder explicativo de la ciencia radica en la exactitud. Sin embargo, hay algunos límites más significativos: aquellos de los que el científico no es consciente. Hasta ahora se han mostrado algunas limitaciones del método científico y, sin ánimo de ser exhaustivos, se pueden añadir algunas más señaladas por Penrose. Para facilitar su comprensión las dividiré en esenciales o circunstanciales. Las primeras siempre estarán ahí, mientras que las segundas pueden resolverse parcialmente con el tiempo.

Entre las limitaciones esenciales se pueden incluir: aquellas preguntas a las que el método científico no puede responder; o los tipos de respuestas válidas que no puede dar, como las filosóficas o religiosas; o también que con el progreso de la ciencia cada vez son más las preguntas sin respuesta. Por otro lado, entre las limitaciones circunstanciales señaladas por Penrose se encuen-

³⁰ Nuestro autor no pretende hacer filosofía a este nivel pero, mientras que en sus planteamientos se observan tanto los supuestos ontológicos como los epistemológicos, sería pretencioso afirmar que hable de supuestos éticos. Para una explicación más detallada de los supuestos ontológicos, epistemológicos y éticos de la ciencia se puede consultar el capítulo II de Artigas 1999b.

tran: que los científicos no se preocupan por las preguntas fundamentales; que la investigación actual no permite la existencia de investigadores aislados y es movida por las teorías que más inversiones atraen; o que las dificultades técnicas y económicas dificultan la realización de experimentos importantes. Algunas de estas limitaciones circunstanciales podrían ser puntualmente mejoradas en el futuro, pero el conjunto de ellas aumentará.

Por tanto, existen unas limitaciones en el método, por su modo sesgado de conocer la realidad, que existirán siempre. Es importante conocerlas, para no caer en reduccionismos. Y además, existen unas limitaciones coyunturales que, conforme desaparezcan, abrirán un campo más amplio de nuevas limitaciones. Quizá por eso nuestro autor, más que señalar los límites del método científico prefiere subrayar la potencialidad: la capacidad de profundización y ensanchamiento sin límites que tiene su uso; y cómo en la práctica se está comenzando a conocer cómo es el mundo. En este intento del método científico por profundizar en el conocimiento de la realidad, Penrose resalta el papel esencial de las matemáticas.

2. *La importancia de las matemáticas*

La armonía que tantas veces se observa entre las matemáticas y la realidad física constituye el primero de los tres *misterios* señalados por Penrose. Se trata del misterio más explorado, por el constante uso de las matemáticas en la investigación científica, y también del misterio que más influencia ha ejercido en el pensamiento de nuestro autor. Debido a esa conexión de las matemáticas con el mundo físico, Penrose evita las posturas demasiado abstractas. Por eso, rechaza tanto el formalismo matemático, que trabaja con símbolos sin necesidad de conectarlos con la realidad³¹, como el intuicionismo, para el que las matemáticas son una creación mental³². Para nuestro autor las matemáticas *están ahí* y se llega a ellas descubriendo más que inventando, mediante cierta intuición directa.

³¹ Weir 2011: «The guiding idea behind formalism is that mathematics is not a body of propositions representing an abstract sector of reality but is much more akin to a game, bringing with it no more commitment to an ontology of objects or properties than ludo or chess».

³² Iemhoff 2012: «Intuitionism is based on the idea that mathematics is a creation of the mind. The truth of a mathematical statement can only be conceived via a mental construction that proves it to be true, and the communication between mathematicians only serves as a means to create the same mental process in different minds».

Desde este enfoque, que quiere conectar las matemáticas con la realidad, Penrose extraerá conclusiones sobre la importancia de la no-computabilidad y de las proposiciones matemáticas verdaderas indecidibles, para aplicarlas a la realidad. Además, se alejará de posiciones mecanicistas, señalará la necesidad de que en la física existan aspectos no-algorítmicos, y sugerirá la conexión de esos nuevos elementos con la actividad consciente.

Pero antes de llegar a estas conclusiones conviene explicar cómo entiende Penrose la relación de las matemáticas con el mundo mental y con el mundo físico. Ya vimos que, según el esquema de los tres mundos, las matemáticas no están dentro de la mente sino que *emergen* misteriosamente de cierta actividad mental; y que la realidad física *emerge* de una pequeña parte de las matemáticas³³. Sin embargo, me parece que este esquema es una elaboración posterior que se origina en la experiencia de Penrose como investigador y en su comprensión gnoseológica y ontológica de las matemáticas. Conviene recordar la experiencia vital que supuso para Penrose resolver el problema del *teselado aperiódico* y descubrir posteriormente su conexión con los cuasi-cristales; y tener en cuenta que en su primer ensayo no se habla del esquema de los tres mundos. Por tanto, sostengo que el esquema de los tres mundos de Penrose es posterior a su declarado platonismo matemático³⁴ y, por eso, conviene enfocarlo desde esta comprensión matemática.

2.1. El formalismo frente al realismo

Roger Penrose afronta el dilema de la realidad matemática preguntándose sobre la posibilidad de la existencia real de los objetos matemáticos. Toma como punto de partida la percepción común de que esos objetos son meras idealizaciones mentales hechas por los físico-matemáticos en su deseo de conocer la realidad³⁵:

«Una opinión común entre muchos físicos de hoy es que la mecánica cuántica ¡no nos ofrece ninguna imagen de la realidad! De acuerdo con esta opi-

³³ Me parece que en la postura de Penrose este concepto se puede aproximar a la cosmovisión emergentista de K. Popper, por la influencia de este autor en su concepción de los tres mundos. Para una profundización se puede consultar Corcó Juvina 1995.

³⁴ Davies 1996, 131-133 y Berto 2009, 147

³⁵ En la base de este impulso se intuye el éxito newtoniano al aplicar las matemáticas como herramienta para conocer la realidad física.

nión, el formalismo de la mecánica cuántica es tan solo eso: un formalismo matemático. Este formalismo, como argumentarán muchos físicos cuánticos, no nos dice nada sobre la realidad cuántica del mundo, sino que meramente nos permite calcular probabilidades para realidades alternativas que podrían ocurrir».

Desde este punto de partida, Penrose se pregunta si los objetos matemáticos pueden ser algo más que construcciones arbitrarias de la mente humana. Y observa que en esos conceptos parece existir una realidad profunda que va más allá de las construcciones mentales de un matemático particular. Es como si el pensamiento matemático estuviese guiado por alguna verdad externa, que tiene realidad por sí misma y que solo se revela parcialmente.

Para sostener su postura, Penrose toma como ejemplo el *conjunto de Mandelbrot*³⁶ y reclama para él una existencia real. En ese conjunto, tras una sencilla fórmula, se esconde una estructura de complejidad inabarcable, tanto para el hombre como para los ordenadores, que se va descubriendo poco a poco. Lo que para Penrose implica que ya *estaba ahí*. Lo mismo opina del sistema de los números complejos. Se trata de una realidad profunda y atemporal que está ahí y que tiene aplicación a la realidad, por ejemplo en los fractales naturales³⁷ y en la mecánica cuántica. La estructura de los números complejos posee una *magia* inherente que se va descubriendo poco a poco.

Por lo tanto, y volviendo un poco atrás, ante las matemáticas cabrían al menos dos posturas básicas. La de pensar que son solo construcciones mentales creadas por los matemáticos, pero que no tienen auténtica realidad. O la de pensar que se están descubriendo verdades objetivas, cuya existencia es independiente de la actividad matemática³⁸. Nuestro autor se adhiere a la se-

³⁶ El *conjunto de Mandelbrot* es el más conocido de los conjuntos fractales. Fue descubierto por el matemático Benoît Mandelbrot en la década de los setenta. Este conjunto se define en el plano complejo así:

Sea c un número complejo cualquiera. A partir de c , se construye una sucesión por inducción:

$z_0 = 0$ es el término inicial

$z_{n+1} = z_n^2 + c$ es la relación de inducción

Si esta sucesión queda acotada, entonces se dice que c pertenece al conjunto de Mandelbrot, y si no, queda excluido del mismo.

³⁷ Un fractal es un objeto geométrico cuya estructura básica, fragmentada o irregular, se repite a diferentes escalas.

³⁸ Para una profundización en filosofía de las matemáticas se puede consultar Benacerraf y Putnam 1964. Y para una descripción más amplia de las distintas posturas filosóficas básicas se puede consultar Horsten 2008.

gunda postura pero, a la vez, afirma que la cuestión no es tan sencilla. Si bien es cierto que, en ocasiones, se descubren estructuras matemáticas en las que sale mucho más de lo que se introdujo (las califica como *obras de Dios*), también es cierto que en otras ocasiones se inventan construcciones matemáticas artificiales para conseguir algún fin muy específico (*obras del hombre*). Por tanto Penrose acepta que haya cierta construcción de las matemáticas, pero siempre remitiendo a una existencia real.

Algo parecido pasaría en otras áreas de la actividad humana como el arte o la ingeniería. Junto al trabajo ordinario, también se dan sencillas novedades que abren grandes panoramas. Estas innovaciones tienen más de inspiración que de transpiración, son más descubrimiento que invención. Sin embargo, a pesar de estas aclaraciones, para Penrose las matemáticas tienen una unidad y unicidad mayor que el arte o la ingeniería. Por eso, adopta la postura del platonismo matemático, donde los objetos y conceptos no son construcciones que existen en la mente de los matemáticos sino realidades inmatrimales y atemporales³⁹.

Esta postura remite a un artículo escrito por Kurt Gödel en 1932, en el que sostiene que los matemáticos solo pueden hacer teorías que se aproximen a las *verdades matemáticas objetivas*, sin llegar a conocerlas completamente⁴⁰. Gödel sostiene que hay un fuerte paralelismo entre teorías viables de los objetos y conceptos matemáticos, por un lado, y las teorías viables de los objetos y propiedades físicas, por otro. Al igual que los objetos y las propiedades físicas no son construidos por el hombre ni se reducen a entes de razón, tampoco los objetos y conceptos matemáticos son construidos por el hombre, ni se reducen a entidades mentales. Los objetos y los conceptos matemáticos son tan objetivos como los objetos y propiedades físicas. Además, de modo análogo a como percibimos los objetos y propiedades físicas, también, mediante la intuición matemática se

³⁹ Linnebo 2011: «Platonism about mathematics (or *mathematical platonism*) is the metaphysical view that there are abstract mathematical objects whose existence is independent of us and our language, thought, and practices. Just as electrons and planets exist independently of us, so do numbers and sets. And just as statements about electrons and planets are made true or false by the objects with which they are concerned and these objects' perfectly objective properties, so are statements about numbers and sets. Mathematical truths are therefore discovered, not invented».

⁴⁰ Un resumen del platonismo matemático de Gödel se puede leer en Horsten 2008 o en el capítulo 9 de Berto 2009. Y para una comprensión más amplia del platonismo se puede leer el artículo de Linnebo 2011

tiene una *percepción* de los objetos y conceptos matemáticos. Y al igual que la percepción física puede ser errónea y corregirse, también la intuición matemática puede equivocarse, corregirse y mejorar. La diferencia de los objetos y propiedades físicas con los objetos matemáticos es que estos últimos no existen en el espacio ni en el tiempo, ni se crean a instancias del espacio o del tiempo.

Lo que Gödel denomina *verdades objetivas* remite a los objetos matemáticos del segmento inferior del Mundo de las Ideas de Platón. En ambos casos se trata de verdades objetivas, sin principio ni fin, universales e inmutables que residen en un mundo al que solo se puede acceder por medio del pensamiento –en el caso de los objetos matemáticos– y que se reproduce de manera imperfecta en el mundo sensible.

Frente a esta postura se encuentra el formalismo matemático, cuyo autor más relevante es David Hilbert⁴¹ y cuya propuesta es definida por Penrose en los siguientes términos:

«La idea del programa de Hilbert consistía en encontrar para cualquier área bien definida de las matemáticas, una lista de axiomas y reglas de inferencia suficientemente amplia que incorporara *todas* las formas de razonamiento correcto apropiadas para dicha área. (...) Si se acepta que semejante amplio sistema de axiomas y reglas de inferencia nos ha sido ya dado para la aritmética, entonces disponemos de un criterio definido para la ‘corrección’ de la demostración matemática de cualquier proposición aritmética. Existía la esperanza de que tal sistema de axiomas y reglas fuera *completo*, en el sentido de que nos permitiera en principio decidir la verdad o falsedad de *cualquier* enunciado matemático que pueda formularse dentro del sistema».

Hilbert intentó desarrollar un sistema que incluyese todos los tipos de razonamiento matemáticamente correctos. Así, sería posible demostrar que un área bien definida de las matemáticas estaba libre de contradicción: este sistema completo permitiría conocer la verdad o falsedad de cualquier nuevo enunciado matemático, sintácticamente correcto, formulado dentro del sistema. Por tanto, estaría perfectamente encerrado en sí mismo y sería inatacable. Las proposiciones bien definidas serían verdaderas o falsas sin necesidad de preocuparse por su significado externo. Toda proposición bien formulada

⁴¹ Para una acercamiento más detallado a la postura de Penrose sobre la verdad matemática, el platonismo y el intuicionismo, se puede leer el capítulo 4, «Verdad, demostración e intuición directa», de Penrose 1999b

sería verdadera cuando su afirmación fuese demostrable dentro del sistema y falsa cuando se demostrase su negación.

En este sistema cerrado, completo e inatacable, el lenguaje natural quedaría sustituido por un conjunto de signos que obedecería a reglas y operaría sobre relaciones. Las proposiciones se formarían mediante una hilera de signos conforme a las reglas. Los enunciados matemáticos expresados así se apoyarían en unos axiomas cuya verdad es evidente⁴² y, a través de las reglas de inferencia, harían derivar unas proposiciones de otras sin remitir necesariamente al significado del lenguaje, sino reduciéndose a operaciones con signos. El formalismo matemático constituiría así un tipo de nominalismo, donde el requisito indispensable es que no hubiese contradicciones.

Frente a esta propuesta de Hilbert se formularon en 1931 los teoremas de incompletitud de Gödel. Según estos teoremas, en un sistema matemático formal –en concreto, en la aritmética, porque son formulados contra el programa de Hilbert– no es posible demostrar o refutar toda proposición bien definida sin caer en una contradicción. Dicho con otras palabras, dentro del sistema se pueden construir proposiciones verdaderas que no tengan demostración. La consecuencia derivada de estos teoremas es que el concepto de verdad matemática no se puede encapsular en ningún esquema formalista.

Esta consecuencia obtiene su fuerza argumental en el hecho de que los teoremas de incompletitud de Gödel están formulados desde dentro del sistema formal. Ahora bien, una vez quedan demostrados los teoremas, es lógico presuponer que, desde otros ángulos y con otros conceptos, también se pueden formular otros argumentos contra la posibilidad de construir un sistema formal. Es lo que hace Penrose:

«En efecto, ¿cómo vamos a decidir qué axiomas o reglas de inferencia adoptar en un caso cualquiera cuando tratamos de establecer un sistema formal? Nuestra guía en la decisión de las reglas a adoptar debe ser siempre nuestra com-

⁴² Por ejemplo en el sistema formal de Hilbert para la aritmética los signos son 0, sucesor, x... y los axiomas son:

0 es un número natural

Si x es un número natural el sucesor de x es un número natural

El 0 no es sucesor de ningún número natural

Para todo x e y si sus sucesores son iguales entonces x e y son iguales

Dada una propiedad, si 0 tiene esa propiedad y si para un número natural cualquiera la tiene él y su sucesor, entonces todo número natural tiene la propiedad.

comprensión intuitiva de lo que es *autoevidentemente verdadero*, dados los significados de los símbolos del sistema. ¿Cómo vamos a decidir qué sistemas formales son razonables para ser adoptados –es decir, que están de acuerdo con nuestras ideas intuitivas sobre *autoevidencia* y *significado*– y cuáles no? Ciertamente, la noción de autoconsistencia no es adecuada para ello. Podemos tener muchos sistemas autoconsistentes que no son *razonables* en este sentido, en los que los axiomas y reglas de inferencia tienen significados que rechazaríamos como falsos, o quizá no tienen significado en absoluto. *Autoevidencia* y *significado* son conceptos que seguirían siendo necesarios aun sin el teorema de Gödel».

Luego, no desde una argumentación formalista como la de Gödel sino desde una argumentación realista, Penrose señala las carencias del formalismo. A la vez, resalta que la comprensión intuitiva, la autoevidencia y el significado de los conceptos matemáticos son necesarios porque la verdad matemática es más que una construcción humana.

Dando un paso más, al igual que Gödel, Penrose identifica el carácter absoluto de la verdad matemática con la existencia platónica de los conceptos matemáticos. Entraríamos de este modo en un amplio tema de debate que no es directamente objeto del presente trabajo. De momento es suficiente con haber visto la crítica que Penrose hace del formalismo para intentar ver ahora qué entiende por comprensión intuitiva. Para ello distinguiré este tipo de conocimiento de la corriente matemática conocida como intuicionismo.

2.2. La intuición directa como método de descubrimiento

El intuicionismo es una corriente iniciada por el matemático holandés Brouwer en 1924 que difiere en bastantes puntos del enfoque matemático clásico⁴³. Según esta corriente, los objetos matemáticos serían productos de la mente humana, cuya existencia equivale a que se puedan construir⁴⁴. Por

⁴³ Para confrontar la postura de Brouwer con el formalismo se puede leer «Intuitionism and formalism» en Benacerraf y Putnam 1964, 66-77

⁴⁴ Horsten 2008: «According to intuitionism, mathematics is essentially an activity of construction. The natural numbers are mental constructions, the real numbers are mental constructions, proofs and theorems are mental constructions, mathematical meaning is a mental construction... Mathematical constructions are produced by the *ideal* mathematician, i.e., abstracted from contingent, physical limitations of the real-life mathematician. But even the ideal mathematician remains a finite being. She can never complete an infinite construction, even though she can complete arbitrarily large finite initial parts of it».

tanto, la verdad de una afirmación matemática solo se podría probar mediante una construcción mental. Mientras que la matemática clásica acepta como objeto de estudio cualquier estructura matemática que pueda definirse consistentemente, el intuicionismo, por el contrario, solo acepta la existencia de aquellos objetos matemáticos que puedan ser construidos paso a paso en el pensamiento del matemático individual. A estos objetos matemáticos solo se les atribuyen aquellas propiedades que puedan ser inmediatamente captadas por la intuición intelectual del matemático.

Esta intuición difiere del concepto de intuición directa usado por Penrose. Mientras que la intuición directa de Penrose es platónica, los intuicionistas utilizan el concepto en sentido kantiano. La intuición constituye el aspecto fundamental de la actividad matemática, no la existencia real de los objetos matemáticos, porque en la intuición se dan las condiciones para construir el objeto matemático. La matemática es concebida por el intuicionismo como una actividad de construcción introspectiva, que se realiza sin palabras ni símbolos, por mera intuición. El lenguaje y la lógica solo sirven para comunicar a los demás y para registrar los resultados de la propia actividad psicológica.

Una de las consecuencias de este planteamiento es que para los intuicionistas el principio de argumentación clásica del *tercero excluso* no sería válido. Según este principio la disyunción de una proposición y su negación es siempre verdadera ($A \vee \neg A$): todo enunciado es verdadero o es falso. Sin embargo, los intuicionistas conciben la verdad como la prueba y la falsedad como la refutación. Luego para ellos, entre afirmar un enunciado o rechazarlo, entre demostrar A o demostrar $\neg A$, cabe una tercera postura que es la no demostración, en la que se encuentran todos aquellos enunciados que no han sido probados ni refutados: estos enunciados no son ni verdaderos ni falsos. Junto a este planteamiento, para los intuicionistas tampoco son válidas las argumentaciones que usan el método de *reductio ad absurdum*. Dicho método consiste en demostrar que una proposición matemática es verdadera probando que si no lo fuera conduciría a una contradicción. Para Penrose estas observaciones del intuicionismo resultan interesantes pero a su juicio son extremas, precisamente porque rechazan métodos tan fructíferos como el de *reductio ad absurdum* o principios tan fundamentales como el del *tercero excluso*. Para Penrose, este principio es una verdad matemática autoevidente.

Frente al intuicionismo Penrose defiende que los objetos matemáticos tienen una existencia atemporal por sí mismos, sin depender de las construcciones mentales ni de los objetos físicos particulares. Aun así, acepta la co-

riente matemática del constructivismo, que es una parte del intuicionismo, porque piensa que cabría en la matemática clásica. Lo que no aclara es cómo relacionar constructivismo con platonismo matemático. Según mi parecer la construcción matemática quedaría dentro del mundo mental, después de que la mente ha alcanzado los objetos matemáticos.

El modo en que la mente alcanza esos objetos matemáticos, según Penrose, es mediante una *inspiración* matemática. Dicha inspiración se produciría por intuición directa, como una súbita *irrupción* en un mundo platónico. Y tal intuición, exclusiva de los seres humanos, permitiría encontrar nuevas realidades en el mundo platónico para luego tener una comprensión más plena de la realidad física⁴⁵.

Para atisbar qué es la intuición directa, Penrose alude al pensamiento no verbal. Como ejemplo afirma que casi todo su pensamiento matemático es *visual*, en términos de conceptos no verbales. Además, cita a varios autores que describen su pensamiento en términos similares y señala la riqueza de las diferencias entre pensamiento verbal y no verbal, pensamiento analítico y geométrico, pensamiento elaborado e intuitivo.

Posteriormente, basándose en su experiencia personal sobre la transmisión de las ideas matemáticas entre científicos, llega a la conclusión de que la mente, al percibir una idea matemática, toma contacto con el mundo platónico de los conceptos matemáticos, *irrumpe* en él. Y cuando los matemáticos se comunican es porque cada uno tiene un *camino directo a la verdad* matemática a través de un proceso de *visión*. La mente posee la capacidad de ese contacto directo con un mundo platónico. Un mundo que es común para todos y que existe en sí. Por tanto las verdades matemáticas serían verdades necesarias, que más que descubrirse se recordarían, al estilo platónico.

Dejaré aquí esta breve aproximación al esquema mental de Penrose sobre las matemáticas, tanto en su aspecto ontológico, platonismo, como gnoseoló-

⁴⁵ Penrose 1994, 50: «According to Plato, mathematical concepts and mathematical truths inhabit an actual world of their own that is timeless and without physical location. Plato's world is an ideal world of perfect forms, distinct from the real world, but in terms of which the physical world must be understood. It also lies beyond our mental constructions; yet, our minds do have some access to this Platonic realm through an 'awareness' of mathematical forms and our ability to reason about them. We shall find that whilst our Platonic perceptions can be aided on occasion by computation, they are not limited by computation. It is this potential for the 'awareness' of mathematical concepts involved in this Platonic access that gives the mind a power beyond what can ever be achieved by a device dependent solely upon computation for its action».

gico, intuición directa. A mi modo de entender, ambos aspectos de las matemáticas, ontológico y gnoseológico, constituyen el punto de partida para el posterior desarrollo del esquema de los tres mundos de Penrose. Un esquema que no pretende representar la realidad de modo estático, sino dinámico, y que se completa desde otros ángulos.

II. NECESIDAD DE UNA NUEVA FÍSICA PARA ENTENDER LA MENTE

El esquema de los tres mundos de Penrose muestra las dificultades que existen para explicar el entrelazamiento entre lo matemático, lo físico y lo mental. Dicho entrelazamiento es calificado como *misterioso* por Penrose, tanto en las relaciones entre los tres mundos como en la entidad de cada mundo. Los tres mundos de Penrose tienen características epistemológicas y ontológicas propias, porque manifiestan distintos tipos de existencia de la misma realidad. Y a la vez, existe un misterio más profundo que subyace en todo el esquema: cómo tres existencias distintas constituyen la única realidad. Este misterio no tiene una solución fácil y remite al problema clásico, ampliamente estudiado, entre la unidad y la multiplicidad de la realidad. Para el objeto del presente trabajo y por el momento bastará con apuntar algunas consideraciones generales al respecto.

Cuando los tres mundos se estudian independientemente prevalece la inclinación a separarlos, mientras que cuando se busca una comprensión global de la realidad se intenta corregir esa inclinación. Incidir sobre la autonomía propia de cada mundo resulta con frecuencia insatisfactorio, porque son más las dificultades que se plantean que las cuestiones que se resuelven. Además, el propio desarrollo de la actividad científica resalta el entrelazamiento entre las matemáticas, la física y la actividad mental del hombre que conoce.

La presente sección se centrará en las matemáticas. Por eso, comenzaré por resaltar la unidad de las matemáticas con el mundo físico y con el pensamiento humano. Para ello, abordaré en el primer apartado la concepción platónica que Penrose tiene del mundo matemático y que, en sus líneas maestras, se identifica con el realismo matemático de Gödel. Esta aproximación al estatuto de las matemáticas también pretende salir al paso de las dificultades para valorar el alcance de la crítica de Penrose a la inteligencia artificial. Se quiere evitar que las conclusiones se vinculen tanto con el realismo platónico que se consideren inválidas para otras filosofías de las matemáticas. Además,

se quiere mostrar que es posible criticar el realismo platónico sin necesidad de rechazar el argumento de Penrose o el teorema de incompletitud de Gödel.

En el segundo apartado se explicarán las cuatro perspectivas en las, según Penrose, se puede encuadrar la relación entre consciencia y computación. Una vez explicadas, en el siguiente apartado se intentará aclarar qué entiende Penrose por computación. De este modo se podrá apreciar mejor el alcance del *nuevo argumento de Penrose*, cuyo contenido e implicaciones se explicarán en el apartado final.

A mi parecer, este enfoque ayuda a comprender tres aspectos de la obra de Penrose. En primer lugar permite contextualizar mejor su argumento contra la viabilidad de una inteligencia artificial con la física conocida. Además, muestra con más claridad uno de los motivos por los que parece necesaria una nueva física. Y por último, abre las puertas a considerar el papel de la consciencia humana en relación a la física y las matemáticas, lo que se estudiará en el siguiente capítulo.

1. *El estatuto de las matemáticas*

Las matemáticas necesitan ser validadas mediante pruebas y dichas demostraciones comienzan a partir de ciertos postulados fundamentales. Sin embargo, asumir estos postulados no es una tarea banal ya que no son obvios ni autoevidentes. Esta problemática conduce a preguntarse sobre cuáles son los métodos adecuados para hacer matemáticas y por qué son esos los métodos apropiados.

Bajo la segunda cuestión subyacen aspectos problemáticos sobre la ontología y la epistemología de las matemáticas, sobre la verdad, la existencia y el tipo de objetividad necesario en matemáticas. Se trata de cuestiones clásicas –relación entre las matemáticas y las ciencias o entre las matemáticas puras y las matemáticas aplicadas– cuya respuesta ha evolucionado históricamente. Intentaré exponer estas cuestiones conforme las explica Penrose, aunque también quedará manifiesta mi preferencia por una postura similar a un *realismo fino* en continuidad argumental con el último libro de Penélope Maddy, *Defending the axioms*.

Maddy es una autora que ha evolucionado en su pensamiento. Siendo siempre una defensora del naturalismo metodológico⁴⁶ aplicado a las mate-

⁴⁶ Paseau 2010: «Methodological naturalism has three principal and related senses in the philosophy of mathematics. The first is that the only authoritative standards in the philosophy of

máticas, se adhirió primero al realismo platónico, para rechazarlo después y evolucionar hacia un realismo fino. Sin embargo, en su última obra, quizá recogiendo algunas de las críticas recibidas⁴⁷, ha abandonado el realismo. En *Defending the axioms* Maddy aplica el método «científico» (naturalista) al estudio metateórico de la matemática y cuando parece que ha llegado a un «realismo» mínimo, da un paso más hasta llegar a una posición que no se compromete con la existencia de ningún objeto matemático, sino solo con los hechos. La justificación fundamental de las matemáticas sería externa, no interna a las matemáticas, y se sustentaría en la fecundidad de los desarrollos matemáticos.

Sin embargo, desde mi punto de vista, no bastaría con afirmar que la justificación externa de las matemáticas es la más importante, como hace la autora, para defender luego un objetivismo matemático que se apoya en los hechos desnudos, como la postura más razonable. Sino que sería necesario reconocer que no vale cualquier tipo de justificación externa, a elegir desde un realismo fino hasta un objetivismo, como sostiene la autora. La realidad de los hechos matemáticos remitiría a una realidad matemática subyacente, a un realismo fino.

Este será el tipo de realismo fino que adoptaré, un *realismo fino a posteriori*, distinto del *realismo fino a priori* del que habla la autora. El *realismo fino a priori*, sería de corte más deductivo, porque proviene de atenuar las exigencias del realismo clásico, mientras que el *realismo fino a posteriori*, sería de corte más inductivo, porque proviene del hecho científico de que las matemáticas funcionan. El realismo viene afirmado *a posteriori*, porque el hecho de que las matemáticas funcionan apunta más allá de un mero funcionalismo o pragmatismo, hacia una existencia mínima pero real de las matemáticas.

En cuanto a la relación entre matemáticas y ciencia, Maddy distingue una evolución histórica en tres etapas. La primera etapa se remonta al esquema platónico, la segunda se asocia con los pioneros de la revolución científica y la tercera sería en la que nos encontramos actualmente. Durante estas

mathematics are those of natural science (physics, biology, etc.). The second is that the only authoritative standards in the philosophy of mathematics are those of mathematics itself. The third, an amalgam of the first two, is that the authoritative standards are those of natural science and mathematics». Maddy no se sitúa con propiedad en ninguno de estos tres tipos.

⁴⁷ Algunos ejemplos de estas críticas se encuentran en Roland 2007, 423-450 y Moore 2007, 141-165.

tres fases se produce una inversión del punto de vista cuando las matemáticas puras se separan de las matemáticas aplicadas⁴⁸. Esta inversión implicaría que en la ciencia actual ya no se estaría descubriendo la estructura matemática subyacente al mundo, como pensaban los pioneros de la revolución científica, sino más bien construyendo modelos matemáticos abstractos e intentando hacer afirmaciones verdaderas sobre su correspondencia con los hechos físicos. Solo en algunos casos esta correspondencia sería prácticamente isomorfa.

Por contraste, la postura de Penrose se podría definir como una *integración* de las tres etapas. Por un lado, se adhiere al realismo platónico y cree en la existencia de los objetos matemáticos. Por otro, opina que las matemáticas constituyen la realidad última del mundo físico. Y por último, se define como un ferviente defensor del método científico y de su capacidad para desvelar la realidad. La imperfección de los modelos actuales se debería a que no utilizan la matemática correcta y, para sostener esta postura, se apoya en los extraños casos de isomorfismo donde se habría conseguido desvelar la verdad matemática intrínseca a la realidad.

A esta postura *integradora* se pueden hacer dos críticas, una relacionada con el platonismo matemático (tendencia a la separación entre mundos) y otra relacionada con la excesiva vinculación entre las matemáticas y la física (tendencia a la unificación de la realidad). Ambos aspectos son importantes a la hora de mantener el equilibrio necesario entre unidad y diversidad. Sin embargo, en este apartado consideraré con detenimiento el platonismo, cuya crítica me parece más necesaria porque Penrose la afirma con más fuerza. Mientras que no incidiré mucho sobre la diferencia entre las matemáticas y la física, ya que Penrose reconoce que los modelos matemáticos tienen sus límites a la hora de representar la realidad. Aun así, conviene recordar que nuestro autor no parece distinguir con claridad entre uno y otro aspecto.

⁴⁸ Maddy 2011. La autora considera esta inversión como un triple proceso. En primer lugar los matemáticos comenzaron a estudiar conceptos, estructuras y teorías puramente matemáticos, sin perseguir unos objetivos con aplicación inmediata al mundo. Además, surgió una amplia variedad de alternativas a la geometría euclídea que podían aplicarse con éxito a la investigación empírica. Y finalmente, los modelos matemáticos de los fenómenos físicos –cada vez más complejos y de difícil comprensión– ya no pretendían reflejar con exactitud la realidad, sino asemejarse al mundo en algunos aspectos concretos y limitados.

1.1. Tipos de realismo matemático

Decir hoy en día que un matemático es realista no es decir demasiado, a no ser que se especifique un poco. Por eso, resulta necesario aclarar qué lugar ocupa Penrose dentro del complejo debate filosófico en torno al realismo. Ya se explicaron algunos aspectos de su pensamiento sobre las matemáticas, pero aquí me gustaría retomarlos con más precisión y aislarlos del esquema de los tres mundos. Por eso, no me ceñiré al modo en que Penrose presenta su pensamiento, sino que lo situaré en un esquema más amplio, el del platonismo clásico.

El platonismo matemático clásico acepta tres condiciones:

1. Las entidades matemáticas existen.
2. Dichas entidades se pueden conocer y nuestra mejor teoría sobre ellas es, por lo menos, aproximadamente verdadera. Los realistas aceptan que una teoría ideal, que cumple con todos los requisitos teóricos y prácticos, puede ser falsa. Aun así, son *optimistas* respecto a la posibilidad del conocer el mundo y respecto a la verdad de los enunciados matemáticos.
3. Tanto las entidades matemáticas como la verdad de los enunciados sobre ellas, son independientes del sujeto. Es decir, las entidades no son construcciones del sujeto, existen con independencia de nosotros. Esto equivale a afirmar que los sujetos descubren las propiedades de las entidades matemáticas y sus relaciones, no las inventan. Las matemáticas no se construyen, se descubren. Además, la independencia de la verdad implica la posibilidad de que existan verdades que trasciendan la evidencia, es decir, enunciados que puedan ser verdaderos o falsos aun cuando no seamos capaces de probarlos, incluso en los casos en los que sepamos que es imposible encontrar una prueba para ellos.

Aparte de estas tres condiciones generales del realismo, todas ellas aceptadas por Penrose, en el caso del platonismo clásico habría que añadir una cuarta condición:

4. Las entidades matemáticas son abstractas: están situadas fuera del espacio-tiempo y son incapaces de interactuar causalmente.

Una vez conocidas estas cuatro condiciones se puede diferenciar entre un platonismo ontológico y un platonismo semántico. El platonismo ontológico afirma que las entidades matemáticas existen y son independientes de los sujetos. Mientras que el platonismo semántico sostiene que los enunciados

de las teorías matemáticas son siempre o bien verdaderos o bien falsos y que su verdad no depende de los sujetos; que es posible defender la verdad aun en casos en los que no se puede probar. Esas condiciones son definitorias del platonismo matemático pero no resulta necesario aceptar las cuatro para ser una realista. De hecho, es lo habitual. En la filosofía de las matemáticas se encuentran defensores de casi todas las combinaciones posibles de estas tesis. Penrose será defensor tanto del platonismo ontológico como del semántico, sin embargo, no aceptará las cuatro tesis en su totalidad. Respecto al punto tres admitirá que sea posible algún tipo de constructivismo y respecto al punto cuatro, aunque lo considera un *misterio*, acepta que pueda haber una interacción causal entre entidades matemáticas y entidades físicas.

Por otro lado, dentro de la filosofía de las matemáticas, también existen dos grandes corrientes anti-platónicas: el formalismo y el intuicionismo. La primera corriente cuestiona la existencia de las entidades matemáticas, mientras que la segunda rechaza la independencia de esas entidades. Para el intuicionismo (o constructivismo en general), las entidades matemáticas son construcciones de la mente humana y, por lo tanto, la verdad de los enunciados matemáticos no puede trascender la evidencia. La verdad, para los intuicionistas, debe ser reducida a la noción de prueba. Penrose, fiel al platonismo, rechaza tanto el formalismo como el intuicionismo. Sin embargo, nuestro autor admite cierta relación de las matemáticas con la realidad física y cierto tipo de constructivismo. Por lo tanto, las entidades matemáticas no pueden ser absolutamente independientes. La diferencia de Penrose con el platonismo clásico será una diferencia de grado.

Esta diferencia de grado en el platonismo suele distinguir entre un realismo moderado y un realismo clásico (Maddy hablaría de realismo fino y realismo fuerte). Ambos realismos defienden la existencia y la independencia de las entidades matemáticas, pero el realismo moderado también postula que los sujetos juegan un papel activo en la configuración de esa realidad. Está a medio camino entre un tipo de constructivismo (y de anti-realismo semántico) y el platonismo tradicional, pero sin comprometerse enteramente con ninguno de ellos. Por su parte el realismo fuerte también sostiene que la verdad puede trascender a la evidencia y que, en el proceso del conocimiento matemático, el sujeto es más bien pasivo. La postura de Penrose será por tanto la de un realismo tradicional que se aproxima al realismo moderado. Su punto de partida es afirmar la existencia de las entidades matemáticas y, sin renunciar al platonismo matemático tradicional, acercarse a un platonismo moderado.

Frente al platonismo matemático clásico, en los años setenta, Benacerraf y Field desarrollaron dos argumentos, que si se ponen en paralelo dan lugar a un dilema epistémico. Sin entrar en los entresijos de esta crítica, conviene señalar que no afecta al planteamiento de nuestro autor. El punto donde Penrose se distancia de esta crítica es cuando afirma que las entidades matemáticas «son incapaces de establecer relaciones causales». Para Penrose existe una relación *misteriosa* entre el mundo platónico de las matemáticas y la realidad física, en la que no queda excluida la causalidad. Luego Penrose esquivo esta crítica porque no es un platónico estrictamente clásico. Sin embargo, eso no significa que su postura sea la más adecuada para afrontar la comprensión de la relación entre física y matemáticas. En el siguiente apartado compararé brevemente la postura *realista a priori* de Penrose con una postura *realista fina a posteriori* que me parece más acertada. Pretendo así exponer que el realismo de Penrose es un *a priori* insuficientemente fundamentado y presentar una alternativa también realista.

1.2. Una alternativa al realismo de Penrose

Intentar explicar mediante el realismo platónico la fuerte interdependencia que científicamente se observa entre la física y las matemáticas, como hace Penrose, añade al problema nuevas dificultades ontológicas y epistemológicas. La dificultad ontológica aumenta porque, al afirmar la existencia real e independiente de las entidades y propiedades matemáticas, se va metafísicamente más allá de lo que las matemáticas implican por sí solas. A su vez, la dificultad epistemológica también aumenta porque el platonismo requiere de un método de intuición directa que salve la distancia creada con el mundo platónico de las matemáticas. En el fondo, el platonismo matemático agranda el *misterio* de la relación entre física y matemáticas.

Una postura realista más atenuada podría afirmar la existencia de las matemáticas como un conjunto de verdades, con sus métodos de descubrimiento propios, partiendo del papel de las matemáticas en la ciencia. Este papel de las matemáticas sería objetivo, tanto en el contraste experimental de la actividad científica con la realidad, como en las elaboraciones matemáticas de los sujetos. Desde esas relaciones objetivas se podría fundamentar el valor último de las matemáticas. Además, en este caso el método no debería salvar la distancia con otro mundo y el tipo de conocimiento matemático no tendría que ser necesariamente la intuición directa de los conceptos matemáticos. Se eliminarían

así las dificultades ontológicas y epistemológicas creadas por el dualismo, sin perjuicio de la objetividad matemática, de su verdad o de su existencia. Aun así permanecería el *misterio*.

En última instancia, lo que justificaría la verdad, existencia y objetividad matemática no sería un mundo paralelo sino su fecundidad tanto matemática (interna) como física (externa). Además, el método de conocimiento matemático tampoco sería un modo de salvar la distancia entre los mundos, sino el instrumento que ayuda a delimitar los contornos en los que las matemáticas trabajan bien. Funcionaría de modo análogo a como los métodos científicos ayudan a delimitar los confines de la investigación científica. La utilidad, efectividad y fecundidad de las matemáticas garantizarían su estatuto existencial, verdadero y objetivo.

Obviamente, la consistencia de estos argumentos sobre la existencia de la realidad matemática parece menor que si se emplease una argumentación platónico-deductiva. Sin embargo, es más acorde con lo que las matemáticas pueden decir de sí mismas. La existencia de un mundo externo que garantice el uso de un método deductivo es algo tentador para cualquier matemático realista, pero lleva a un dualismo de difícil salida.

Por tanto, el estatuto de las matemáticas sería en la práctica el de una ciencia, o por lo menos asimilable a una ciencia⁴⁹. La justificación última de las matemáticas sería fundamentalmente externa, pero no platónica. En línea con los teoremas de incompletitud de Gödel no tendría una fundamentación interna, pero tampoco dispondría de una roca firme externa donde apoyarse. Sería una justificación *retroactiva*, de modo análogo a como sucede en otras ciencias. Con la investigación, el conocimiento matemático sería aclarado, perfeccionado y profundizado. Las matemáticas abrirían nuevos campos de desarrollo, se justificarían mejor internamente y recibirían más apoyos externos de otras áreas como la física. De este modo, por un lado, aumentaría intensiva y extensivamente la realidad conocida y, por otro, se reforzaría la apertura e interdependencia de las matemáticas con el resto de la realidad.

Esta postura es distinta del planteamiento de Penrose-Gödel. Coincide con ambos autores en que la justificación última de las matemáticas no puede ser interna, conforme se deduce de los teoremas de incompletitud de Gödel, pero difiere en que para buscar esa justificación haya que acudir al realismo platónico.

⁴⁹ Para profundizar en este argumento se puede consultar Maddy 2011.

Según Gödel los axiomas matemáticos no son proposiciones autoevidentes que no requieran de una demostración formal y, por eso, no sirven de fundamento último para el desarrollo de más verdades. Los axiomas, para justificarse, «tendrían que ser asumidos *como mínimo* en el mismo sentido que cualquier teoría física bien establecida»⁵⁰. Sin embargo, ¿cuál es el motivo para ir más allá de ese mínimo hasta afirmar la existencia de un mundo platónico? Ese salto no parece justificado y, al igual que en la ciencia, bastaría con un mínimo de realismo conforme a una matemática *bien establecida*. Es cierto que el objetivo de las matemáticas es lograr teorías consistentes, modos efectivos de organizar y extender el pensamiento matemático, métodos heurísticos útiles para generar nuevas hipótesis productivas... y por lo tanto, que las justificaciones intrínsecas son importantes, pero la justificación última y fundamental es externa.

Me parece que con estas aclaraciones, el paralelismo entre las matemáticas y la física queda resaltado e integrado con más profundidad. Se eliminan los aspectos menos científicos sin ir en detrimento de la solidez de los argumentos matemáticos. A la vez, se resalta la cuestión sobre la eficacia de las matemáticas aplicadas. Por tanto, para el argumento que se va a exponer –que se apoya en el teorema de incompletitud de Gödel– no es necesario afirmar la existencia de un mundo platónico. Basta con señalar que las matemáticas requieren de una justificación externa y que dicha justificación entrelaza las matemáticas con el mundo físico.

2. *Relación entre computación y consciencia*

Las conclusiones a las que se ha llegado en el apartado anterior, sobre la justificación externa de la existencia, verdad y objetividad matemáticas, estimulan a formularse preguntas análogas sobre el mundo mental: ¿resultaría necesario afirmar la existencia de un mundo mental separado?, ¿bastaría considerarlo como un epifenómeno?, ¿qué otras alternativas hay?... Las respuestas son muy variadas. Van desde la defensa del dualismo (mente-cerebro, persona-naturaleza, software-hardware, espíritu-materia, yo-sensibilidad), hasta la negación de la premisa mayor (existencia del espíritu, del yo...), pasando por una colorida variedad de posturas intermedias.

Según Penrose, es difícil aceptar una visión enteramente dualista de la mente y del cuerpo como obedeciendo a distintos tipos de leyes: físico-mate-

⁵⁰ Gödel 1947, 521. El subrayado es mío.

mático-deterministas por un lado y libres por otro. Por eso, considera que lo que controle o describa el funcionamiento de la mente debe ser una «parte» integral del esquema que gobierna las propiedades materiales de nuestro universo⁵¹. Sin embargo, la física conocida o la actividad computacional serían insuficientes para describir el funcionamiento de la mente. Por lo tanto, debe haber algo más que esté fuera de la física conocida y que sea de naturaleza *no-computacional*.

Para explicar esta postura comenzaré por definir las distintas perspectivas que, según nuestro autor, se observan con relación a la posibilidad de crear artificialmente seres conscientes. A partir de ahora cobrará especial relieve la primera parte de *Shadows of the Mind*⁵², donde Penrose profundiza sobre este argumento que ya había expuesto en *The Emperor's New Mind*.

Esta profundización sigue dos vías. Una negativa, de crítica frente a quienes piensan que nuestra mentalidad consciente puede ser, en principio, plenamente concebida en términos de modelos computacionales. Y otra positiva, de búsqueda de los medios físicos donde se pueda dar esa actividad no computacional. La primera, más rigurosa, se expondrá en lo que queda de este capítulo. La segunda, más especulativa, se afrontará en el siguiente.

2.1. Cuatro perspectivas: A B C D

Penrose agrupa los diversos argumentos sobre la relación entre pensamiento consciente y computación en cuatro perspectivas:

A. Todo pensamiento es computación. Con solo realizar los cálculos computacionales adecuados se evocarán los sentimientos de una conciencia consciente⁵³.

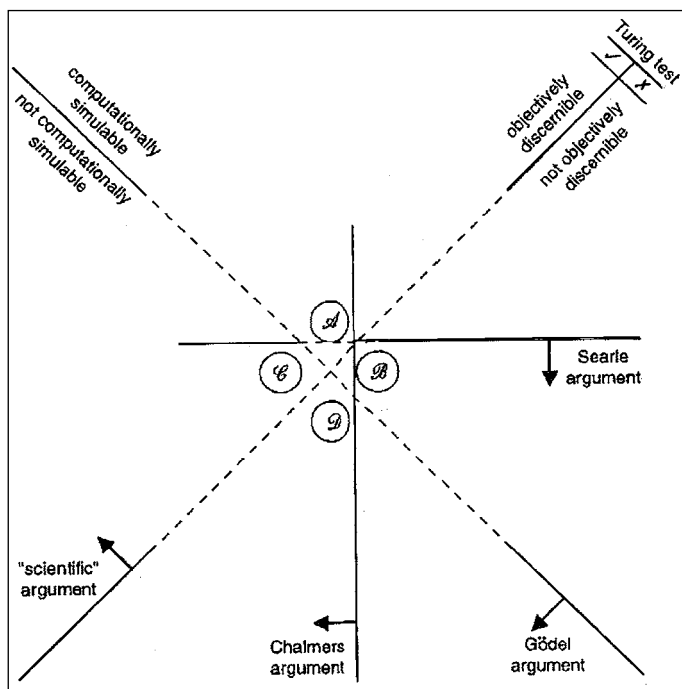
⁵¹ Penrose 1994, 213.

⁵² Penrose 1994, v: «The present volume provides what I believe to be a much more powerful and rigorous case for this general conclusion, and it applies to any kind of computational process whatever». También en el prólogo se observa una analogía entre lo que Penrose pretende explicar y el mito de la caverna platónico.

⁵³ Penrose emplea los dos términos «awareness» y «consciousness» para referirse al fenómeno de la conciencia y, aunque no los define, intenta aclarar la terminología. Además, sostiene que su postura coincide con la percepción intuitiva común que se tiene del significado de esos conceptos. En su esquema: «(a) 'intelligence' requires 'understanding' and (b) 'understanding' requires 'awareness'». Además, 'awareness' sería el aspecto *pasivo* del fenómeno de 'consciousness', mientras que 'free will' sería el aspecto *activo*. Penrose 1994, 37-40.

- B. La consciencia es una característica de la acción física del cerebro. Cualquier acción física se puede simular computacionalmente, pero la simulación por sí misma no puede evocar la consciencia.
- C. Una adecuada acción física en los cerebros evoca la consciencia, pero esa actividad física no puede ser correctamente simulada.
- D. La consciencia no se puede explicar en términos físicos, computacionales o de cualquier otra ciencia.

Ninguno de estos cuatro tipos de relaciones entre pensamiento consciente y computación sería excluyente. Es más, la mayoría de los autores se situarían entre varios planteamientos o adoptarían posturas más flexibles. Pero el objetivo de Penrose no es analizar todas las posibilidades sino las más paradigmáticas. Por eso, se centra en esas cuatro posturas, que las asocia con los planteamientos de cuatro autores: Turing, Searle, Penrose y Gödel respectivamente. Y somete esas posturas a cuatro críticas a las que denomina: argumento de Searle (contra la postura A), argumento de Chalmers (contra la postura B), test de Turing o argumento «científico» (contra las posturas B y D) y argumento de Gödel (contra las posturas A y B).



En el esquema se puede ver cómo dichos argumentos afectan a las distintas posturas y cuáles quedan excluidas. Las letras incluidas en los círculos representan las cuatro posturas. Las líneas horizontal, vertical y las dos diagonales representan los cuatro argumentos de crítica. Y las flechas perpendiculares a cada línea indican qué posturas «sobreviven» a la crítica. Así por ejemplo, el argumento «científico», que mediante el test de Turing distingue si la consciencia es o no es objetivamente discernible, separa las posturas B y D de las posturas A y C. De tal modo que solo las posturas A y C sostienen que la presencia de una consciencia se puede detectar científicamente.

2.2. El argumento de John Searle

La perspectiva A correspondería con la *Inteligencia Artificial fuerte* y sería defendida por Turing. Según dicha postura la actividad mental es simplemente la realización correcta de una secuencia de operaciones bien definidas, como las que realiza cualquier dispositivo con un algoritmo simple. De este modo un ordenador bien programado (o los propios programas) podrían comprender el lenguaje natural y tendrían otras capacidades mentales similares a los seres humanos, cuyas habilidades imita. De acuerdo con la *IA fuerte*, un ordenador puede jugar al ajedrez de forma inteligente, hacer un movimiento inteligente, o entender el lenguaje. De modo análogo, la mente dispondría de un algoritmo extremadamente sofisticado ejecutado con exquisita sutileza, pero nada más. Por tanto el ordenador que poseyese un algoritmo de estas características sería consciente.

Sin embargo, según Penrose, el proceso de entender es mucho más rico que un algoritmo que da la respuesta adecuada. Contra la *IA fuerte*, se dirige el famoso argumento de la *habitación china* de Searle⁵⁴. Este argumento propone un experimento mental por reducción al absurdo, cuyo elemento central es un ser humano realizando una simulación imaginaria de lo que hace un ordenador. El ser humano dentro de una *habitación* sigue unas instrucciones para ordenar y manejar símbolos chinos, aunque no sabe su significado, de modo similar a como un ordenador sigue las instrucciones algorítmicas de un programa. Así, mientras el ser humano manipula los símbolos chinos siguiendo las instrucciones podría parecer que entiende chino, pero en realidad no

⁵⁴ Searle 1980, 417-457.

comprende nada. Lo único que hace es manipular símbolos sobre la base de su sintaxis, pero nunca llega a comprender el idioma.

Otro modo de presentar este argumento es como lo hace Penrose. Nuestro autor presenta una habitación en la que se encierra a una persona sin conocimientos de chino pero con las reglas gramaticales del idioma y un perfecto dominio de ellas. Después, a esa persona se le formulan unas preguntas en chino cuyo significado no entiende, pero a las que puede dar una respuesta adecuada con la ayuda de las reglas. En este caso esa persona podría responder bien pero seguiría sin entender lo que ha respondido.

Searle empleó este argumento para criticar la IA fuerte, a la vez que defendía una *Inteligencia Artificial débil*, según la cual los cerebros serían equivalentes a máquinas de pensar. Para Searle todos los aspectos del entendimiento se podrían simular, pero la simulación por sí misma no implicaría entendimiento. Por tanto, para la *IA débil* los ordenadores serían un elemento útil para áreas como la psicología o la lingüística, porque podrían simular las capacidades mentales, pero eso no significaría que los ordenadores fuesen inteligentes. Frente a la *IA débil* defendida por Searle, Penrose presenta una crítica de David Chalmers.

2.3. El argumento de David Chalmers

La postura B de Penrose, que se aproxima a la versión clásica de la *IA débil*, sostiene que las acciones del cerebro podrían ser simuladas computacionalmente. Aun así, un comportamiento externo similar no sería suficiente para saber lo que el ordenador entiende o siente; y, por lo tanto, para saber si está consciente, porque la consciencia, según Searle, estaría en lo que *siente* y no en cómo *actúa*. Actuar como un sujeto consciente no sería suficiente para garantizar que se es consciente. Por tanto, la presencia de la consciencia no sería objetivamente discernible.

Según Penrose esta postura ha sido criticada por David Chalmers en un argumento que se dirige solo contra la postura B (IA débil) y deja intactas el resto de posturas. El argumento parte del supuesto, que aceptaría Searle, de que en un cerebro humano cada una de sus neuronas podría ser reemplazada en el futuro por un *chip* que funcionase exactamente igual. Si se realizase este cambio individualmente, ante cada nueva sustitución de una neurona, las experiencias internas de la persona deberían permanecer sin cambios. No existiría una enésima neurona cuyo reemplazo provocase la pérdida de cons-

ciencia. Por tanto, concluye Penrose con Chalmers, tampoco la postura B es correcta.

En resumen y una vez consideradas estas dos críticas por separado, si se combina el argumento de Chalmers con el argumento de la habitación china, resulta que la Inteligencia Artificial quedaría excluida en su conjunto, tanto en la versión fuerte (postura A) como en la versión débil (postura B). Luego la Inteligencia Artificial sería insuficiente para explicar el fenómeno de la consciencia. Esta misma conclusión es a la que quiere llegar Penrose con su propio argumento. Un argumento que se conoce como Nuevo Argumento de Penrose, quizá porque constituye una nueva profundización en los argumentos tipo Gödel que había desarrollado Lucas⁵⁵.

Por mi parte, como Penrose ha desarrollado un argumento propio contra la Inteligencia Artificial, no me detendré a considerar con detenimiento los argumentos de Searle y Chalmers. Me centraré, más bien, en explicar el nuevo argumento de Penrose, porque es el más relevante en nuestro autor y porque pretende ser más consistente y completo que los de Searle y Chalmers⁵⁶. Sin embargo, antes de afrontar esta tarea, expondré el último de los argumentos que Penrose presenta contra la postura D y aclararé qué entiende nuestro autor por computación.

2.4. El argumento «científico»

Según la postura D –al igual que la B– la presencia de una consciencia no se podría detectar científicamente, porque no tendría manifestaciones experimentalmente comprobables. Lo que diferencia la postura D de la B es que según la primera el comportamiento de una mente humana no se podría simular computacionalmente, mientras que para la segunda sería posible dicha simulación, aunque eso no signifique la presencia de una consciencia. Penrose está de acuerdo con el *mentalismo* –como él denomina a la postura D– en su afirmación de que la mente humana no se puede simular, pero lo rechaza por sostener que no es científicamente posible detectar si un ser es consciente o

⁵⁵ Penrose reconoce que su argumento se remonta al de John Lucas. «I believe that our positions are very broadly in agreement, although the emphasis that I am placing on the role of the Gödelian argument may be a little different from his». Penrose 1997b, 7 [*Collected Works*, Vol. 6, 70]

⁵⁶ Penrose 1997b, 9 [*Collected Works*, Vol. 6, 72]

no. Nuestro autor defiende que la consciencia se puede detectar científicamente, de un modo análogo a como pretende detectarla el test de Turing.

Dicho test es una prueba propuesta en 1950 por Alan Turing para descubrir la existencia de inteligencia en una máquina⁵⁷. Desde una postura tipo A (IA fuerte) presupone que si una máquina actúa en todos los aspectos como inteligente entonces es inteligente. Durante el test de Turing, un investigador situado en una habitación debe formular preguntas a una máquina y a un ser humano que están situados en otras habitaciones. Su objetivo es descubrir quién es el ser humano y cuál es la máquina, a pesar de que ambos le puedan mentir. La tesis de Turing es que si el jugador y la máquina son suficientemente hábiles el investigador no podrá distinguir quién es quién⁵⁸.

Penrose hace suyo el test de Turing y lo generaliza con un nuevo desarrollo al que denomina argumento «científico». Según su postura mediante métodos científicos se podría llegar a detectar la presencia de una consciencia. El motivo que aduce es que si bien el fenómeno de la consciencia es difícil de explicar en el conjunto de los conocimientos científicos actuales, eso no significa que sea ajeno a la actividad científica. Rechaza la postura *mentalista* porque no es científicamente comprobable, pero considera que el enigma de la consciencia ya contiene suficiente misterio sin necesidad de buscar soluciones fuera de la ciencia. Sin embargo, de lo que no parece consciente Penrose es de que el argumento usado para rechazar el *mentalismo* también se puede usar en contra del platonismo. Ya hay suficiente misterio en la relación entre las matemáticas y la física sin necesidad de buscar soluciones fuera: el mundo platónico matemático tampoco es científicamente comprobable.

Como alternativa científica al *mentalismo*, Penrose sostiene la postura C. Según esta perspectiva, los ordenadores nunca podrán simular eficazmente el comportamiento consciente de un ser humano. Es decir, en el test de Turing, siempre habría algún juez que se daría cuenta de que el ordenador no entiende⁵⁹.

⁵⁷ Para profundizar sobre el test de Turing se puede leer Oppy y Dowe 2011

⁵⁸ Durante las dos últimas décadas se ha realizado una competición anual entre programas de ordenador que sigue el estándar establecido en el test de Turing. Hasta ahora ningún programa ha conseguido ganar la medalla de oro del Premio Loebner, que se entrega a aquella pareja (ser humano-ordenador) que consiga engañar al juez.

⁵⁹ Penrose 1994, 14-15: «But viewpoint C, on the other hand, would not even admit that a fully effective simulation of a conscious person could ever be achieved merely by a computer-controlled robot. Thus, according to C, the robot's actual lack of consciousness ought ultimately to reveal itself, after a sufficiently long interrogation».

Ningún objeto inconsciente se podría hacer pasar por un sujeto consciente; aun así, la presencia de un ser consciente sería científicamente detectable.

Por último, tras este boceto sobre algunas perspectivas en la relación entre computación y consciencia, así como sobre los distintos argumentos que nuestro autor presenta en contra de algunas de estas posturas, afrontaré el nuevo argumento de Penrose contra la Inteligencia Artificial. Pero antes, intentaré aclarar qué entiende Penrose por computación, para poder apreciar mejor el alcance y profundidad de su argumento.

3. *Procedimientos de computación*

La computación es una parte pequeña de las matemáticas, que, en múltiples contextos, sirve para explicar los fenómenos de la realidad física. Si esta afirmación se pone en paralelo con el esquema de Penrose, dónde una pequeña parte del mundo matemático platónico sirve de base para la emergencia del mundo físico, entonces se podría pensar que la computación constituye esa pequeña parte de las matemáticas que puede dar razón de la realidad física. Sin embargo, no es así. Según Penrose, esa parte de las matemáticas que sirve de base para la física no es exclusivamente computacional: la explicación científico-matemática de la realidad es más amplia que la explicación computacional⁶⁰. Hay aspectos científicos y matemáticos que no se pueden simular: ¿Cuáles son? y ¿por qué no se pueden simular?

Para responder a estas preguntas conviene detenerse en qué elementos matemáticos incluye Penrose en el concepto de computación. Ahí se encierran aspectos tan diversos como ordenadores paralelos, redes neuronales artificiales, técnicas heurísticas, capacidad de aprendizaje, sistemas caóticos, aleatoriedad computacional, interacción con el medio u ordenadores analógicos⁶¹.

⁶⁰ Penrose 1997a, 13 [*Collected Works*, Vol. 6, 136]. Penrose no distingue entre explicación matemática y explicación científica. Sin embargo, no parecen que sean lo mismo. Personalmente opino que la explicación científica es más amplia que la explicación matemática. Además, al igual que se puede hacer una crítica matemática para mostrar la no computabilidad de la consciencia, también se podría hacer una crítica científica para mostrar la no-matematicidad de la consciencia. Si afirmamos que hay características físicas no computacionales, también podríamos afirmar que hay características físicas no matemáticas. Sin embargo, así como la computación se puede englobar por completo en las matemáticas, no está tan claro que las matemáticas se puedan englobar en las ciencias.

⁶¹ Penrose 1997a, 14ss [*Collected Works*, Vol. 6, 137ss]

Algunos de estos aspectos, en la actualidad, son difícilmente computables, sin embargo, Penrose prefiere incluirlos en un concepto más amplio de «lo que podría ser computable en el futuro». Por eso, amplía el contenido del término «computacional» para que su crítica tenga mayor alcance. Para él, computación sería todo aquello que puede realizar una máquina de propósito general (una máquina de Turing)⁶². Dicha máquina, en su aspecto matemático, se puede considerar como un ordenador idealizado con capacidad ilimitada de memoria que ejecuta algoritmos sin detenerse ni equivocarse nunca. De este modo, todo aquello que puede ser simulado en una máquina de propósito general sería computacional o, lo que es lo mismo para Penrose, algorítmico.

Veamos la explicación de Penrose sobre aquellos aspectos matemáticos de dudosa computabilidad que se suelen asociar con la inteligencia artificial, como el caos, la aleatoriedad o los procedimientos *top-down* y *bottom-up*. El matemático inglés quiere profundizar en ellos para aclarar que no dejan de ser simulaciones computacionales.

3.1. Top-down y bottom-up

A grandes rasgos, los procedimientos computacionales se pueden agrupar en dos tipos, los que siguen unas reglas para obtener resultados y los que a partir de los datos reelaboran las reglas. Se denomina *top-down* a los primeros y *bottom-up* a los segundos.

El diseño *top-down* fue promovido en los años sesenta por los investigadores de IBM Harlan Mills y Niklaus Wirth, y dominó en la ingeniería de software hasta los ochenta. Entonces llegó la programación orientada a objetos. Ese nuevo enfoque, que se denominó *bottom-up*, pretendía programar módulos en un nivel inferior sin saber exactamente cómo se iban a integrar entre ellos o cómo se iban a conectar al conjunto del sistema.

En el desarrollo de *software* moderno se suelen combinar ambas técnicas. Por un lado, se busca un conocimiento completo del sistema que se considera necesario para un buen diseño (enfoque *top-down*) y, por otro, se usan módulos de programación existentes (enfoque *bottom-up*) para que el sistema *crezca* hasta cumplir con los requisitos iniciales del proyecto. En un procedimien-

⁶² Para una profundización en el concepto de máquina de Turing, se puede leer Barker-Plummer 2012

to *top-down* todo está perfectamente diseñado y especificado para resolver un tipo de problemas. Mientras que en un procedimiento *bottom-up* las reglas de funcionamiento no se especifican con tanta claridad, sino que el algoritmo dispone de un sistema de *aprendizaje* que le permite mejorar los resultados y cambiar las reglas conforme a la experiencia que va adquiriendo. Un ejemplo de procedimiento *top-down* es la búsqueda de números primos, mientras que un ejemplo de procedimiento *bottom-up* pueden ser las redes neuronales artificiales que se usan para el reconocimiento de voz. En la actualidad existen sistemas con resultados cada vez más prometedores que combinan ambos procedimientos. Sin embargo, el hecho de que el sistema pueda *aprender* sigue siendo una cuestión computacional que viene especificada de antemano y que, por lo tanto, puede ser simulada. Es lo que sucede, por ejemplo, con procedimientos que mejoran los resultados de búsqueda o la capacidad de computación; aunque *aprenden* a realizar mejor su tarea, no dejan de ser íntegramente simulados.

3.2. Caos

Según la acepción común, los sistemas caóticos son aquellos donde una mínima diferencia en las condiciones iniciales provoca grandes variaciones en los resultados finales. A esta propiedad, que se denomina «dependencia sensible», se suman otras dos propiedades, no-linealidad y determinismo. En conjunto, las tres propiedades hacen que los sistemas caóticos resulten *controladamente impredecibles*: a largo plazo y en la práctica no se puede predecir bien el resultado, aunque también se sabe que el resultado estará dentro de unos límites. Por lo tanto, los sistemas caóticos se comportan como si no estuvieran determinados, aunque en realidad sí lo están. Un ejemplo de sistema caótico sería la predicción del tiempo⁶³.

El comportamiento de los sistemas caóticos, a pesar de su dificultad intrínseca, se estudia mediante técnicas de computación. Dichos estudios muestran, por un lado, que si las condiciones iniciales son las mismas el resultado final será el mismo y, por otro, que aunque en algunos casos no se pueda predecir el resultado exacto, sí que se pueden predecir resultados *factibles*. Esto implica, según Penrose, que los sistemas caóticos son computables ya que se podría programar en una máquina de Turing de propósito general que dispu-

⁶³ Para una profundización en el concepto de caos se puede leer Bishop 2009.

siera de la suficiente precisión. Por tanto, conviene distinguir entre el procedimiento computacional en sí y las circunstancias que lo limitan. Aunque en un sistema caótico, en la práctica y debido a las limitaciones, no se pueda obtener el resultado con precisión, sin embargo, si se eliminasen las limitaciones, sí que se podría. Luego, con una máquina de propósito general idealizada se obtendría la precisión buscada. El caos sería ejemplo de una no-computabilidad práctica pero no teórica⁶⁴.

Lo mismo se podría decir de la computación analógica. Se puede considerar como una cuestión de precisión, inalcanzable en la práctica pero no teóricamente. En la actualidad, una computación con parámetros físicos continuos se puede simular adecuadamente si los algoritmos tienen la suficiente precisión. Aun así, respecto a la computación analógica, Penrose hace una excepción: distingue entre una posición fuerte y otra débil de su punto de vista, la postura C. Según la postura C débil el fenómeno de consciencia se podría deber a una computación analógica, que en la práctica es imposible de simular con una computación digital. Mientras que según la postura C fuerte, sostenida por Penrose, el fenómeno de la consciencia no se debe ni a una computación analógica ni a una computación digital sino a algún aspecto no computacional de la realidad que todavía no ha sido desvelado por la física conocida⁶⁵.

3.3. Aleatoriedad

Un tercer aspecto matemático que se puede asociar con la inteligencia artificial son los fenómenos estrictamente aleatorios; aquellos que con pro-

⁶⁴ Algunos autores como John Polkinghorne, observan en el caos un tipo de indeterminación que daría cabida a la acción libre humana. Cfr. Monserrat 2005, 363-393.

⁶⁵ Penrose 1994, 25-26: «According to the *weak* version of C, there would have to be physical actions underlying the behavior of the conscious human brain that are non-computable in the standard sense of discrete Turing computability, but which can be entirely understood in terms of present-day physical theories. For this to be possible, it would appear that these actions would have to depend upon continuous physical parameters in such a way that they cannot be properly simulated by standard digital procedures. According to the *strong* version of C, on the other hand, the non-computability would have to come from some non-computable physical theory –as yet undiscovered– whose implications are essential ingredients of conscious brain actions. Although this second possibility might seem far-fetched, the alternative (for C -supporters) is, indeed, to find a role for some continuous action from amongst the known laws of physics, which cannot be properly simulated in any computational way. However, the expectation for the moment must surely be that, for any reliable analogue system of any type that has been seriously envisaged to date, it would be possible –in principle at least– to provide an effective digital simulation of it».

iedad son debidos al azar⁶⁶. Estos fenómenos no existen en el ámbito computacional, pero se pueden simular mediante sistemas *pseudo-aleatorios*. Las diferencias técnicas que existen entre un proceso y otro indican que en los fenómenos estrictamente aleatorios hay un elemento no-computacional. Sin embargo, la pura aleatoriedad, según Penrose, no aportaría diferencias significativas respecto a lo que se puede simular con un ordenador, por lo que tampoco constituiría el tipo de no-computabilidad que está buscando.

Un fenómeno análogo sería la interacción con el medio ambiente. En este caso, cada experiencia del entorno es única e irrepetible. El hecho individual no se puede simular. Aun así, para Penrose, esto no impediría que, como sucedía en los sistemas caóticos, se pueda realizar una simulación típica o plausible del entorno⁶⁷. Se trataría una vez más de una no-computabilidad práctica pero no teórica.

Por otro lado, si se considerase que el fenómeno de la consciencia surge cuando la computación interactúa con el medio ambiente, entonces se podría pensar que hay algo no computable en el entorno. Penrose ve viable esa posibilidad pero considera que, al ser interiorizado por el ordenador, se perdería su aspecto no-computacional.

En resumen, todos aquellos aspectos de las matemáticas de dudosa computabilidad, como la aleatoriedad o el entorno, para Penrose se podrían simular sin que se diese un cambio significativo. Por lo que la pregunta sobre qué tipo de realidad podría ser radicalmente no-computable seguiría abierta⁶⁸.

4. *No-computabilidad en el pensamiento matemático*

Para encontrar una respuesta a la pregunta abierta en el apartado anterior sobre qué tipo de realidad podría ser radicalmente no-computable, Penrose acude a las matemáticas. Su objetivo es demostrar que el entendimiento humano no puede ser una actividad algorítmica y para argumentarlo se apoya en algo que

⁶⁶ Para una profundización se puede leer Eagle 2012.

⁶⁷ Estas últimas consideraciones abren nuevas perspectivas sobre la posibilidad de que la consciencia sea inducida o despertada por un agente externo. Sin embargo, no constituyen un argumento central en el planteamiento de Penrose, por lo que no se estudiará en el presente trabajo.

⁶⁸ Un ejemplo de problema no-computable es el *teselado aperiódico* descubierto por Penrose. Ningún ordenador podría haber encontrado la solución a pesar de ser un problema bien definido. Solo el pensamiento humano era capaz de hallar la respuesta. Penrose 1994, 29-33.

pueda servir de puente entre el entendimiento y la computación. Ese algo es un pequeño aspecto de la actividad cerebral: el pensamiento matemático.

Según Penrose, hay tres razones para buscar ahí. La primera es que con ese argumento se atacaría el punto de vista de la inteligencia artificial en su propio terreno. La segunda es que, según Penrose, solo desde dentro de las matemáticas se podría encontrar cierta demostración rigurosa de algún aspecto de esa actividad consciente no-computacional. Y la tercera es que la computabilidad tiene naturaleza matemática y, por lo tanto, solo desde las matemáticas se podría probar la existencia de algo no-computable⁶⁹.

Su argumento consistirá en mostrar que cuando se realizan juicios matemáticos conscientes sobre la verdad de algunas proposiciones matemáticas bien formuladas sucede algo no algorítmico. Dicho en otros términos, existen clases de problemas matemáticos, bien formulados y con solución, que no encuentran respuesta mediante medios completamente computacionales. Esto se debe, según Penrose, a que los ordenadores no poseen la cualidad humana del entendimiento y no pueden apreciar el contenido de las pruebas matemáticas.

En su crítica se pueden distinguir tres niveles entrelazados. El primero centrado sobre el programa de Hilbert y el formalismo matemático, el segundo que desciende al problema computacional, y el tercero que extrae conclusiones sobre el entendimiento y la consciencia.

4.1. Gödel y el programa de Hilbert

El primer nivel se remonta al formalismo de Hilbert y a los teoremas de incompletitud de Gödel. Es un argumento que afecta al conjunto de las matemáticas.

Hilbert buscaba un sistema formal matemático bien definido: un conjunto suficientemente amplio de axiomas autoevidentes y reglas de inferencia que incorporase *todas* las formas de razonamiento correcto para un área de las matemáticas. Si ese sistema formal fuese *completo* se podría decidir la verdad o falsedad de cualquier proposición matemática, sintácticamente correcta, formulada dentro del sistema. Esa área de las matemáticas estaría libre de contradicción.

⁶⁹ De modo análogo a lo que pretende Penrose, si se admite la existencia de una instancia superior donde se puede incluir todo el pensamiento matemático, desde esa instancia se podría argumentar que la consciencia puede ser no-matemática. Esa instancia superior sería el pensamiento en sí.

Sin embargo, Gödel demostró en 1930, mediante los teoremas de incompletitud⁷⁰, que se pueden hacer enunciados aritméticos verdaderos que van más allá del alcance del sistema formal especificado. No existe un sistema suficientemente amplio donde sea posible demostrar o refutar toda proposición bien definida sin caer en una contradicción. El concepto de verdad matemático no se puede encapsular en ningún esquema formalista.

Por tanto, en cualquier sistema de axiomas y reglas suficientemente amplio y bien definido se pueden encontrar familias de proposiciones matemáticas indecidibles. De esas proposiciones no se puede afirmar su verdad o falsedad, desde dentro del sistema, aunque se conozca que son verdaderas o falsas desde instancias externas al sistema. En esos casos, concluirá Penrose, el entendimiento humano, incluso cuando está limitado al área de los enunciados matemáticos, no puede ser encapsulado en ningún sistema de reglas⁷¹. Aceptar un conjunto de reglas permite trascenderlas.

4.2. Turing y el problema de decisión

El programa de Hilbert se plasma computacionalmente en el *problema de decisión*⁷² y la respuesta de Gödel encuentra su paralelo en varios enfoques análogos que se desarrollaron prácticamente a la vez. De entre estos destacan dos, el que se debe a Alan Turing y su máquina de propósito general y el que expuso Alonzo Church con su esquema de *cálculo lambda*. Ambos autores demostraron que el argumento de Gödel se aplica a cualquier sistema de reglas que puedan ser programadas en una máquina idealizada de propósito general y llegaron a la conclusión de que el problema de decisión no tenía respuesta. A partir de ahora me centraré, como hace Penrose, en la máquina universal de Turing. En este caso, el problema de decisión se reformula en el *problema de parada*⁷³.

⁷⁰ Para el caso particular de la aritmética de Peano (PA) el primer teorema se pueden enunciar: Si PA es consistente, entonces existe un enunciado G tal que ni él ni su negación son demostrables en PA. Y el segundo: Si PA es consistente, entonces el enunciado que representa en PA la consistencia de PA no es demostrable en PA.

⁷¹ Penrose 1998, 157-158 [*Collected Works*, Vol. 6, 256-257].

⁷² Consiste en buscar «si existe un método que permita decidir sobre cualquier problema matemático». Es decir, resolver cualquier problema matemático conjugando simplemente axiomas y teoremas.

⁷³ Un enunciado del problema de parada puede ser: «No hay una función computable $H(x,y)$ que permita determinar si la x -ésima función computable f_x finaliza arrojando un resultado cuando computa el input y ».

Una máquina universal de Turing sería aquella capaz de ejecutar individualmente cualquier tipo de algoritmo⁷⁴ sin error. Siempre daría una respuesta válida o, en el caso de que no existiese, continuaría trabajando sin bloquearse. Ahora bien, ¿esta máquina podría *decidir* si para cierto algoritmo habrá alguna ejecución en la que no se va a detener porque no tiene respuesta válida? Es decir, ¿existe algún algoritmo que programado en la máquina podría decidir sin error qué algoritmos van a encontrar siempre respuesta y cuáles no se van a detener en algún caso⁷⁵? Turing llega a la conclusión de que no puede existir tal algoritmo de decisión.

Como se puede apreciar, en este segundo nivel el argumento matemático de Gödel contra el formalismo de Hilbert se convierte en un argumento computacional contra la posibilidad de encontrar un algoritmo que dé una respuesta al problema de parada. La pregunta por la verdad o falsedad de una proposición bien definida se desplaza a la pregunta sobre si una máquina universal de Turing se parará cuando actúe sobre la *n*-ésima entrada; y la existencia de proposiciones matemáticas gödelianas indecidibles encuentra su paralelo en la imposibilidad de determinar si la máquina de Turing se parará ante la *n*-ésima entrada.

4.3. El nuevo argumento de Penrose

Sobre lo expuesto hasta ahora y en un tercer nivel, aparece el *nuevo argumento de Penrose*⁷⁶. Dicho argumento toma elementos de Gödel, Turing y Lucas para obtener conclusiones sobre el entendimiento humano y la inteligencia artificial⁷⁷.

⁷⁴ Por algoritmo se puede entender todo procedimiento que puede ser realizado por una máquina de Turing. En su estructura más básica consiste en ejecutar unas instrucciones paso a paso a partir de unas reglas completamente especificadas. Estos pasos permiten la evolución desde un estado interno de las variables del sistema a otro, hasta que se llega a una instrucción concreta en la que se detiene. Después de mostrar la respuesta la máquina queda lista para ejecutar otra operación.

⁷⁵ Por ejemplo, no se sabe si un algoritmo con la conjetura de Goldbach encontrará siempre respuestas válidas. Dicha conjetura se enuncia: «Todo número par mayor que 2 puede escribirse como suma de dos números primos»; aunque fue escrita en 1742 por Goldbach en una carta dirigida a Euler con un enunciado equivalente: «Todo número entero mayor que 5 se puede escribir como suma de tres números primos».

⁷⁶ Así es denominado por primera vez en Lindström 2001, 241-250.

⁷⁷ Como se resume en Penrose 1994, 49: «The argument I shall present in the next chapter provides what I believe to be a very clear-cut argument for a non-computational ingredient in our conscious thinking. This depends upon a simple form of the famous and powerful theorem of mathematical logic, due to the great Czech-born logician Kurt Gödel. I shall need only a very

A partir de ahora me centraré en la versión más sencilla del argumento, para lo cual resulta inevitable aludir a los otros dos niveles con los que se entrelaza. El procedimiento será por *reductio ad absurdum* y se apoyará en el *corte diagonal de Cantor*, tal y como hizo Turing al demostrar que el problema de parada tenía una respuesta negativa.

Penrose comienza por denominar $C(n)$ a una computación que se aplica repetidamente y por separado para cada número natural n . Por tanto, $C(n)$ sería la acción de una máquina de Turing –u ordenador idealizado– sobre el número n , que es la entrada a partir de la cual la máquina trabaja⁷⁸. La cuestión a resolver –dado un valor de n – es si la máquina se parará o no.

Para encontrar la respuesta se realiza otro procedimiento computacional (algoritmo A) que contiene todos los argumentos posibles para demostrar convincentemente y sin error que $C(n)$ no se va a detener. Es decir, A es matemáticamente *consistente*⁷⁹. Lo que en este caso significa que:

- Si A se detiene, entonces $C(n)$ no se detiene.

simplified form of this argument, requiring only very little mathematics (where I also borrow from an important later idea due to Alan Turing). Any reasonably dedicated reader should find no great difficulty in following it. However Gödel-type arguments, used in this kind of way, have sometimes been vigorously disputed. Consequently, some readers might have gained an impression that this argument from Gödel's theorem has been fully refuted. I should make it clear that this is not so. It is true that many counter-arguments have been put forward over the years. Many of these were aimed at a pioneering earlier argument –in favour of mentalism [D] and opposed to physicalism [C]– that had been advanced by the Oxford philosopher John Lucas (1961). Lucas had argued from the Gödel theorem that mental faculties must indeed lie beyond what can be achieved computationally (others, such as Nagel and Newmann (1958), had previously argued in a similar vein). My own argument, though following similar lines, is presented somewhat differently from that of Lucas – and not necessarily as support for mentalism [D]. I believe that my form of presentation is better able to withstand the different criticisms that have been raised against the Lucas argument, and to show up their various inadequacies».

⁷⁸ Penrose pone dos ejemplos prácticos: $C(n)$ puede ser «encuentra un número que no sea la suma de n números cuadrados». Y otro «encuentra un número impar que sea la suma de n números pares». En el segundo caso, la respuesta es que la máquina nunca se parará, sea cual sea el valor de n . Y en el primer caso la máquina sólo se parará cuando n sea 0, 1, 2 o 3, dando como resultado 1, 2, 3 y 7. Sin embargo, para probar que no se parará en ningún caso más hace falta una formidable demostración matemática.

⁷⁹ Alonso 2001, nota 30: «En general Penrose utiliza los términos ‘sólido’ [*sound*] y ‘solidez’ [*soundness*] para referirse a la ‘consistencia’ de los sistemas formales, i.e., a sistemas formales en los cuales no debería ser posible demostrar tanto una proposición como su negación. Sin embargo, Penrose utiliza el término sólido de modo más específico para referirse a la noción de ω -consistencia, la cual es más fuerte que la noción de consistencia. Más adelante veremos qué significa esta noción. Por lo pronto podemos simplemente considerar el concepto de sólido

Se generaliza A para que se pueda ejecutar sobre otras computaciones $C_1(n)$, $C_2(n)$, $C_3(n)$... y se denomina $A(q, n)$ a la ejecución de A sobre la computación $C_q(n)$ ⁸⁰. De modo que:

- Si $A(q, n)$ se detiene, entonces $C_q(n)$ no se detiene.

Sobre esta afirmación, Penrose aplica el *corte diagonal de Cantor* –como hizo Turing– y considera el caso en que $q=n$.

- Si $A(n, n)$ se detiene, entonces $C_n(n)$ no se detiene.

Como $A(n, n)$ depende de una sola variable y es un procedimiento computacional, al igual que C, se puede considerar que $A(n, n)=C(n)$. Para no confundirla con el resto de $C(n)$, se puede considerar que es la computación k-ésima:

- $A(n, n)=C_k(n)$

Sobre esta igualdad, como hace la segunda parte del corte diagonal de Cantor, se puede examinar el caso especial donde $n=k$. Así se obtiene:

- $A(k, k)=C_k(k)$

De modo que la afirmación «Si $A(n, n)$ se detiene, entonces $C_n(n)$ no se detiene» se convierte en:

- Si $A(k, k)$ se detiene, entonces $C_k(k)$ no se detiene.

Pero como $A(k, k)=C_k(k)$, entonces resulta que:

- Si $C_k(k)$ se detiene, entonces $C_k(k)$ no se detiene.

como equivalente al de consistente con respecto a los sistemas formales. En la versión castellana de *Shadows of the mind*, Javier García Sanz tradujo el término inglés *sound* como ‘válido’.

De hecho en Penrose 1997b, 10 [*Collected Works*, Vol. 6, 73], el autor reconoce la relación entre ambos términos: «Another advantage is that the notion of ‘soundness’ of a system F, when this notion is restricted to F’s ability to establish Π_1 -sentences, is equivalent to F’s *consistency*». La apreciación es debida a Hilbert. Feferman 1995, 249-256

⁸⁰ Es decir, hay dos algoritmos. El primero se llama C (de computación) que es el que se detendrá o no al resolver un problema. Este algoritmo nunca falla. Si se detiene es que ha resuelto el problema y si no se para es que el problema no tiene solución.

El segundo se llama A (de algoritmo) y actúa sobre C. Contiene todos mecanismos posibles para realizar bien su tarea (indicar cuando C no se va a detener). Si A se detiene es que en algún caso C no se va a detener y si A no se para es que C va a encontrar siempre una solución.

Ambos mecanismos se ejecutan en un ordenador idealizado, que nunca falla. Dicho ordenador podría estar constituido por dos ordenadores idealizados pero físicamente distintos, que al estar interconectados funcionarían como un único ordenador idealizado.

Luego $C_k(k)$ *no se detiene*. Lo que significa que $A(k,k)$ tampoco se detiene. Por tanto, en este caso concreto, el procedimiento A es incapaz de determinar que $C_k(k)$ no se detiene cuando de hecho *ya se sabe que no se detiene*.

En resumen, A tiene la capacidad de determinar que $C_k(k)$ no se detiene porque es consistente. Pero hay un caso concreto en el que A no puede determinarlo, aunque sabemos que $C_k(k)$ no se detiene⁸¹. Por tanto nosotros sabemos algo que A no sabe y que debería saber porque ha sido programada con todas las herramientas matemáticas necesarias. Luego la conclusión es que el entendimiento humano no se puede contener en A ⁸².

Ningún conjunto consistente y cognoscible de reglas computacionales será suficiente para determinar qué computaciones no se paran. Por tanto no se puede encontrar semejante algoritmo. El algoritmo A no puede ser una formalización de los procedimientos que los matemáticos siguen para determinar si una computación no se detiene. El nuevo argumento de Penrose concluye: *Los matemáticos humanos no están usando un algoritmo consistente cognoscible para determinar la verdad matemática*⁸³.

4.4. Alcance del argumento

Algunos autores posteriores a Penrose, como McCall, han desarrollado también sus propios argumentos de tipo gödeliano para sostener la imposibilidad de que un ordenador pueda simular el razonamiento humano. Posteriormente Haim Gaifman ha publicado un artículo en el que analiza el alcance de dichos argumentos, tanto el de McCall como el de otros autores. Para ello inicialmente defiende la posibilidad de partir desde los teoremas de incompletitud y, tras una argumentación correcta, llegar hasta la conclusión ya señalada, aunque por ejemplo McCall no lo consiga. Sin embargo, su artículo se centra en reformular el problema.

La tesis de Gaifman es que *si un ordenador pudiese, de hecho, simular todo nuestro razonamiento matemático*, entonces no podríamos entender completamente cómo trabaja. Es decir, lo que pone en duda es la *consistencia matemática*

⁸¹ Más brevemente: «la consistencia de A implica que A es incompleto».

⁸² En este argumento hay varios puntos que podrían sembrar la duda. Penrose analiza en *Shadows of the Mind* veinte contraargumentos para los que da respuesta y termina reafirmando su tesis. No es el objetivo del presente trabajo analizar cada uno de ellos sino considerar las conclusiones.

⁸³ «Human mathematicians are not using a knowably sound algorithm in order to ascertain mathematical truth».

del algoritmo que hace las veces de un matemático; y precisamente de dicha consistencia dependen tanto las conclusiones finales de los teoremas de Gödel como las del nuevo argumento de Penrose. Para cualquier sistema deductivo formal mínimamente adecuado y conforme al segundo teorema de Gödel, esa consistencia se puede expresar pero no se puede probar dentro del sistema⁸⁴. Esto significa que las conclusiones de los teoremas de incompletitud de Gödel son válidas para rechazar el formalismo matemático de Hilbert, porque en la propuesta de este último la consistencia matemática es un *a priori*. Sin embargo, ¿sucede lo mismo en el nuevo argumento de Penrose? Nuestro autor da por supuesto que sí, pero no está tan claro que un algoritmo que haga las veces de un matemático sea consistente. El problema no está tanto en si el argumento de Penrose está bien elaborado como en asumir la premisa de consistencia matemática. Veámoslo un poco más despacio.

Una primera idea es que un algoritmo que prueba teoremas, genera pruebas dentro del sistema formal. Pero si el algoritmo solo puede saber lo que puede probar, entonces no puede conocer que es consistente porque no lo puede probar. Lo mismo sucede con un algoritmo, que desde un nivel superior (semántico, no sintáctico) pudiese probar la consistencia del nivel inferior: no podría probar la consistencia de sí mismo. La apelación al concepto de consistencia se desplazaría a un nivel superior. También sucedería algo parecido en una red de ordenadores donde el algoritmo de un ordenador prueba la consistencia del algoritmo de otro ordenador.

Sin embargo, para Penrose todos estos supuestos y otros similares se englobarían dentro de una máquina de Turing idealizada, capaz de ejecutar un algoritmo que contuviese todos los razonamientos matemáticos. El punto de partida constituye por tanto un *a priori* hipotético. El algoritmo que se ejecutase en la máquina de Turing, *a priori*, sería consistente. Pero cabría ar-

⁸⁴ El primer teorema de incompletitud afirma que *toda teoría aritmética recursiva que sea consistente es incompleta*. Es decir, no hay ninguna teoría matemática formal capaz de describir los números naturales y la aritmética con suficiente expresividad que sea a la vez consistente y completa. Eso implica que si los axiomas de dicha teoría no se contradicen entre sí, entonces existen enunciados que no pueden probarse ni refutarse. Las teorías aritméticas para las que el teorema es válido son aquellas en las que la deducción de teoremas puede realizarse mediante un algoritmo. El segundo teorema de incompletitud afirma que *en toda teoría aritmética recursiva que sea consistente, Consis T no es un teorema*. Se trata por tanto de un caso particular del primer teorema. Se afirma ahora que una de las sentencias indecidibles de dicha teoría es aquella que «afirma» la consistencia de la misma. Es decir, que si el sistema en cuestión es consistente, no es posible probarlo dentro del propio sistema.

gumentar, como hace Gaifman, que un algoritmo capaz de realizarlo sería en realidad un ser humano.

Por tanto, surge un problema: al formalizar el marco computacional del razonamiento matemático ideal se está aceptando implícitamente la consistencia. Penrose acepta *a priori* que el algoritmo *probador de teorías* es consistente. Sin embargo, no hay motivos suficientes para adoptar esa postura. Más bien, lo lógico sería pensar que no hay una representación exacta del pensamiento matemático, que ningún sistema algorítmico puede representarlo fielmente.

Dando un paso más, la cuestión se puede desplazar a si yo soy intrínsecamente capaz de demostrar la consistencia de mi pensamiento matemático. O más bien estoy sometido a ella, pero no puedo darme cuenta. Si eso es así, entonces –concluye Gaifman– nosotros seríamos máquinas⁸⁵.

Aparece de este modo la cuestión sobre la auto-reflexión matemática. En un primer paso, el matemático reflexiona sobre su pensamiento matemático y se da cuenta de que sus conclusiones se pueden formalizar en un sistema deductivo. En un segundo paso se infiere que el sistema formalizado es consistente, lo que implica que la auto-reflexión realizada es parte del razonamiento matemático. Sin embargo, Gödel muestra que la auto-reflexión no puede abarcar todo el razonamiento porque no se puede probar a sí misma dentro del sistema. Luego, el sistema solo podría reflejar parcialmente nuestro razonamiento, porque el acto de auto-reflexión queda fuera de la imagen que muestra el sistema.

Por tanto, las conclusiones de los teoremas de incompletitud de Gödel dan argumentos para admitir que una plena comprensión de nuestro propio razonamiento matemático es imposible de alcanzar. Podemos confirmar aspectos de cómo funciona, pero no podemos tener una teoría completa y detallada. La razón de la imposibilidad se puede encontrar en que la auto-reflexión introduce al sujeto en la teoría.

4.5. Conclusiones

A mi parecer, el *nuevo argumento de Penrose* se desenvuelve en dos niveles de razonamiento: el nivel del algoritmo en sí (algoritmo A) y nivel del entendimiento humano consciente. Del contraste entre ambos niveles, nuestro autor

⁸⁵ Gaifman 2000, 470: «The upshot is hauntingly reminiscent of Spinoza's conception, on which humans are predetermined creatures, who derive their sense of freedom from their incapacity to grasp their own nature».

infiere la diferencia. El entendimiento humano razonaría de un modo no-algorítmico porque no sería asimilable al proceso algorítmico de un ordenador.

Además, como para Penrose el fenómeno de entender (*understanding*) requiere de la consciencia (*consciousness*), también el fenómeno de la consciencia tendría un aspecto no algorítmico. Pero ese aspecto no estaría sólo en la dimensión pasiva (*awareness*) de la consciencia, sino también en su dimensión activa (*free will*). Según Penrose, con la física conocida, la consciencia (*consciousness*) tanto en su dimensión pasiva (*awareness*) como activa (*free will*) no podría darse en un ordenador, sino que entendimiento, consciencia y libertad serían atributos meramente humanos.

Para recalcar que esos atributos son humanos, Penrose señala que, en la programación de un ordenador, son los seres humanos quienes determinan qué algoritmo es el que se debe ejecutar en cada circunstancia. Y también son ellos quienes proporcionan las reglas que debe seguir y quienes analizan los resultados. Suplantar el elemento de entendimiento humano mediante acciones enteramente algorítmicas es imposible. Aunque la computación proporcione una inestimable ayuda instrumental eso no la desvincula del entendimiento humano matemático, el cual no puede ser eliminado por completo. La diferencia no estriba, por tanto, en la ejecución mecánica de procesos, que podría ser superior en los ordenadores, sino en la capacidad de entender, propia de un sujeto consciente. Ese es el paso que no puede dar un ordenador, aunque tenga todo el conocimiento básico. Hay algo esencial en el proceso de aprendizaje del entendimiento humano que no es posible simularlo mediante ningún medio computacional⁸⁶.

Penrose deduce, en ese contraste de niveles entre el algoritmo y el entendimiento humano, que el entendimiento matemático está más allá de lo que se puede expresar en un sistema computacional. Pero en su argumento quedan algunas rendijas que tanto él como otros autores⁸⁷ reconocen. Esas rendijas hacen que el *nuevo argumento* no sea completamente concluyente, si bien los autores difieren en la profundidad y en los motivos. Mientras que

⁸⁶ Aunque del esquema de Penrose se deduciría que los ordenadores no poseen la intuición necesaria para llegar al mundo platónico, personalmente considero que el proceso de aprehensión matemática se podría realizar mediante un proceso de abstracción, que tampoco sería computable.

⁸⁷ Se puede consultar por ejemplo: Alonso 2001, 139-164, Shapiro 2003, 19-42 o Lindström 2006, 231-237.

Penrose considera que el argumento sí concluye porque esas rendijas serían insuficientes para contrarrestar la solidez de su argumento, los defensores de la inteligencia artificial consideran que las rendijas tienen suficiente entidad para mantener su opinión. El *statu quo* del *nuevo argumento* de Penrose se puede resumir en que no es *matemáticamente* concluyente, como pretendía nuestro autor.

Las conclusiones sobre el argumento de Penrose se podrían dejar aquí, con más razón si se considera que no es un argumento sencillo y ha sido ampliamente tratado por muchos autores competentes. Aun así me gustaría hacer una primera consideración personal sobre el aspecto auto-referencial de la crítica de Penrose –asociado con el corte diagonal de Cantor– para más adelante entrelazarlo con la fundamentación última de las matemáticas. Comenzaré distinguiendo entre una auto-referencia computacional que denominaré recursividad y una auto-referencia humana que se puede denominar reflexividad.

La principal diferencia entre recursividad y reflexividad reside en que la reflexividad es consciente, realiza el juicio y formula la conclusión del argumento matemático de Penrose. A su vez también hay una diferencia en el modo en que ambos términos se relacionan con el concepto de *infinito*⁸⁸. En el caso de la recursividad se relaciona con un concepto de infinito matemático que remite a Georg Cantor, mientras que en el caso de la reflexividad se relaciona con la apertura ilimitada del conocimiento humano. A mi modo de ver solo la reflexividad remitiría a un infinito realmente abierto, mientras que el infinito recursivo se quedaría encerrado en el ámbito matemático.

Para entender mejor a qué me refiero con la apertura ilimitada del pensamiento humano, puede servir la respuesta que Leonardo Polo dio en una entrevista en la que preguntaban sobre este argumento:

«Si comparamos esas dos tesis: por una parte, que el conocimiento es limitado y, por otra, que es infinito (lo que significa que no hay un último objeto, o que en cuanto se llega a alguno, desde él se entiende más), si cabe detectar el límite mental de una manera estricta, *eo ipso* cabe abandonarlo, es decir, ir más allá del conocimiento limitado (...) Para cualquiera que se haya dedica-

⁸⁸ Este concepto resulta muy interesante desde distintos puntos de vista, pero no es el objeto de este trabajo. Para una profundización reciente en el campo de la física se puede consultar Ortiz de Landázuri 2010, 101-114

do algún tiempo a pensar es palmario: nuestro conocimiento no se detiene definitivamente; no es posible que el conocimiento intelectual tropiece con lo conocido de manera que tras de él no haya nada más».

Luego, en la actividad consciente del pensamiento humano auto-reflexivo resulta imposible que exista un «telón de fondo intelectual», porque ese límite se supera al reconocerlo conscientemente. Sin embargo, en la recursividad matemática ese telón de fondo siempre permanece. La actividad matemática no puede salir de su hacer matemático. Solo el matemático con su entendimiento puede sobrepasar el límite. Es decir, cuando Penrose pone en paralelo la recursividad del algoritmo capaz de resolver competentemente todos los problemas matemáticos (algoritmo A), con la reflexividad humana sobre el pensamiento matemático, es el matemático con su pensamiento humano consciente el que juzga tanto la reflexividad propia como la recursividad del algoritmo.

Una segunda consideración personal es que la computación es un área de las matemáticas suficientemente amplia y compleja y, por lo tanto, se le puede aplicar los teoremas de incompletitud de Gödel. Lo que implica que la computación no puede ser un sistema formal que se auto-fundamente en sí misma: no puede ser completo y consistente a la vez. Solo se puede fundamentar desde un ámbito más amplio como puede ser el conjunto de las matemáticas. A la vez el conjunto de las matemáticas se apoya en el fundamento último de las matemáticas, que para Penrose sería el mundo platónico. Sin embargo, como ya se ha argumentado al principio de este capítulo, ese fundamento más bien estaría tanto en la propia actividad matemática (fundamento interno), como en la actividad del pensamiento humano y en la confrontación de las matemáticas con otras ciencias (fundamento externo). Luego para que el nuevo argumento fuese concluyente requeriría de un fundamento exterior a las matemáticas. El argumento de Penrose contra la inteligencia artificial sería insuficiente porque al ser matemático requeriría de dicho fundamento externo. Remitiría tanto al razonamiento humano consciente, es decir, a un razonamiento filosófico a favor o contra la inteligencia artificial, como a una justificación científica, es decir a una confirmación o refutación experimental de que es posible crear inteligencia artificial.

Por tanto, el argumento no puede ser en sí mismo concluyente, aunque puede ser válido si se juzga desde una instancia superior como el entendimiento humano. Es lo que hace Penrose al poner en paralelo dos niveles

auto-referenciales distintos: la reflexividad humana y la recursividad computacional. En realidad su argumento acaba siendo filosófico, porque desde las matemáticas no se pueden extraer las conclusiones que Penrose extrae. Sino que mediante la reflexividad el entendimiento humano se compara a sí mismo con la recursividad computacional para concluir que no son lo mismo. Aunque personalmente pienso que hay más distancia entre reflexividad y recursividad, según Penrose, lo que diferencia a una de otra es el carácter *no-algorítmico* de la reflexividad.

Por contraste, para la mayoría de los defensores de la inteligencia artificial no existiría una diferencia sustancial entre reflexividad y recursividad. Ambos procesos serían auto-referenciales⁸⁹. Así por ejemplo, Alonso sugiere la necesidad de profundizar en la plasticidad, la identidad y la *referencia* entre algoritmos, para saber hasta dónde pueden llegar, porque no se sabe qué puede suceder cuando interactúan todos a la vez. Además, como estas cuestiones se relacionan con la *identidad personal*, tampoco sabríamos si un ordenador podría llegar a tener identidad personal. Penrose lo ve de otro modo. Para él los aspectos de plasticidad, identidad y referencia serían radicalmente distintos en un ordenador que en un ser humano, porque la reflexividad humana no es igual a lo que hacen los ordenadores, es *no-algorítmica*.

En resumen, desde el punto de vista que he presentado la reflexividad humana consciente posee una apertura esencial para la que no puede existir un «telón de fondo intelectual», como diría Polo. Esto permite que el pensamiento humano se juzgue y juzgue otras cosas, de modo que crezca su autoconocimiento y su conocimiento de la realidad de modo ilimitado. Frente a este modo de conocer, la computación podrá crecer mediante recursividad y complejidad relacional, pero siempre en la medida en que participa de la apertura del pensamiento humano. Es decir, en la medida en que el ser humano *introduce* nuevos elementos a los sistemas computacionales. El crecimiento de los algoritmos, según Penrose, se produce por la contribución de los seres humanos con el descubrimiento de patrones y técnicas de computación. El hecho de que un algoritmo crezca en complejidad o en interacciones no le va a hacer consciente.

Además, el nuevo argumento de Penrose tropieza con un problema desde el inicio, cuando pretende hacer una demostración *matemática* inatacable. En

⁸⁹ Hofstadter 2008 considera la percepción del yo como un sujeto *creado* (que no existe realmente) para fundamentar una auto-referencia infinita a partir de las percepciones.

ese intento va más allá de lo que las matemáticas pueden decir de sí mismas, por lo que no concluye. Esto se debe a que las matemáticas no se fundamentan en sí mismas, sino en la actividad del pensamiento humano y en su entrelazamiento con la realidad. Dichos entrelazamientos hacen que las matemáticas sean inabarcables en cuanto alcanzan un mínimo de complejidad, como sucede con la computación matemática.

Por último, la conclusión del argumento de Penrose, en cuanto conclusión filosófica, parece válida. No es un argumento que posea el grado de certeza que quería darle Penrose aunque proporciona un grado de certeza suficiente. Se une de este modo a otros razonamientos de índole filosófica o técnica para negar que lo que hace un cerebro humano sea esencialmente igual a lo que hace un ordenador. Aun así sostener una postura distinta, sea mecanicista o mentalista, sigue siendo posible dentro del marco establecido por Penrose.

4.6. Necesidad de un elemento no-algorítmico

Las conclusiones a las que se ha llegado en el apartado anterior y que se derivan del nuevo argumento de Penrose serían básicamente compatibles tanto con el *fisicalismo* defendido por Penrose (postura C) como con el *mentalismo* (postura D). De hecho algunos autores como Alonso, tras analizar con solvencia el *nuevo argumento*, lo asocian con el *mentalismo*. No es esta la conclusión a la que quiere llegar Penrose. Con su *nuevo argumento* nuestro autor pretende rechazar las posturas que defienden la posibilidad de una inteligencia artificial, pero con el argumento «científico» también quiere rechazar el *mentalismo*. O más bien la postura D, porque nuestro autor prefiere no usar la terminología mentalista-fisicalista sino las posturas definidas por él como A, B, C y D.

En una de sus consideraciones para criticar el mentalismo (postura D) Penrose se pregunta cómo es posible que nuestras mentes parezcan estar tan asociadas con un objeto físico como es el cerebro. Según nuestro autor, observar la cercanía entre el fenómeno de la consciencia y la actividad cerebral es un signo más de la dificultad para separar la consciencia del substrato físico que la sustente. Más bien considera que debe haber una íntima conexión entre consciencia y física y, por eso, busca signos de una acción no-algorítmica, característica de la consciencia, en la física actual.

En esta búsqueda observa que dentro de la teoría física actual casi todo parece computable, salvo la presencia de una aleatoriedad esencial en el proce-

so de medida cuántico y la acción de los cerebros humanos. Por eso, intentará tender un puente entre estos dos mundos buscando una acción cuántica en el cerebro.

La cuestión que queda pendiente es si la física actual es suficiente para dar razón del fenómeno de la consciencia o si por el contrario sería necesario algún cambio radical en nuestra comprensión actual de la física. Si solo prestamos atención a que el fenómeno de la consciencia parece ser exclusivo de los cerebros humanos y a que el pensamiento matemático es no-computable, entonces se podría sostener la postura C débil. Según esta postura para buscar la causa del fenómeno de la consciencia sería suficiente con investigar en la actividad cerebral. La causa última podría deberse a la complejidad de la estructura cerebral, pero en cualquier caso no requeriría de nuevos descubrimientos en la física, sino de una profundización en lo ya conocido.

Sin embargo, a Penrose no le convence la postura débil ya que observa serias dificultades ontológicas y científicas en la comprensión del proceso de medida cuántico. Dicho proceso sería esencialmente inconsistente por su doble dimensión determinista y probabilista. En la situación actual, no sería posible saber con precisión en qué momento tiene lugar el proceso de medida, porque se trata de un proceso esencialmente aleatorio. Y esa aleatoriedad pura, como ya se ha apuntado, es no-algorítmica. Por todo esto, Penrose rechaza la postura C débil y suscribe la postura C fuerte. Para esta última, no basta con buscar la causa del fenómeno de la consciencia en la actividad cerebral, sino que requiere de una nueva física que en sí misma contenga un elemento radicalmente no-algorítmico.

Con su postura, Penrose une dos aspectos que se encuentran en otros muchos autores, aunque normalmente separados. Por un lado, esos autores suelen sostener que bajo el fenómeno de la consciencia debe subyacer algún elemento o alguna característica radicalmente no-computacional, pero consideran que la respuesta se encuentra en la complejidad del cerebro humano. Y por otro, esos mismos autores, también pueden sostener la necesidad de nuevos avances importantes en la física para comprender mejor la realidad. Sin embargo, Penrose es de los pocos que conecta la necesidad de una nueva física con su potencial explicativo respecto al fenómeno de la consciencia. En la siguiente sección se expondrán los argumentos de Penrose para sostener la necesidad de una nueva física y qué aspectos de esta pueden estar influyendo en el fenómeno de consciencia de los cerebros humanos.

III. UNA APROXIMACIÓN HEURÍSTICA PARA UN NUEVO PARADIGMA CIENTÍFICO

La comprensión del mundo representado por la física ha experimentado dos revoluciones en el siglo pasado. La primera unió el espacio y el tiempo en una única entidad geométrica que incorpora el fenómeno de la gravedad. La segunda cambió por completo nuestra comprensión de la materia. Estos cambios se han visto estimulados por el progreso de las matemáticas, que a su vez han servido de catalizador heurístico para el desarrollo de nuevas áreas de investigación. Sin embargo, la simbiosis entre la dimensión física y la dimensión matemática de la realidad se enfrenta a nuevos retos. Estos retos manifiestan los límites actuales de la comprensión de la realidad y, conforme se aclaran, desplazan esos límites hacia nuevas fronteras, de modo similar a como el área de un círculo y su perímetro aumentan simultáneamente. Hasta ahora se han apuntado algunas de estas fronteras y en la presente sección se afrontará uno de los enigmas que opone más resistencia a la comprensión del conocimiento humano: la consciencia.

Se trata de un tema muy amplio y el objetivo concreto consistirá en contextualizar la perspectiva de Penrose al respecto. Nuestro autor ha encontrado dos escollos básicos en los intentos de reducir la consciencia o el pensamiento humano a una *máquina mental*: el *problema de medida* de la mecánica cuántica y el teorema de incompletitud de Gödel en clave computacional. La conclusión a la que llega es que en la naturaleza debe haber algo fundamentalmente *no computacional* que se escapa a la comprensión actual de la física. Por eso, se requiere de un nuevo enfoque. Dicho enfoque debe tener en cuenta los logros alcanzados, los problemas abiertos y las sugerencias heurísticas de muchos autores. La interacción entre estos elementos y las nuevas contribuciones de la ciencia puede desembocar en la aparición de una nueva física que permita dar una explicación más completa de la realidad conocida.

En esta perspectiva encuentran acogida las propuestas de Penrose que presentaré a continuación. Son propuestas que en muchas ocasiones no están exentas de polémica y en otras tantas han sido calificadas por el mismo autor como *no convencionales*. Sin embargo, la necesidad de nuevas y sugerentes aproximaciones es cada vez más patente a los ojos de muchos científicos. Comenzaré señalando algunos motivos físicos y matemáticos que apuntan hacia esa nueva aproximación para después señalar qué efectos podría tener en la consciencia y terminar con un breve análisis crítico de la postura de Penrose al respecto.

1. *Hacia la gravitación cuántica*

Los intentos de unificación de la física en una única teoría siguen caminos muy diversos. Así por ejemplo, los esfuerzos por unificar las cuatro fuerzas fundamentales de la naturaleza se denominan genéricamente *Teorías del Todo* (en inglés Theories of Everything o TOE) o, cuando excluyen la gravedad en su intento, *Teorías de la Gran Unificación* (en inglés Grand Unification Theories o GUT). Entre los que siguen este camino, la perspectiva común es que, para llegar a una *teoría del todo*, primero hay que pasar por una *gran unificación*.

Un enfoque alternativo es el de aquellas propuestas que se centran en unificar las dos grandes teorías de la física conocida. En este caso predomina la denominación genérica de *Gravedad Cuántica* (en inglés Quantum Gravity o QG) y engloba diferentes líneas de investigación como las teorías de cuerdas, la supergravedad, la gravedad cuántica de bucles, la geometría no conmutativa, o la teoría de *twistores* de Penrose, entre otras. En este apartado no se describirán esas posturas, sino simplemente algunos de los motivos que llevan a buscar una teoría unificadora. Así que se mantendrá el término genérico de *gravedad cuántica* para la futura teoría, sea cual sea su configuración final.

Los motivos para buscar esa nueva teoría derivan de algunas deficiencias observadas en las teorías actuales. A grandes rasgos, el paradigma físico actual encuentra dificultades en la explicación de:

- Las condiciones de contorno y las singularidades de la teoría de la relatividad.
- Los aspectos determinista y probabilista del proceso de medida de la mecánica cuántica.
- Los aspectos no-locales del entrelazamiento cuántico.
- La esencial asimetría temporal del universo.
- El aspecto no algorítmico que se percibe en el pensamiento de los seres humanos.

Una nueva teoría de *gravitación cuántica* que englobase las dos grandes teorías desarrolladas en el siglo pasado –la mecánica cuántica y la teoría de la relatividad general– debería responder a alguna de estas dificultades.

A este propósito, la postura de Penrose se hace nítida. Por un lado, está firmemente convencido de la necesidad de buscar esa teoría y, por otro, se distancia del punto de vista convencional. Mientras que la mayoría de los autores sugieren que la relatividad general se debe integrar en la mecánica cuántica, Penrose sostiene la postura inversa: es la mecánica cuántica la que se debe in-

tegrar en la teoría de la relatividad general. Simplificando mucho, esta postura implica que si la teoría de la relatividad general se ampliase hasta incluir las condiciones de contorno, entonces podría dar una explicación de los fenómenos cuánticos que se dan en las singularidades y de ese modo englobar los fenómenos cuánticos ya conocidos.

La mayoría de la comunidad científica sostiene la postura contraria, según la cual, la existencia de singularidades implica que la teoría de la relatividad debe ser reemplazada por una nueva teoría lo suficientemente sólida. La única teoría en condiciones de aportar respuestas implicaría una generalización de la mecánica cuántica que incluyese la gravedad. Esta nueva teoría se apoyaría sobre una unidad mínima de tiempo –es decir, un tiempo discreto– o desprovería de sentido al concepto de tiempo⁹⁰.

Los motivos que aduce Penrose para sostener su enfoque poco convencional son fundamentalmente técnicos, pero se pueden reducir a dos grandes consideraciones. Por un lado, observa que la precisión experimental de la teoría de Einstein es del orden de 1000 veces mayor que en la mecánica cuántica y, por otro, resalta que la solución actual al *problema de medida* es esencialmente insatisfactoria. La nueva teoría de la gravedad cuántica debería dar una respuesta plausible y realista, no meramente funcional, de los aspectos determinista y probabilista del proceso de medida cuántico ya que, según Penrose, los procesos actuales serían aproximaciones generales a un procedimiento único más preciso.

Desde el punto de vista físico, este es el nervio central de su argumentación. Por eso lo consideraré brevemente antes de volver sobre el punto de vista matemático. En esa segunda parte, el aspecto central será la dimensión no algorítmica del pensamiento matemático.

1.1. Ecuaciones dinámicas y condiciones de contorno

En el nivel macroscópico la teoría de la gravedad describe con gran éxito el comportamiento dinámico del universo. Sin embargo, como cualquier modelo matemático tiene sus límites de aplicabilidad, que en este caso son singularidades. Al principio, esos límites se asociaron solo con el modelo matemático

⁹⁰ Lapidra 2008, 244-245. Veremos más adelante estas consideraciones sobre la desaparición o discretización del tiempo.

co pero con el tiempo los datos empíricos han asentado la existencia de dichas singularidades, confirmando la cercanía del modelo con la realidad. Aun así el modelo no explica lo que sucede más allá de las condiciones de contorno. Por eso, se necesita de una nueva teoría de la gravedad que englobe la teoría de la relatividad general y dé explicación de lo que sucede en las singularidades.

a) En los agujeros negros

Las singularidades de la teoría de la relatividad son de dos tipos: la singularidad inicial y los agujeros negros. Sobre el conocimiento de las singularidades del segundo tipo se ha avanzado algo. Así por ejemplo se supone que la entropía de los agujeros negros es un efecto de la *gravedad cuántica* porque su valor depende de la geometría del agujero y es proporcional a las constantes de Newton (gravedad) y de Planck (cuántica)⁹¹. Además, se sabe que los agujeros negros emiten energía en forma de radiación, denominada *radiación de Hawking*⁹² y se investiga la conjetura de la *censura cósmica* formulada por Penrose.

La conjetura de Penrose pretende acotar las soluciones posibles a las ecuaciones de Einstein ya que, en su formulación actual, son compatibles con la existencia de singularidades visibles, o *agujeros blancos*. Sin embargo, según los requisitos de la censura cósmica, dichas singularidades no podrían existir. Ninguna singularidad formada por colapso gravitacional se podría observar directamente ya que todas serían *vestidas* por el *horizonte de sucesos* (delimitado por el momento en que la velocidad de escape se hace mayor que la velocidad de la luz)⁹³. En palabras de Penrose:

«La censura cósmica es básicamente una conjetura matemática –todavía no demostrada ni refutada– concerniente a las soluciones generales de la ecuación

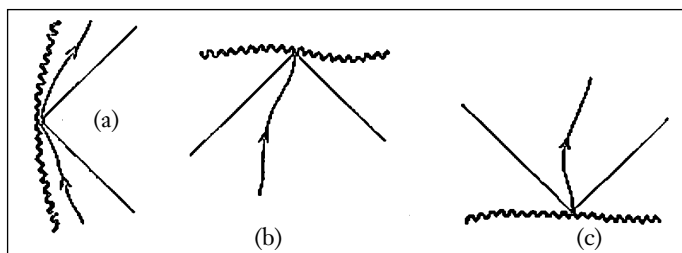
⁹¹ Hawking llegó al valor de la entropía y de la temperatura de un agujero negro a partir de consideraciones sobre la teoría cuántica de campos en un *espaciotiempo* curvo. Es decir, haciendo que la cuántica se adapte a la relatividad y no al revés. La nueva teoría física propugnada por Penrose se inclina por este enfoque, aun así busca una teoría que supere e integre ambas. Penrose 2006, 1105; o Thorne 1995 si se quiere profundizar más en los agujeros negros.

⁹² Como consecuencia del principio de indeterminación de Heisenberg, se pueden crear pares de partícula-antipartícula a partir de un aporte puntual de energía. Estos pares se desintegran rápidamente entre sí, devolviendo la energía usada para su formación. Sin embargo, si ese par se forma en el horizonte de sucesos de un agujero negro, es posible que uno de los componentes del par escape del agujero, produciendo una radiación y disminuyendo la masa del agujero negro. Esta pérdida de masa por efecto cuántico sería tanto más rápida cuanto más pequeño sea el agujero negro y supondría que a la larga todos los agujeros negros desaparecen.

⁹³ En Hod 2008, 346-349 se recoge una de las últimas aportaciones al debate sobre esta conjetura.

ción de Einstein. Si aceptamos esta conjetura, entonces las singularidades espaciotemporales físicas tienen que ser de *género espacio* (o quizá *nulas*) pero nunca de *género tiempo*. Hay dos tipos de singularidades de *género espacio* (o *nulas*), a saber, *iniciales* o *finales*, dependiendo de si las curvas de género tiempo pueden escapar de la singularidad hacia el futuro o entrar desde el pasado».

Para entender esta cita se pueden traer a la memoria los conos de Minkowski. En una singularidad de *género espacio* (o nula), dicho cono sería cortado en horizontal por el centro (o en diagonal junto al borde). Y el resultado obtenido serían dos mitades: una superior(c) y otra inferior (b). Mientras que con una singularidad de *género tiempo* el cono sería cortado en vertical (a)⁹⁴.



La *censura cósmica* niega que existan singularidades de *género tiempo* o *agujeros blancos* (a) y solo admite la existencia de singularidades de *género espacio* (o nulo). Entre estas últimas habría dos tipos: las finales o *agujeros negros* (b), que corresponderían con el cono inferior del «reloj de arena»; y las iniciales o *big bang* (c) que serían el cono superior del «reloj de arena».

Los esquemas considerados inducen a pensar que cuando se atraviesan los límites de contorno de una singularidad final el tiempo desaparece, mientras que a partir de la singularidad inicial el tiempo se crea. Luego parecería que la *censura cósmica* introduce una condición de asimetría temporal en las ecuaciones de Einstein. Lo que se observa más claramente cuando se considera que entre las singularidades inicial y final no puede haber una simetría especular.

⁹⁴ En el esquema de *diagramas de Penrose* se ven las singularidades (con líneas onduladas), la trayectoria de una partícula con masa (con flechas) y los bordes del cono de Minkowski en una sola dimensión (las rectas a 45°).

A esto se añade que, en un universo que no cesa de expandirse, el pequeño efecto de la *radiación de Hawking* acabaría siendo predominante. Por tanto los agujeros negros, serían singularidades *intermedias* que tenderían a desaparecer. Al final existiría el tiempo y una vasta extensión desoladora de un universo que se enfría y se expande cada vez más rápidamente.

b) En la singularidad inicial

Por contraste, en la singularidad inicial las cosas serían distintas. Por una parte, el *espaciotiempo* inicial estaría máximamente concentrado en un universo muy caliente. Y por otra parte, los efectos derivados de esa fase inicial podrían observarse directamente –aunque hasta cierto límite– en los rastros del universo visible. Eso permite acercarse a la condición de contorno inicial donde los efectos de la gravedad cuántica se hacen predominantes. Veamos tres enfoques distintos ante este problema.

En primer lugar, la explicación más aceptada sobre las condiciones iniciales de las singularidades se debe a Hartle y Hawking⁹⁵. Su propuesta elimina las *condiciones de contorno* mediante la hipótesis de que, por debajo del tiempo y de la longitud de Planck⁹⁶, el universo se redondea y se cierra suavemente en una geometría puramente espacial⁹⁷. El universo deja de tener una geometría de *espaciotiempo* para adquirir una geometría de *espacio cuatridimensional* cerrado sobre sí mismo.

⁹⁵ Y la única sería según Penrose 2006, 1026. Posteriormente nuestro autor ha publicado *Ciclos del tiempo*, donde propone una nueva teoría.

⁹⁶ Los parámetros conocidos como límites de Planck definen los límites de validez de la física conocida. Entre estos, se encuentra la temperatura de Planck ($T_p = 1.416785(71) \times 10^{32}$ K) que es la temperatura del universo durante el primer instante del *Big Bang* y el límite por encima del cual cualquier partícula subatómica se rompería. Esa temperatura se alcanza cuando el universo tiene las dimensiones espaciales de la longitud de Planck ($l_p = 1.616252(81) \times 10^{-35}$ metros) y la dimensión temporal del tiempo de Planck ($t_p = 5.39124(27) \times 10^{-44}$ s).

⁹⁷ La geometría lorentziana que rige en el comportamiento del universo, tendría su origen en una geometría riemanniana (con cuatro dimensiones espaciales). El vector tiempo, que en la geometría lorentziana era ortogonal y se plegaba sobre una *superficie* 3-espacial, sería una variable espacial más.

Para comprender este cambio de geometrías ayuda considerar una dimensión temporal y solo dos dimensiones espaciales, e imaginarse una pelota de Badminton. En esa pelota, la semiesfera correspondería con el 4-espacio riemanniano (volumen con 3 dimensiones espaciales porque el tiempo sería una variable espacio más) y las plumas con la geometría lorentziana. Las plumas definen un volumen formado por un círculo espacial horizontal que se expande y se desplaza en el tiempo a lo largo de la vertical.

Estas consideraciones apuntan a que en el substrato de la gravedad cuántica no predominaría ni un *espaciotiempo* fijo, como en la mecánica cuántica actual, ni un *espaciotiempo* relativo, como se observa en la relatividad general, sino un *espacio cuatridimensional*. Sin embargo, esta hipótesis se enfrenta con dos dificultades no menores. En primer lugar, no está claro que en dicho contorno la transición entre espacios sea suave. Y en segundo lugar, tal y como resalta Penrose, esa hipótesis sigue sin ser asimétrica respecto al tiempo.

Ante estas dificultades Penrose propone una segunda alternativa. En su último libro, *Ciclos del tiempo*, formula su teoría de la *Cosmología Cíclica Conforme* (CCC) en la que la asimetría temporal se convierte en el elemento central. En esta teoría, las condiciones de contorno de la singularidad inicial del eón actual enlazan con las condiciones finales de un eón anterior del universo y el inicio del *espaciotiempo* actual coincide con el final del *espaciotiempo* del eón anterior. Luego el tiempo no *desaparece* más allá de los límites de Planck sino que mediante una cadena interminable de eones, enlaza con los tiempos de ciclos anteriores y posteriores. Sin embargo, este enfoque adolece de una clara dificultad explicativa ya que pasa instantáneamente de un universo muy expandido a un universo muy concentrado, lo que matemáticamente implica un redimensionamiento discontinuo del espacio de fases⁹⁸.

Por último, la postura más común en relación a las condiciones de contorno es que constituyen los límites de la continuidad. Más allá de los límites de Planck no es que el tiempo desaparezca sino que junto con el espacio se haría discreto. Las formulaciones actuales de la relatividad general y de la mecánica cuántica consideran que el *espaciotiempo* es continuo, mientras que las nuevas teorías presuponen que en la escala de Planck el *espaciotiempo* se hace discreto⁹⁹. Por debajo de los límites de Planck la geometría del universo dejaría de ser clásica y continua debido a los efectos de la gravedad cuántica. Las motivaciones que llevan a este cambio serían de dos tipos. Se apoyarían, por una parte, en razones físicas como los *saltos cuánticos*, y por otra en motiva-

⁹⁸ Penrose 2010, 141-225. Esa postura contrasta con la sostenida en Penrose 2006, 1130 donde afirma que el espacio de fases *está en equilibrio* entre dos efectos. Uno, debido a los agujeros negros en los que se pierde información y disminuye el espacio de fases y otro por el que, con la reducción del vector de estado, aumenta el espacio de fases. El equilibrio entre un proceso y otro es una característica global que no implica una asociación directa de cada partícula que cae en un agujero con la reducción de estado de otra partícula, como si estuvieran conectadas por efectos no locales.

⁹⁹ Así sucede con las redes de espín de Penrose.

ciones heurísticas, ya que es más fácil tratar con elementos discretos que con los infinitos grados de libertad de las variables continuas.

Sin embargo, los datos aportados por las imágenes de Campo Ultra Profundo del telescopio Hubble ponen en duda esa visión discreta del *espaciotiempo*. Si la longitud y el tiempo de Planck fuesen las unidades mínimas del espacio y del tiempo, se supone que su influencia se apreciaría en las observaciones cosmológicas: dichas observaciones tendrían un máximo de precisión a partir del cual se vuelven borrosas. Pero como la hipótesis de discretización del *espaciotiempo* no concuerda con la nitidez de las observaciones del Campo Ultra Profundo, la estructura básica del universo podría seguir siendo continua.

En resumen, una nueva teoría física debería incluir las ecuaciones dinámicas del universo y explicar qué hay más allá de las condiciones de contorno sin que existiese una transición abrupta. La existencia de singularidades indica el límite del modelo matemático actual que define el universo conocido. Este universo es continuo, temporal y observable, pero el nuevo modelo tendría que ir más allá. Quizá hacia un universo discontinuo, atemporal o inobservable.

Una hipótesis posible es suponer que el *espaciotiempo* descrito por la relatividad «se hace visible» en el límite porque el tiempo surge a partir de la atemporalidad¹⁰⁰; o porque pasa de un sustrato discreto a otro continuo; o porque se pasa de un eón previo al eón actual; etc. Si la nueva teoría no pudiese superar ese límite la única alternativa, según Penrose, sería afirmar la necesidad de un acto divino como el responsable de esas condiciones de contorno. Y sobre ese acto no se puede hacer ciencia¹⁰¹.

La respuesta sobre el sustrato de una futura teoría de la gravedad cuántica queda abierta y su comprensión es relevante para la ciencia y para la filosofía. Las alternativas básicas oscilan entre un *espaciotiempo* continuo y un *espaciotiempo* discreto. En el siguiente apartado se verán algunos de los motivos que llevan a muchos científicos a inclinarse por esta última postura. Entre estos

¹⁰⁰ Según Michael Heller en las condiciones de contorno se daría el paso de la atemporalidad a la temporalidad. La herramienta matemática que describiría el nuevo modelo matemático podría ser una geometría no conmutativa. Heller 2003, 79-99.

¹⁰¹ Penrose 1999b, 437. Este comentario y otras perspectivas, como la de un Dios afinando en las condiciones iniciales del universo, dan a entender que Penrose ve a Dios como un *God-of-the-gaps* que tapa los agujeros explicativos de la ciencia.

motivos resaltan los procesos de cuantización y segunda cuantización¹⁰² que se emplean en la mecánica cuántica y que tienen un papel fundamental para tender puentes entre aspectos continuos y discretos de la teoría. Sin embargo, el punto central será que la nueva teoría debería dar razón de los aspectos determinista y probabilista de la mecánica cuántica.

1.2. Determinismo y probabilismo

En los experimentos cuánticos se observa que al realizar una medida se produce un *salto cuántico*, un cambio discontinuo del estado de una partícula. Al principio, la ecuación de Schrödinger gobierna la evolución continua y determinada, **U**, de una partícula. Pero cuando se realiza una medida, esta evolución se ve reemplazada por un proceso puntual y aleatorio, **R**, que introduce la indeterminación y la probabilidad en la teoría cuántica. Y que localiza a la partícula en un determinado lugar, sin que aparentemente haya una razón objetiva. Es lo que constituye el problema de medida: ¿qué sucede cuando se realiza una medida? Y ¿qué sucede entre medida y medida?

a) Formalismo matemático subyacente

Una primera aproximación podría indicar que, cuando se realiza una medida, se manifiestan los aspectos corpusculares de la partícula; mientras que entre medida y medida se manifiestan los aspectos ondulatorios. Sin embargo, a un nivel más profundo parece que la dualidad onda-corpúsculo, o la más general *complementariedad* de Bohr, es una consecuencia de la estructura matemática de la mecánica cuántica. El principio básico que subyace en este problema es la *no conmutatividad* entre las medidas de posición y momento¹⁰³.

¹⁰² La cuantización es un procedimiento matemático general para construir el modelo cuántico (discreto) de un sistema físico a partir de su descripción clásica (continua). Mientras que la segunda cuantización es el formalismo concreto que se utiliza en la teoría cuántica de campos. Su importancia y utilidad reside en que permite estudiar los campos físicos (continuos) desde el punto de vista cuántico (discreto) y facilita extender la mecánica cuántica no relativista a sistemas con un número de partículas variable. Así permite dar razón de algunos efectos relativistas como los procesos de creación de pares partícula-antipartícula.

¹⁰³ La no conmutatividad implica que el orden de los factores sí altera el resultado. Así por ejemplo, para un álgebra no conmutativa se cumple que $3 \cdot 7 \neq 7 \cdot 3$. Sería algo análogo, aunque con bastantes matices, a lo que sucede con el orden entre dos términos semánticos. Así por ejemplo, en castellano no es lo mismo decir «perfecto blanco» que «blanco perfecto».

En el caso de la mecánica cuántica, esto supone que el orden con el que se aplica una serie de operadores sobre una función no es indiferente¹⁰⁴. Si x es el operador posición de la función de onda de partícula y p es el operador momento, estos dos operadores serían conmutativos cuando aplicados en distinto orden sobre una función diesen el mismo resultado. Este sería el comportamiento de la mecánica clásica:

- $xp - px = 0$

Sin embargo, en mecánica cuántica el resultado que se obtiene es distinto:

- $xp - px \neq 0$, en concreto $xp - px = i\hbar$ ¹⁰⁵

Lo que implica que el comportamiento de esos dos operadores no es conmutativo y que, por lo tanto, requieren de otro tipo de herramienta matemática distinto al que usa la mecánica clásica. Esta nueva herramienta puede ser algún tipo de álgebra no conmutativa.

Transcrito a nomenclatura matemática, $xp - px$ es igual que $[x, p]$ y se denomina *conmutador de posición y momento*. Para una partícula clásica dicho conmutador sería cero ($[x, p] = 0$) mientras que para una partícula cuántica sería mayor o igual que la constante de Planck ($[x, p] = i\hbar$).

Por otro lado, en un álgebra conmutativa se cumple que:

- $[a, b] = 0$

Mientras que en un álgebra no conmutativa:

- $[a, b] \neq k$

donde k es el *parámetro de deformación*.

Por tanto llegamos a dos conclusiones. En primer lugar, desde el punto de vista matemático la mecánica cuántica se puede considerar como un tipo de álgebra no conmutativa, cuyo parámetro de deformación solo depende de la constante de Planck. Y, en segundo lugar, cualquier álgebra conmutativa se puede convertir en un álgebra no conmutativa, si se le introduce una perturbación adecuada. Es decir, sería posible pasar del modelo matemático de comportamiento clásico (conmutativo) a un modelo cuántico (no conmutativo) mediante una adecuada perturbación, introducida por la constante de Planck.

¹⁰⁴ Al inicio de la mecánica cuántica el sistema se consideraba de modo clásico, formado por partículas con sus variables, pero con la teoría de campos cuánticos esos conceptos fueron sustituidos respectivamente por función y operador.

¹⁰⁵ Si se sustituye $xp - px$ por la nomenclatura usada hasta ahora ($\Delta x \Delta p$), con los ajustes convenientes se obtiene la relación de indeterminación de Heisenberg: $\Delta x \Delta p \geq \hbar/2$.

Esta es una de las actuales líneas de investigación de la gravedad cuántica. Se engloba dentro de la búsqueda de las condiciones en las que la realidad física pasa de un sistema cuántico a un sistema clásico. En términos físicos, eso significa buscar cuándo y cómo el parámetro de deformación (la constante de Planck) se hace cero o tiende a cero.

Los efectos de la no conmutatividad también se aprecian en la veintena de parámetros que se emplean en mecánica cuántica. Algunos de estos son conmutativos mientras que otros, como posición y momento o energía y tiempo, no. En concreto, estas dos parejas de parámetros se asocian con las dos relaciones de indeterminación de Heisenberg¹⁰⁶. Cuando se realiza una medida en la que los dos parámetros están involucrados se observa la consabida falta de precisión.

Por otro lado, entre las dos parejas hay un paralelismo consistente con la teoría de la relatividad. Según esta existe una única variable *espaciotiempo* y también una única variable *momentoenergía*. Luego si hay una no conmutatividad entre espacio y momento, también debería haber una no conmutatividad *paralela* entre tiempo y energía, como así sucede. Combinadas, ambas parejas remitirían a una no conmutatividad entre *espaciotiempo* y *momentoenergía*. Los cuatro parámetros que se usan en mecánica cuántica se comportan de un modo acorde con la conclusión relativista de que a gran escala esos cuatro parámetros son dos.

Se observa en estos comentarios otra línea de investigación y otro punto de enlace entre cuántica y relatividad, en el que queda mucho por hacer. Si bien la mecánica cuántica se puede dotar de un formalismo matemático con álgebras no conmutativas, en la actualidad resulta difícil reformular la relatividad, que es una teoría geométrica, en un formalismo no conmutativo. La investigación se inclina hacia geometrías no conmutativas, análogas a la representada por el teselado aperiódico de Penrose.

Estas consideraciones sobre los distintos formalismos matemáticos que subyacen en las teorías lleva al dilema radical de la relación entre matemáticas y física: ¿se trata simplemente de un formalismo que no nos ofrece ninguna imagen de la realidad o hay algo más? La respuesta de Penrose ya se conoce: si las matemáticas que se emplean son las adecuadas, entonces reflejan con precisión la realidad.

¹⁰⁶ Para posición y momento, $\Delta x \cdot \Delta p \geq \hbar/2$; y para tiempo y energía, $\Delta t \cdot \Delta E \geq \hbar/2$

b) Ontología de la mecánica cuántica

La pregunta anterior sobre si las teorías reflejan la realidad es similar a la duda sobre la visión determinista y probabilista del proceso de medida: ¿esta visión de la mecánica cuántica refleja la realidad? Ante la dificultad de unificar dos enfoques tan distintos en una visión coherente de la realidad, Penrose se inclina a pensar que hace falta una nueva teoría física, mientras que otros autores defienden interpretaciones alternativas¹⁰⁷.

Desde los primeros instantes de la mecánica cuántica se puede distinguir algunos autores, como Bohr, para los que la función de onda (ψ) que describe el proceso **U** no describe ninguna realidad en el nivel cuántico, sino que es una convención matemática útil para hacer predicciones. Frente a estos, surgen los que consideran que el vector de estado (ψ) sí describe la realidad física en el nivel cuántico y que el fenómeno de reducción del vector de estado o colapso de la función de onda (**R**) es una aproximación o ilusión. Según esta postura, la superposición cuántica también se daría en el nivel macroscópico, aunque por alguna razón no se podría detectar. Por último estarían aquellos para quienes la ecuación de Schrödinger es una aproximación a algo todavía bastante desconocido que está entre los niveles cuántico y clásico. Es aquí donde algún efecto cuántico se haría presente y la *no computabilidad* postulada por Penrose jugaría un papel esencial¹⁰⁸.

En estas tres grandes líneas interpretativas, que se acaban de exponer, se engloban las ontologías actuales de la mecánica cuántica. Penrose las agrupa en dos tipos: convencionales y no convencionales. Examinaré brevemente las del primer tipo (interpretación de Copenhague, multiversos y decoherencia por el entorno) y comentaré en cuarto lugar la postura no convencional sostenida por el propio Penrose¹⁰⁹.

¹⁰⁷ Así por ejemplo, Ramón Lapiedra defiende una intervención subjetiva creadora en la observación del proceso de medida. Según él «no hay bastante realidad objetiva en el mundo para dar siempre cuenta de nuestra experiencia cuántica porque, a menudo, esta experiencia crea una parte de esta realidad» (Lapiedra 2008, 167). El colapso de la función de onda *aportaría* a la realidad objetiva en el nivel cuántico y sería la base de la acción de la libertad en los hombres. La *decisión* cuántica de dónde colapsar, estaría amplificada al nivel macroscópico en los cerebros humanos para dar cabida a la libertad humana. Lapiedra 2008, 176-177. Según esta postura, hay dos niveles de *acausación* en la realidad: la indeterminación cuántica y la libertad humana.

¹⁰⁸ Penrose 1998, 170-171 [*Collected Works*, Vol. 6, 268-269]

¹⁰⁹ Para una descripción más detallada de las distintas posturas que considera Penrose se puede consultar Penrose 2006, 1049-1092.

Según la lectura que Penrose hace de la *interpretación de Copenhage*, el vector de estado (ψ) no representa la *realidad* del nivel cuántico, sino el *conocimiento* que el experimentador tiene del sistema. Luego el *salto* que ocurre en el procedimiento de reducción del vector de estado (**R**) es simplemente un *salto* en el conocimiento que el experimentador tiene del sistema y no un cambio físico.

Para los defensores de los *multiversos*, por el contrario, el vector de estado representaría la realidad (ψ), mientras que la reducción del vector de estado (**R**) no sería un fenómeno real. En consecuencia, los resultados de una medida *coexisten* entrelazados en una gran superposición lineal cuántica¹¹⁰. De esta postura, según Penrose, se deduce que la verdadera realidad no serían los mundos alternativos sino un *ómnium* que contendría todas las superposiciones en un vector de estado (ψ). Ese vector englobaría todo el universo, incluyendo los diferentes estados mentales del experimentador. Por lo tanto, al igual que en la interpretación de Copenhague, ψ tampoco estaría describiendo la realidad *observada*. En ambas posturas son las percepciones del experimentador humano las que dan sentido a la realidad observada y, por tanto, ni el proceso **U** ni el proceso **R** serían reales.

La tercera postura, *decoherencia por el entorno*, es más pragmática que ontológica y está considerada como la interpretación predominante entre los teóricos de la mecánica cuántica. Si para algunos autores de las dos interpretaciones anteriores el proceso **R** tendría lugar cuando una *consciencia* observa la realidad, para los defensores de la decoherencia por el entorno la reducción de estado **R** se puede entender como el resultado del entrelazamiento inextricable del sistema con su entorno.

Dicho entorno, constituido por infinidad de partículas en movimiento aleatorio, resulta inobservable en la práctica. Pero los efectos del entorno sobre un sistema concreto pueden ser *ignorados* para obtener una descripción

¹¹⁰ Como se recoge en Lapiedra 2008, 100, resulta llamativo que algunos científicos sostengan esta propuesta: «Lo menos que puede decirse de una interpretación de la mecánica cuántica como ésta, que puebla el mundo total de una miríada exponencialmente creciente de diversos mundos, (...) es que no es precisamente una interpretación muy económica a la hora de intentar describir la realidad. (...) El caso ha de tener su interés para la sociología de la ciencia, a la hora de plantearse por qué una propuesta tan inverosímil ha podido gozar de una audiencia tan grande. En cualquier caso, para nosotros esta propuesta de interpretación de la mecánica cuántica debería ilustrar, una vez más, la profunda desorientación de la comunidad científica en la materia».

adecuada del sistema. El objeto matemático que se emplea para describir el sistema se denomina *matriz densidad*¹¹¹ y sustituye al vector de estado.

La ontología de la matriz densidad no está clara pero su uso permite redefinir el estado del sistema cuántico como un conjunto de probabilidades. En términos técnicos, eso significa que si el entorno se modela de una forma adecuada, entonces en un período de tiempo muy corto la matriz densidad se hace diagonal. Lo que a su vez implica que el proceso de reducción de estado **R** puede entenderse como una consecuencia del entrelazamiento con el entorno. La matriz densidad no sería por tanto un proceso real sino una aproximación a la verdad cuántica de lo que está sucediendo. Se trata de una solución pragmática o *para todo propósito práctico* (FAPP, *for all practical purpose*), que se conforma con que la teoría prediga bien los resultados.

Penrose, por el contrario, considera que la cuestión sobre la ontología tiene una importancia crucial. Por eso, frente a estas tres posturas, presenta su propia interpretación de la paradoja de la medida. Su postura se resume en considerar que la formulación convencional de la mecánica cuántica es provisional y que se necesita una *nueva teoría física* para dar un sentido *real* a la paradoja de la medida. Las interpretaciones convencionales resultan insatisfactorias porque difuminan la realidad de lo que está pasando. La nueva teoría no puede ser un mero retoque, sino que debe suponer un cambio sutil pero sustancial¹¹².

En el fondo del planteamiento de Penrose subsiste la creencia de que la mecánica cuántica está descubriendo algo real que las matemáticas pueden reflejar con precisión. Las matemáticas son algo más que una herramienta para conocer la realidad, son el substrato mismo de la realidad. Este enfoque le lleva a rechazar la *matriz densidad* –en cuanto a su poder explicativo aunque no predictivo– y a sostener que tanto el proceso **U** como el proceso **R** son objetivos. En el siguiente apartado veremos qué motivos aduce para sostener esta postura y qué tipo de fenómeno puede aportar ese cambio sutil de perspectiva.

¹¹¹ La matriz densidad combina las probabilidades clásicas con las amplitudes de probabilidad cuánticas, sin distinguir directamente unas de otras. Resulta muy útil en la práctica, aunque no queda claro su estatuto ontológico.

¹¹² Penrose 2000, 223-224 [*Collected Works*, Vol. 6, 446-447]

1.3. Gravedad, reducción del estado cuántico y asimetría temporal

Un precedente de cambio sutil aunque radical de perspectiva es el que surgió entre la teoría gravitatoria de Newton y la de Einstein. Desde el enfoque del primero las fuerzas gravitatorias se suman linealmente, mientras que desde la perspectiva del segundo dicha linealidad es sustituida por una no linealidad muy sutil debida a la geometría del espaciotiempo. Este tipo de cambio radical con unos efectos sutiles es el que Penrose augura para la mecánica cuántica.

Uno de los lugares donde esa *no linealidad* sutil se podría manifestar es en la paradoja de la medida, porque se trata del fenómeno cuántico que más interrogantes suscita: ¿qué está pasando entre \mathbf{U} y \mathbf{R} ?, ¿en qué sentido se puede hablar de salto discontinuo?, ¿ese salto es real o solo una consecuencia del formalismo matemático empleado?, ¿cuál es la causa de ese salto?, ¿se debe a algún proceso atemporal o no local?, ¿se relaciona con la no conmutatividad de los operadores?...

En su búsqueda de alguna respuesta nuestro autor considera que ambos procesos son un efecto de la gravedad cuántica y valora como fundamental la hipótesis –sostenida también por autores como Ghirardi, Rimini o Weber¹¹³– de que \mathbf{R} es una reducción *objetiva* (**OR**) del vector de estado. Este planteamiento difiere de las ontologías convencionales, ya que estas no consideran que la reducción del vector de estado sea algo objetivo, que acontezca en la realidad, fruto de la gravedad (*OR gravitatoria*), sino más bien una consecuencia de los modelos matemáticos que se usan.

Para defender la necesidad de una nueva postura donde la gravedad juegue un papel más esencial, Penrose observa que el teorema CPT¹¹⁴, todavía no comprobado pero fundamental para todas las teorías cuánticas de campos, presenta algunas dificultades con relación a otros conocimientos básicos de la física actual. Dicho teorema, en primer lugar, postula que el *espaciotiempo* debe ser plano, lo que implica una esencial incompatibilidad con el *espaciotiempo* curvo que se

¹¹³ Ghirardi, Rimini y Weber 1986, 470-491 y Ghirardi, Grassi y Rimini 1990, 1057-1064

¹¹⁴ El teorema CPT de la teoría cuántica de campos afirma que toda interacción física es invariante si se le aplican las tres operaciones de conjugación de carga (C), inversión de paridad (P) e inversión temporal (T). Simplificando mucho, esto significa que existe un comportamiento especular básico en la física cuántica. Sin embargo, dicha simetría que afectaría a la carga, a la paridad y al tiempo, se sabe que no es válida cuando se aplica por separado a C, a P, a T y a CP. Un ejemplo de ruptura de simetría es la que se observa en la fuerza electrodébil.

observa en la teoría de la relatividad general. Además, en segundo lugar, sostiene la existencia de una simetría cuántica esencial que se apreciaría especialmente en los primeros instantes del universo y que, con el transcurso del tiempo, se habría roto dando lugar a las asimetrías que se observan en la actualidad. Sin embargo, tal simetría no se asocia bien con la asimetría esencial que se observa tanto en el proceso **U/R** como en el universo a gran escala. Se llega por tanto a la existencia de un conflicto entre los principios de la cuántica y los de la relatividad, que tiene como elemento central el tiempo. El dilema está abierto y Penrose se inclina en favor de la curvatura *espaciotemporal* y de la asimetría temporal.

Sin embargo, al intentar poner en contacto la asimetría temporal con la curvatura *espaciotemporal* se hace manifiesto que la formulación actual de la relatividad no es asimétrica. Son los datos empíricos y las consecuencias de la teoría –el estudio del *espaciotiempo* relativo y de sus singularidades– las que remiten a la asimetría temporal¹¹⁵. Dicha asimetría sería una de las características de la nueva teoría de la gravedad cuántica y supondría que tanto el teorema CPT de la mecánica cuántica como la formulación actual de la relatividad general son falsas, en la medida en que no son asimétricas. Dicha asimetría sería un efecto de la gravedad ya que esta es la única fuerza que se comporta de forma diferente al resto.

Por tanto, el camino para buscar la unificación de las cuatro fuerzas no pasaría tanto por buscar simetrías en la naturaleza, aunque resulte útil, como por estudiar el modo sutil en que los efectos gravitatorios se manifiestan en la asimetría temporal cuántica. Veamos qué tipos de efectos gravitatorios son los que se podrían dar en el proceso de medida.

Una primera aproximación a la superposición cuántica manifiesta que la noción de evolución unitaria (**U**) de Schrödinger requiere un único operador

¹¹⁵ La hipótesis de *censura cósmica* propuesta por Penrose sería un intento de introducir la asimetría temporal en las leyes simétricas de la relatividad. Mientras que la hipótesis de la inversión de la flecha del tiempo sostenida por Hawking sería un intento de conservar la simetría. Cfr. Hawking y Penrose 2010, 65-67, donde en la última página Hawking escribe:

«I wrote a paper claiming that the arrow of time would reverse when the universe contracted again. But after that, discussions with Don Page and Raymond Laflamme convinced me that I had made my greatest mistake, or at least my greatest mistake in physics: the universe would not return to a smooth state in the collapse. This would mean that the arrow of time would not reverse. It would continue pointing in the same direction as in the expansion».

Las recientes observaciones a gran escala de un universo en expansión acelerada sostienen mejor la postura de Penrose e impulsaron a Hawking a reformular su postura, eliminando la inversión temporal y su teoría del *Big Crunch*.

de traslación tiempo. Mientras que, visto desde la relatividad, cada uno de los *espaciotiempos* involucrados en los estados superpuestos debería tener una traslación de tiempo propia. Identificar los dos estados como si fueran uno solo requeriría una correspondencia punto a punto entre los *espaciotiempos* superpuestos y eso supondría una violación del *principio de covarianza general* de la relatividad general¹¹⁶. Por tanto, **U**, no es un proceso compatible con la relatividad general, ya que solo define una variable tiempo y , por lo tanto, un *espaciotiempo* fijo o plano como el definido por la relatividad especial.

Además, considerar que **U** es real y que los *espaciotiempos* de los estados superpuestos no se pueden unir con una sola variable tiempo, implica la existencia de una incerteza esencial en el operador tiempo-derivativo de Schrödinger. Esta incerteza se interpreta como una energía incierta, E_G , en el estado superpuesto. Y por el principio de indeterminación de Heisenberg esa energía es consistente con una superposición fundamentalmente inestable que tendría un tiempo de vida del orden de $t_G = \hbar/E_G$. Lo que significa que cuanto mayor fuese la energía incierta (E_G) menor es el tiempo de vida (t_G) de la superposición, según un factor de escalado de la constante reducida de Planck (\hbar).

Dicha energía, en el caso de la superposición de dos geometrías *espaciotemporales*, se puede definir como la autoenergía gravitacional de las diferencias entre las distribuciones de masas de los dos estados en cuestión, es decir, la energía que se gana al reunir una distribución de masa a partir de las masas dispersas. Concepto que se aproxima al de energía de interacción gravitatoria.

Esto significa que la superposición sería inestable de modo análogo a como es inestable un átomo de uranio-238. La diferencia reside en que mientras la vida media del isótopo 238 del uranio es unos 4500 millones de años, la de una partícula superpuesta sería mucho mayor ya que la E_G es muy pequeña. A la vez, cuando más grandes fuesen las moléculas, menor sería el tiempo que tardaría en decaer el estado de superposición. De tal modo que para el caso de una pequeña gota de agua en estado de superposición el tiempo de decaimiento (la decoherencia) sería de aproximadamente medio segundo.

Más adelante consideraré de nuevo este punto pero, de momento, conviene señalar que junto a esta decoherencia automática del estado de superpo-

¹¹⁶ El principio de covarianza establece que las leyes de la física deben tener la misma forma en todos los marcos de referencia.

sición, Penrose también considera real la decoherencia por el entorno debido a efectos no locales. Eso significa que para poder observar la *reducción objetiva* (OR) se requeriría mantenerla en un estado de superposición evitando que decaiga por efectos externos.

Por otro lado, según la postura de Penrose, la creación de superposiciones, con sus líneas de flujo asociadas, sería una consecuencia de la gravitación cuántica. Se trataría de un efecto análogo al que se produciría en los primeros instantes del *big bang*, donde los grados de libertad de la gravedad no estaban en equilibrio térmico junto con el resto de fuerzas, sino que se iban desatando poco a poco, causando el aumento de entropía y la expansión del universo. Ese fenómeno sería el responsable de crear alternativas cuánticas y, en la formulación actual de **U**, no se observaría porque la ecuación de Schrödinger utiliza una única variable tiempo.

Además, también el colapso de la función de onda (**R**) sería efecto de la gravedad, conforme al tiempo de decaimiento estimado t_G . La causa última remitiría a la *hipótesis de la curvatura de Weyl*. Según esta hipótesis, el tensor de curvatura de Weyl, que es nulo en el origen del universo y tiende a infinito en los agujeros negros, sería el responsable de hacer converger las líneas de flujo de las superposiciones, destruyendo las posibilidades creadas (alternativas abiertas o grados de libertad habilitados).

Llegados a este momento y desde mi comprensión de los argumentos, considero importante resaltar que Penrose da al problema de medida una solución similar a la que encuentra con la asimetría temporal para la cosmología. En su esquema, la gravedad constituye el elemento motor tanto de la asimetría temporal que se observa en la cosmología, como de la que se aprecia en la mecánica cuántica. Penrose establece cierto paralelismo, no explícito, entre la cosmología y la paradoja de la medida. A nivel cosmológico, la gravedad no termalizada desata grados de libertad y genera la expansión del universo en los instantes iniciales; y, más adelante, mediante un efecto de otro tipo, esa gravedad también produce el colapso de la materia en los agujeros negros. Mientras que a nivel cuántico, la gravedad sería la responsable de provocar inicialmente la superposición y el desdoblamiento en múltiples caminos alternativos de la partícula durante su evolución unitaria (**U**); y, más adelante, mediante un proceso de reducción objetiva (OR), esa gravedad también sería la causante del colapso de la función de onda. Siempre y cuando no se produzca antes una decoherencia por el entorno.

En resumen, para Penrose la gravedad y la asimetría del tiempo deben jugar un papel fundamental en el desarrollo de una nueva teoría que englobe

la relatividad general y la mecánica cuántica. Por eso, no sería sensato despreciar el efecto gravitatorio a pequeña escala. Si bien se trata de un efecto muy sutil comparado con el resto de fuerzas, podría ser relevante en alguno de los múltiples sistemas caóticos que se encuentran en la naturaleza, como podría ser en el problema de medida, y que son extremadamente sensibles a las condiciones iniciales. Se trata de afinar y no tanto de redondear o asumir que las fluctuaciones se cancelan.

Por otro lado, la geometría gravitatoria y la asimetría temporal irían a la par, en lo que personalmente denominaría un *espaciotiempo asimétrico*. La evolución dinámica de dicho *espaciotiempo* estaría gobernada tanto por sus propias leyes locales como por un elemento no algorítmico que lo gobernaría globalmente. Algo parecido a lo que pasa con los teselados de Penrose cuyo *isomorfismo local* está gobernado por una *decisión* no local de colocar las teselas aperiódicamente. Dicha *decisión* no algorítmica gobierna desde un nivel superior y crea las diferentes configuraciones que se parecen localmente. Veamos ahora qué características puede tener ese elemento de decisión no algorítmico.

1.4. Elementos no algorítmicos

Penrose observa que en las matemáticas hay muchas realidades no algorítmicas y solo unas pocas, como los cuasi-cristales, han encontrado una aplicación física. Su visión integradora de las matemáticas y de la física no implica que todas las matemáticas tengan un eco en la realidad física pero sí que una realidad como el pensamiento matemático, que existe en la realidad y que se comporta no algorítmicamente, debe encontrar su reflejo físico.

Como ya se vio, el punto de partida teórico de Penrose hacia ese algo no computacional son las matemáticas¹¹⁷. Desde ahí ha efectuado su crítica más ambiciosa y desde las consecuencias de esa crítica busca esas estructuras no algorítmicas en la realidad. En una primera aproximación destaca que en las

¹¹⁷ Aunque no es el único, como se resalta en Gherab-Martín 2011, 82: «Lo cierto es que, mucho antes de dedicarse a difundir el concepto de *no-computacionalidad*, en la vida de R. Penrose hizo temprano acto de presencia el concepto de *aperiodicidad*». Y lo hizo no solo mediante el descubrimiento de los teselados aperiódicos sino también mediante la influencia de la tesis sostenida por Schrödinger, en el libro *¿Qué es la vida?*, de que la vida es información y que las fibras cromosómicas deberían estar formadas por cristales aperiódicos.

matemáticas se pueden observar algunas estructuras cuya resolución no es, en principio, computable¹¹⁸. Pero lo que realmente busca es una *no computabilidad* relevante para la física y para la acción del cerebro.

Con ayuda de su argumentación matemática, Penrose concluye que la consciencia que subyace en el entendimiento en general – y en el entendimiento matemático en particular– es algo que está fuera de cualquier descripción computacional y que, en consecuencia, no puede surgir de un universo enteramente computacional¹¹⁹. Por lo tanto, el siguiente paso consiste en buscar si dentro de las leyes físicas conocidas hay espacio para una acción no computacional o si dicha acción hay que buscarla fuera de la física actual. En el caso de que esté fuera, lo siguiente sería investigar cómo es posible que esa física no computacional dé lugar a una acción relevante para el fenómeno de la consciencia en los cerebros humanos. El único espacio que Penrose encuentra para este tipo de actividad no computacional está fuera de la física conocida y se encontraría en la respuesta a la naturaleza del *problema de medida*¹²⁰.

En el *problema de medida* se observan dos partes bien definidas: una en el nivel clásico de los fenómenos a gran escala y otra en el nivel cuántico gobernado por la ecuación de Schrödinger. El comportamiento de los entes o partículas que integran ambos niveles parece perfectamente determinado, lo que implicaría que cada uno de esos niveles es computable por separado. Sin embargo, al pasar de un nivel a otro, cuando entran en contacto y se realiza una medida u observación, el colapso de la función de onda (o reducción del vector de estado) introduce una aleatoriedad fundamental en el sistema.

Dicha aleatoriedad es de carácter no-computacional pero, según Penrose, no constituye un obstáculo insalvable para la física actual. Desde un punto de vista teórico, la aleatoriedad se podría simular con una máquina de Turing de propósito universal y, desde un punto de vista práctico, se puede modelar mediante la *matriz densidad*. Sin embargo, dicha aleatoriedad sí que constitu-

¹¹⁸ Se describen varios ejemplos en Penrose 1997a, 22-23 [*Collected Works*, Vol. 6, 145-146]

¹¹⁹ Penrose 1998, 167 [*Collected Works*, Vol. 6, 265]

¹²⁰ Algunos autores identifican directamente el problema de medida con el *ubi* para las acciones libres conscientes. Rosenblum y Kuttner 2010, 17: «El papel crucial de la consciencia al que nos referimos incluye nuestra impresión de que podríamos haber *elegido* obrar de otra manera en vez de como lo hemos hecho. La percepción de que tenemos *libre albedrío*, lo que los físicos llaman el *problema de la medida*, es básica para el enigma cuántico».

ye un obstáculo para comprender realmente qué está pasando y podría estar enmascarando el proceso real. Ese oscilar de la mecánica cuántica entre la representación unitaria (**U**) y el proceso de reducción (**R**) del vector de estado, sin explicar qué sucede entre medias, hace que la teoría convencional sea básicamente insatisfactoria.

En la explicación de este fenómeno aleatorio es donde hay que buscar las raíces del comportamiento no computacional esencial. Lo que inicialmente podría parecer una genuina aleatoriedad, en realidad podría esconder un sistema caótico perfectamente organizado y extremadamente sensible. El estudio en profundidad del *proceso de medida* aclararía si se trata de una aleatoriedad real o solo aparente. En línea con esta tesis, Penrose observa que la reducción del vector de estado debe ser no local porque incluye sutiles efectos no locales debidos a la decoherencia por el entorno. Pero no se conforma con aceptar esta explicación sin más y apela a una respuesta más precisa¹²¹, como la *reducción objetiva gravitatoria* que se explicó en el apartado anterior.

Por último y como sugerencia heurística para profundizar en esta investigación, Penrose subraya dos aspectos de las matemáticas que ya están presentes en algunas teorías físicas pero que todavía no han alcanzado suficiente relevancia. El primero es el papel de los números complejos y el segundo es el papel de la simetría.

Los números complejos se introdujeron en la física como una argucia matemática para encontrar soluciones a algunos problemas y acabaron convirtiéndose en el fundamento elegante e inesperado de la mecánica cuántica. Así se observa, por ejemplo, en las amplitudes de probabilidad que rigen el principio de superposición cuántico. El hecho de que los números complejos ocupen un lugar tan privilegiado en la física, según Penrose, no es un mero formalismo sino que más bien manifiesta su íntima conexión con el diseño básico de la naturaleza. Además, como los números complejos todavía no tienen un eco en la teoría de la relatividad, constituirían una de las vías de trabajo para la nueva *gravedad cuántica*¹²².

¹²¹ Penrose 1997a, 25 [*Collected Works*, Vol. 6, 148]

¹²² Por ejemplo, los números complejos juegan un papel fundamental, es en la teleportación cuántica (transmisión de información cuántica). Penrose 2006, 804-816.

Las amplitudes cuánticas, debido a su relación de raíz cuadrada de una probabilidad, tienen asociaciones cercanas con la geometría del *espaciotiempo*. Así por ejemplo, cada estado puro del espín de una partícula másica se asocia con una dirección en el espacio. De modo que si me-

Por otro lado, las simetrías matemáticas han resultado de gran utilidad para la de la mayoría de las teorías físicas actuales. Sin embargo, según Penrose, muchas de esas simetrías estarían rotas desde el principio y serían simplemente aproximaciones a las asimetrías fundamentales. Por eso, también sugiere que se explore esta posibilidad.

Junto al papel de los números complejos y de la simetría, Penrose considera otros dos motores heurísticos como son: la belleza o elegancia de las teorías matemáticas y los «milagros», que simplifican y conectan sorprendentemente dichas teorías. Sin embargo, llega a la conclusión de que ni la belleza de las formulaciones matemáticas ni los «milagros» son en sí mismos guías seguras hacia una comprensión de la realidad. Requieren de su conexión con el mundo físico. Conexión que en el caso de los números complejos y de la simetría ya parece garantizada.

En resumen, según la postura de Penrose, la reducción del vector de estado sería una reducción objetiva: un hecho real y no imaginado o aproximado, que se debería a efectos gravitacionales. Para denominar tal efecto usa el acrónimo **OR** que indica, por un lado, que se trata de una reducción objetiva (*Objective Reduction*) y, por otro, que representa un proceso donde una superposición cuántica entre dos alternativas clásicas se convierte en una u otra (*one OR the other*). Dicho fenómeno (*OR gravitatorio*) no sería aleatorio sino debido a algún proceso no computable que se manifestaría en el pensamiento humano mediante el fenómeno de la consciencia.

2. *El fenómeno de la consciencia*

Hasta el presente apartado los temas tratados han girado en torno a la física y a las matemáticas, aunque trascendidas por especulaciones filosóficas e hipótesis científicas. A partir de ahora me centraré en un nuevo fenómeno que también aborda Penrose, el de la consciencia. Cuando en una entrevista le

dimos el espín de la partícula en esa concreta dirección, obtenemos la respuesta «sí» con una probabilidad del cien por cien. La esfera de Riemann proporciona una asociación directa entre (los radios) de amplitudes complejas de la mecánica cuántica y la geometría de direcciones en el espacio.

Esta amplia consideración se podría combinar con un enfoque más restrictivo. En concreto, como se ha visto al principio de estos apartados, las matemáticas no conmutativas podrían desempeñar un papel muy especial en el desarrollo de nuevas teorías.

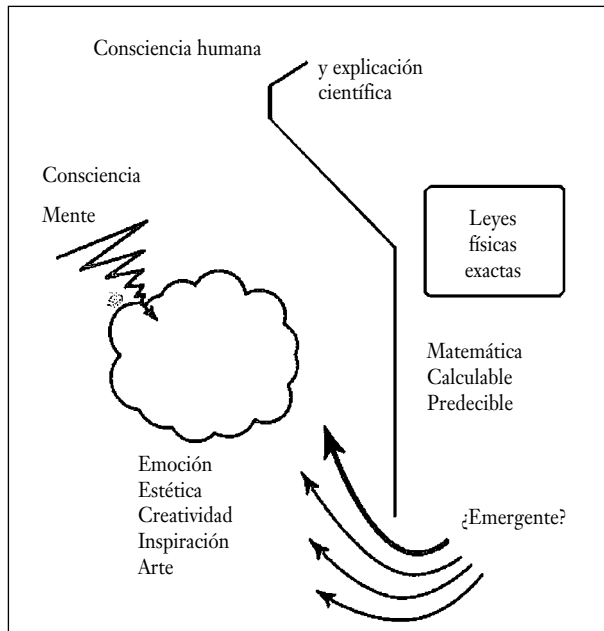
preguntaron qué es lo que le llevo a traspasar las fronteras de la física y de las matemáticas para indagar en el funcionamiento de la consciencia, su respuesta fue:

«Es un punto de vista que he formulado ya cuando estaba en la universidad, en la década de 1950. Y fundamentalmente me inspiré en el teorema de Gödel, que demuestra que las verdades matemáticas no pueden reducirse solamente a cálculos, y que para comprender las realidades matemáticas necesitamos ir más allá, salir de las meras normas de computación. Es decir, que ningún sistema consistente se puede usar para demostrarse a sí mismo. Lo que Gödel hace es mostrar cómo se pueden establecer ciertas verdades matemáticas que están fuera del alcance de las normas matemáticas. Así que la manera en que nosotros entendemos esas normas nos permite trascender más allá de las normas mismas. Lo que eso me indica es que nuestra comprensión está fuera de las normas. Este es un aspecto de la cuestión que nos lleva a la fase siguiente, nuestro cerebro y la capacidad de pensar conscientemente, que es lo que nos separa para siempre de las computadoras: la más potente y perfeccionada de ellas puede realizar cálculos de asombrosa complejidad con vertiginosa rapidez, pero jamás ‘entenderá’ lo que hace. Es el resultado de cómo operan las leyes físicas, y esas leyes físicas tienen que estar fuera de la actividad computacional. La física clásica y la física cuántica tal como la entendemos hoy podrían verse reducidas a computación. Así que tenemos que ir a buscar más allá de estas dos disciplinas. Entonces, yo me pregunto dónde está el enlace más débil en la manera en que entendemos las leyes físicas. Y considero que está en la mecánica cuántica, que incorpora dos procedimientos que son incompatibles: uno es la evolución unitaria de la ecuación de Schrödinger, y el otro es qué es lo que se hace cuando se realiza una medición. Porque de la ecuación de Schrödinger no se puede obtener una medición, y esto es una paradoja, porque un artefacto con el que podamos medir se está construyendo con ingredientes cuánticos, claro, ¿y cómo puede comportarse de manera distinta a lo que postula la ecuación de Schrödinger? Así que todo esto nos indica que hay algo más allá. Y estoy de acuerdo con Einstein, con el mismo Schrödinger, e incluso con Paul Dirac: todos ellos afirmaban que la mecánica cuántica de nuestros días está incompleta. Y necesitamos por tanto completarla. Así que lo que yo especulo es que hace falta fundar las bases para la revolución teórica que permita a la física incluir en su campo el fenómeno de la consciencia».

Se ha querido recoger la cita íntegra, aunque sea larga, porque resume bien las explicaciones que se han dado hasta este momento. Ahora queda pro-

fundizar en el estudio del tipo de actividad física que según Penrose puede darse en el cerebro y que esté relacionada con la consciencia.

La consciencia, según el esquema de los tres mundos de Penrose, se encuentra entre dos misterios: uno que la relaciona con el mundo físico y otro que la asocia con su capacidad de conocer la verdad matemática. Eso implica, por un lado, que las matemáticas tienen cierto apoyo en la actividad mental consciente y, por otro, que la consciencia se puede tratar en términos del mundo físico. Sin embargo, no se reduce a esto, como las matemáticas tampoco se reducen a lo que el matemático puede realizar. En el siguiente esquema se pueden observar algunas características de la relación entre el mundo físico y el mental.



El mundo físico, estaría gobernado por leyes físicas y matemáticas precisas, aunque todavía no se hayan desvelado completamente. Se trata por tanto de un mundo predecible y calculable, que quizá sea determinista y también computable. Por su parte, el mundo mental alberga la consciencia y es el lugar donde algunos términos como alma, espíritu, arte o religión adquieren sentido. Del mundo físico *emergerían* algunas cualidades mentales como la emoción, la estética o la inspiración. Esto significa para algunos autores que

dichas cualidades mentales pueden emerger de una actividad computacional. Penrose, que podría alinearse con la primera idea de *emergencia*, rechaza de plano que dicha *emergencia* sea fruto de una actividad computacional.

La consciencia tendría dos ámbitos de manifestaciones: uno pasivo que se asocia con el conocimiento, en sentido amplio, y otro activo que se relaciona con la libertad y la voluntad. Sin embargo, Penrose no aclara los términos ni el modo en que se relacionan, en parte porque no se considera capaz de precisiones filosóficas y en parte porque se conforma con las acepciones propias del sentido común. Para su argumentación le basta considerar que la comprensión humana requiere de la consciencia.

En cuanto a las relaciones entre consciencia y computabilidad, Penrose adopta una postura *C fuerte*, según la cual, una adecuada acción física de los cerebros es capaz de evocar la consciencia, pero esta acción no puede ser simulada porque reclama la presencia de un elemento no computacional.

«Esta es mi versión de la *C fuerte*: buscamos la no computabilidad en la física que forma un puente entre los niveles cuántico y clásico, lo cual es pedir bastante. Estoy diciendo que no solo necesitamos una nueva física, sino que también necesitamos una nueva física que sea relevante para la actividad del cerebro».

Se distancia así de la postura *C débil* para la cual el motivo de la imposibilidad de dicha simulación se podría deber a la *no computabilidad* de los fenómenos de aleatoriedad pura que se observan en el colapso de la función de onda (**R**). Sin embargo, para Penrose, integrar una aleatoriedad pura (imposible de simular) en un sistema de computación no añadiría una diferencia significativa respecto a la pseudo-aleatoriedad de las simulaciones actuales. Este enfoque sería insuficiente y se requiere de algo diferente que debe ser no computable.

El argumento central de Penrose sobre el porqué de la *no computabilidad* ya se ha explicado, con sus pros y sus contras. También se ha explicado que dicha *no computabilidad* se podría encontrar en el fenómeno de *OR gravitatoria*¹²³. Queda por tanto aclarar cómo se relaciona este fenómeno con la consciencia y señalar dónde se podrían encontrar sus efectos.

¹²³ En Penrose 1999a, 97-99 se señalan dos hipótesis sobre la gravedad cuántica en los que se incluye un elemento no computable. La primera es de Geroch-Hartle y la segunda de David Deutsch.

A partir de este punto los argumentos de Penrose se tornan altamente especulativos y a mi parecer poco convincentes. De entre todas sus hipótesis, esta es la más criticada y la menos respaldada. Como se verá es una argumentación llena de saltos explicativos y con muchos huecos por rellenar. Sin embargo, tienen un valor heurístico intrínseco y el mérito de intentar desarrollar las teorías hasta que tengan un contacto con la realidad. Cabe resaltar también que en sus dos últimos libros Penrose no ha tratado de este tema y entre su primer y su segundo libro ha cambiado de tesis. Por tanto aquí me centraré en el contenido de *Shadows of the mind* y en el de alguno de sus artículos.

Para una comprensión más profunda del esquema y como planteamiento de fondo conviene recordar el teselado aperiódico de Penrose. Dicho teselado consiste en un esquema muy simple con dos tipos de teselas cuya evolución es completamente determinista y no computable. Robert Berger demostró que la evolución de este esquema no puede ser simulada por ningún ordenador, porque no hay un algoritmo capaz de decidir si un conjunto de teselas cubrirá una superficie. Por tanto, dicho esquema está gobernado por reglas no locales que escapan a la comprensión de un ordenador ideal, de modo análogo a lo que sucede con la *escalera sin fin* de Penrose. En dicha escalera los tramos están bien contruidos localmente y, sin embargo, se requiere una visión de conjunto para percibir que la escalera no es real.

La *no computabilidad* de Penrose se situaría en un nivel superior e influiría sobre el nivel inferior, de modo análogo a como la *aperiodicidad*, que es una regla superior, influye sobre el nivel inferior sin que se pueda detectar localmente. A nivel local todo podría parecer determinado y computable; solo cuando se observa desde un nivel superior aparece la *no computabilidad*. Así también la comprensión se daría en un nivel superior al de la computabilidad y ese nivel es el de la consciencia. Es decir, consciencia-comprensión, aperiodicidad y *no computabilidad* pertenecen a niveles superiores, en relación con la computabilidad.

Estas consideraciones de Penrose sobre la consciencia se integran con argumentos sobre la reflexividad y la incapacidad de crear bucles cerrados en cuanto alcanzan un mínimo de complejidad. Así por ejemplo, el argumento de Gödel aplicado a una parte bien definida y suficientemente compleja de las matemáticas¹²⁴, implica buscar su justificación en unas matemáticas par-

¹²⁴ Como puede ser la aritmética de Peano.

cialmente externas (o superiores) a ese esquema. Dicho esquema se vuelve recursivo sobre cada conjunto de matemáticas e implica en última instancia y por consideraciones filosóficas que las matemáticas no tienen una justificación interna última. Este tipo de argumentos le llevan a considerar que podríamos tener «tipos de no-computabilidad de orden superior» involucrados, por ejemplo, en la forma en que evoluciona el Universo o en nuestra libertad.

Sostiene por tanto la existencia de varios niveles en los que cada uno de esos niveles podría tener las características de un «determinismo no computable» con relación al nivel superior. Desde mi punto de vista, esto implica que cada uno de esos niveles es un sistema abierto, que no se puede encerrar o justificar internamente. Ahora bien, hay una pregunta que quedaría sin resolver: ¿el esquema global es cerrado o abierto?

En el esquema de los tres mundos de Penrose hay un intento de cerrar el conjunto de la realidad. Cada mundo *justifica* la existencia del anterior, de modo que permanecen abiertos, aunque el conjunto sea cerrado. Además, hay un intento de fundamentar la realidad última en la existencia real del mundo platónico de las matemáticas. Sin embargo, la intensidad con que Penrose defiende cada una de estas dos posturas es diferente. Mientras que en la cuestión sobre el fundamento último no admite ninguna duda, en la cuestión sobre la apertura del esquema global, sí que admite otras alternativas. Por lo tanto, se podría sostener que la realidad de Penrose se fundamenta en las matemáticas aunque no tiene por qué ser completamente cerrada.

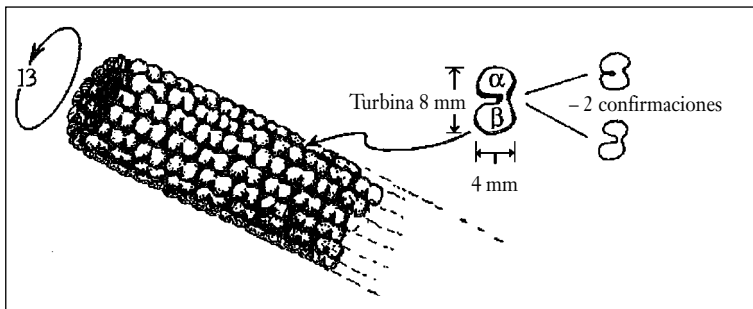
Por mi parte, como argumenté, me parece que el mundo platónico de las matemáticas no es el fundamento último de la realidad y que las matemáticas, la física y la consciencia son tres sistemas abiertos, tanto considerados en sí mismos como tomados en conjunto. A su vez esos sistemas se pueden subdividir en subsistemas abiertos. Dicho esto centrémonos en lo que sucede en los cerebros humanos para ver dónde tienen cabida las consideraciones sobre la no computabilidad.

3. *Hacia la base física de la consciencia*

El fenómeno **OR** es el que, bajo su aparente aleatoriedad pura, podría esconder la *no computabilidad* buscada. Sin embargo, las escalas de tiempo previstas para dicho fenómeno son muy superiores a las escalas de tiempo

que se manejan en los procesos cerebrales responsables de la consciencia¹²⁵. Luego, tomado aisladamente, dicho fenómeno no sería suficiente para ejercer una acción relevante en el cerebro. Hay que buscar dónde o bajo qué fenómeno se podría dar una influencia relevante de **OR** para la actividad cerebral.

La primera propuesta de Penrose, explicada en *Emperor's New Mind*, giraba en torno al concepto de gravitón¹²⁶. Sin embargo, nuestro autor cambió de opinión ante los descubrimientos de Stuart Hameroff sobre la actividad del citoesqueleto¹²⁷ de las neuronas. En dicho citoesqueleto se hallan unos microtúbulos hechos de proteínas llamadas tubulinas, que pueden adoptar dos conformaciones, abierta o cerrada.



¹²⁵ En Penrose 1997a, 25 [*Collected Works*, Vol. 6, 148] se explica cuál sería el valor de tiempo implicado en la OR:

«The claim (or suggestion) is that you consider the energy E that it would take to displace one instance of this mass in the gravitational field of the other instance of the mass. The reciprocal in Planck units, of that energy gives you a time $(\hbar/E=t)$, and that time t is the time it takes for the state to jump from the superposed state to one or the other».

Para los protones o neutrones este proceso supondría cientos de millones de años, pero para diminutas gotas de agua, cuya masa (y también su energía asociada a la superposición) es notablemente mayor, el proceso OR tendría lugar en escalas de tiempo adecuadas a las de los procesos biológicos.

¹²⁶ Cfr. nota 53 para ver la diferencia entre la primera y segunda propuesta de Penrose respecto a la reducción de estado. Un gravitón, según algunas teorías de gravedad cuántica, sería la partícula elemental de la interacción gravitatoria, de modo similar a como el fotón lo es de la fuerza electromagnética, los bosones W y Z de la interacción débil y el gluón de la interacción fuerte. En la actualidad los gravitones no se han descubierto.

¹²⁷ El citoesqueleto es un entramado tridimensional de proteínas que provee soporte interno en la célula, organiza las estructuras internas de la misma e interviene en los fenómenos de transporte, tráfico y división celular.

Según Stuart Hameroff los microtúbulos se podrían considerar como *autómatas celulares* y las tubulinas en sus dos conformaciones hacer el papel de los «ceros» y «unos» de los transistores electrónicos. Por tanto a través de los microtúbulos se transmitiría una enorme cantidad de información digitalizada y la complejidad de los procesos que tienen lugar en las neuronas se elevaría considerablemente¹²⁸.

El primer aspecto de los microtúbulos que le llamó la atención a Penrose es que son tubos lo suficientemente pequeños para que puedan albergar efectos cuánticos. Su estructura de tubo implica que su actividad interior se podría aislar de la actividad aleatoria del entorno. De esta manera, dentro de los tubos se podrían conservar masas en estado de superposición cuántica, que al encontrarse aisladas, no decaerían por el entorno. A la vez, la estructura de tubo podría permitir la existencia de algún tipo de coherencia cuántica que podría estar involucrada en las señales que se transmiten a lo largo de esos microtúbulos. Dicha coherencia se podría deber a la actividad electromagnética de moléculas de agua ordenadas tanto en el interior como en el exterior de la tubulina, o bien a fenómenos similares a la superconductividad¹²⁹.

En una primera fase las moléculas de agua del interior de la tubulina estarían ordenadas coherentemente entre ellas y entrelazadas cuánticamente. Esta superposición se mantendría hasta que, en una segunda fase en la que actuaría la consciencia, las tubulinas se verían implicadas en el estado de coherencia cuántica superpuesta. Al incorporarse las tubulinas a la superposición, la masa desplazada –y su energía E_G – aumentaría hasta ser lo suficientemente grande como para provocar la OR en un tiempo breve t_G . Durante ese pe-

¹²⁸ Por cuestiones de entrelazamiento cuántico, la cantidad de bits de información que contendrían 500 tubulinas entrelazadas sería de 2^{500} . Lo que significa que para duplicar la cantidad de información bastaría con entrelazar una tubulina más.

¹²⁹ La analogía es sugerente, porque durante muchos años se pensó que la superconductividad era un fenómeno sólo posible a temperaturas cercanas al cero absoluto. Sin embargo, recientemente se han descubierto fenómenos de superconductividad a alta temperatura (bastante por debajo de la temperatura corporal, pero que podrían alcanzarla). Dichos fenómenos se observaron inicialmente a bajas temperaturas en los *condensados Bose-Einstein* (BEC) y muestran un estado de coherencia cuántica macroscópico.

En este caso, al bajar la temperatura se aumenta la longitud de onda de las partículas hasta que sus ondas asociadas se acoplan entre sí dando lugar a una gigante onda de materia. Las ondas de materia superpuestas, están en fase, y cada fotón –cuya longitud de onda ha aumentado– se extiende sobre todo el BEC. Este fenómeno normalmente asociado con los bosones, también puede darse con pares de fermiones (quarks y leptones).

riodo de tiempo las moléculas de agua exteriores a la tubulina mantendrían el aislamiento hasta que, por efecto de la OR, se conformase un patrón de tubulinas dando lugar al fenómeno de consciencia. Esa selección tendría lugar por efectos locales gravitatorios y desde un nivel superior no local por efectos conscientes no computacionales. Posteriormente, el nuevo patrón de tubulinas, aportaría su información a las neuronas, influyendo en estas y en sus conexiones sinápticas.

Este esquema requeriría de algún tipo de coherencia global a lo largo de extensas regiones del cerebro para que la *no computabilidad* tenga la necesaria importancia en la acción consciente del cerebro. Dicha coherencia global, susceptible de una *OR orquestada*, sería posible por el entrelazamiento cuántico de distintas partículas en distintas partes del cerebro.

En resumen, la propuesta de Hameroff y Penrose requiere en primer lugar un estado de *coherencia cuántica a gran escala* en infinidad de microtúbulos de neuronas globalmente distribuidas a lo largo del cerebro. Además, los microtúbulos, que podrían albergar moléculas de agua ordenadas, proporcionarían algún tipo de escudo contra el ruido cuántico del ambiente para evitar la decoherencia por el entorno y mantener la superposición mediante un fenómeno similar al de la superconductividad. Con ayuda de este substrato y en una segunda fase se producirían los sucesos conscientes. En esta fase las tubulinas se verían involucradas en el estado de coherencia cuántica y se produciría una superposición cuántica de diferentes estructuras de tubulinas. Este proceso sería una *OR orquestada* de modo que resultase escogido un patrón de estructuras de tubulinas de un modo *no computacional*. Cada una de esas tubulinas sería portadora de información digitalizada. Las escalas de tiempo estimadas por Penrose supondrían la acción coordinada de entre 10 y 100 mil neuronas para un solo fenómeno de consciencia.

4. *Dificultades de la propuesta*

Como ya se avanzó, el grado de especulación de la propuesta de Penrose sobre cómo puede actuar la consciencia en el cerebro humano es muy elevado. Hasta tal punto que resulta difícil escoger el tipo de argumento en contra. Además, incluye tal cantidad de presupuestos filosóficos e interpretaciones científicas que no se pueden abordar todos ellos. Así que me centraré solo en algunas dificultades que él mismo reconoce y en algunos presupuestos filosóficos.

Un primer aspecto es que el fenómeno **OR** estaría normalmente enmascarado por la aleatoriedad cuántica y sería prácticamente inobservable entre la maraña de fenómenos de decoherencia por el entorno. Solo un experimento *ad hoc* como el experimento FELIX podría dar indicios de una **OR** gravitatoria, aunque en última instancia también podrían ser provocados por el entorno. Todo esto apunta a que la interpretación más plausible a día de hoy es la de *decoherencia por el entorno*, que es compatible con la defensa de un procedimiento objetivo de reducción del vector de estado.

Por otro lado, el fenómeno de consciencia sería inobservable, porque la *OR orquestada* se derivaría del entrelazamiento cuántico y de los efectos de la gravedad en un sistema de gran cantidad de partículas, moléculas y estructuras. Solo en un sistema cuántico coherente y aislado del entorno, donde el vector de estado se aplicase a escala macroscópica a un enorme número de partículas actuando en concierto, se podría observar la consciencia. Pero para el punto de vista convencional todo esto es altamente improbable¹³⁰ y desde un punto de vista científico reclama algún tipo de experimento que resultaría imposible de realizar.

En tercer lugar el fenómeno de la consciencia requeriría de la presencia de ciertas estructuras biológicas que no son específicas de los cerebros humanos y que se pueden observar en otras especies. Por eso, o se concluye que también otros seres son capaces de tener algunos grados inferiores de consciencia, como sostiene Penrose, o se afirma que la consciencia deriva en última instancia de la complejidad del cerebro, postura que el mismo Penrose rechaza en su crítica de la posibilidad de una inteligencia artificial.

Además, después de la explicación de Penrose, la conclusión personal a la que llego y que no aparece explicitada por Penrose, es que la acción de la gravedad –es decir, la geometría del *espaciotiempo*– en unas estructuras del cerebro, sería la «glándula pineal» del fenómeno de la consciencia. Postura desde la que es fácil deslizarse hacia el mentalismo rechazado por nuestro autor.

En definitiva, Penrose elabora una teoría altamente especulativa en la que se integran una infinidad de elementos científicos sobre los que resulta necesario profundizar. Es cierto que quizá por esas vías estén las soluciones a las dificultades explicativas de tantos fenómenos reales que no conocemos

¹³⁰ Penrose 1998, 172-173 [*Collected Works*, Vol. 6, 270-271]

pero, como el mismo Penrose reconocería, necesitamos profundizar más antes de llegar a una teoría convincente.

Por último considero que en la postura de Penrose hay un par de puntos que conviene resaltar y que convendría explorar con más detenimiento. El primero es el aspecto recursivo de muchos fenómenos físicos y su posible relación con la consciencia. Y el segundo es la conclusión de que la consciencia se sitúa en un nivel superior al de la física aunque pudiese *emerger* de esta. Penrose, no intenta incluir la consciencia en el mismo nivel de las causas físicas o matemáticas, ni tampoco separarla completamente de estas. Deja por tanto el camino abierto a una consciencia que interacciona con los niveles de las ciencias y de las matemáticas, aunque el modo en que se da esa relación siga siendo el mayor de los *misterios*.

CONCLUSIONES

En este trabajo se ha presentado la postura de Penrose con relación al mundo físico, a las matemáticas y a la consciencia. Se ha pretendido ser fiel al contenido y al espíritu del autor, evitando caer en un primer nivel de crítica más superficial para llegar a otro más profundo. Por eso, frente al objetivo central de la valoración filosófica, se han supeditado algunas apreciaciones físicas, matemáticas o biológicas.

En el marco general de la obra de Penrose se aprecian dos motivaciones. La primera constituye la defensa de una postura filosófica según la cual la *no computabilidad* sería un rasgo de nuestro pensamiento consciente; y la segunda expresa el deseo pedagógico de escribir sobre temas científicos para que llegue a un público más amplio. Lo que más le motivó para escribir su primera obra fue:

«(...) ver en el programa *Horizon* de la BBC a Marvin Minsky y otros haciendo unas afirmaciones de lo más escandalosas. Sentí que había un punto de vista, el mío, que nunca había visto expresado en ningún sitio y que era necesario sacar a la luz (...) Supongo que lo que estaba haciendo en ese libro [*The emperor's new mind*] era filosofía, pero alguien se quejó de que yo me remitía a un único filósofo –lo que seguramente es cierto–. Esto es porque las cuestiones que interesan a los filósofos tienden a ser bastante diferentes de las que interesan a los científicos; los filósofos tienden a ocuparse de sus propias controversias internas».

Encontramos por tanto en Penrose un afán por transmitir su punto de vista, junto con una necesidad de hacer filosofía de la naturaleza para contrarrestar las afirmaciones *escandalosas* de otros científicos.

Con estas motivaciones convive un modo de realizar la tarea en el que se manifiesta tanto la preferencia por la aproximación científica como el deseo de alcanzar una visión de conjunto¹³¹. La primera aproximación, se quiera o no, es reductiva ya que no escapa a la dimensión legaliforme de la ciencia. Esta busca leyes que expliquen el comportamiento de una realidad que, sin embargo, las trasciende. Mientras que el segundo enfoque puede desembocar en la unificación de aspectos demasiado distantes, como cuando Penrose amalgama la solución de varios puntos inexplicados por la ciencia actual en torno a la resolución del *problema de medida*.

He aquí por tanto dos puntos matizables de su enfoque. El primero es la pretensión implícita y explícita de que la ciencia no tenga límite explicativo para todo lo que tenga substrato material: desde el colapso de la función de onda, hasta los fenómenos de la consciencia o de la libertad. Y el segundo es la aparente precipitación que le lleva a interconectar muchos puntos inexplicados en propuestas excesivamente pretenciosas, como la del fenómeno de consciencia.

Este segundo punto requeriría algunos matices ya que se tratan de propuestas heurísticas que podrían en algún caso ser verdaderas. Mientras que el primer punto es una asunción filosófica fuertemente sostenida por Penrose tanto a nivel teórico como práctico. Aun así, su obra no carece de interés filosófico, sino que requiere de una lectura atenta que permita discernir la paja del grano.

En los ensayos científicos de Penrose resalta un sano inconformismo tanto con las teorías físicas reinantes como con la pretensión de atenerse a lo que funciona sin preguntarse por el porqué último de las cosas. Nuestro autor no se conforma con el poder predictivo de las teorías – con un funcionalismo al que le bastan unos buenos resultados – sino que aspira al poder explicativo científico y filosófico. Además, confía en la capacidad humana, que subyace en el método científico, de conocer la realidad.

En los aspectos vitales y filosóficos que marcan el modo de pensar de Penrose se aprecia la influencia de autores como Kurt Gödel en lo matemático,

¹³¹ Como afirma Lee Smolin en Penrose 1996, 224-241: «Todas las reflexiones de Roger están conectadas. Las ideas técnicas desarrolladas en su teoría de los operadores de torsión, su pensamiento filosófico, sus ideas acerca de la mecánica cuántica, sus ideas acerca del cerebro y la mente, todas están conectadas».

Albert Einstein y Denis Sciama en lo físico, o Karl Popper en la cosmovisión científica. Junto a estas influencias conviven los intereses, compartidos desde la infancia con su padre, por la geometría, las matemáticas o la consciencia.

En su metodología resalta el papel fundamental tanto del *instinto* científico como del sentido común. Ambos catalizan los esfuerzos para acertar con las vías que llevan a las teorías válidas y en última instancia verdaderas. Además, manifiesta una pasión por conocer la realidad bajo un enfoque simplificador que busca la belleza intrínseca de lo real. Aun así no basta, por ejemplo, con que las matemáticas empleadas sean bellas o simples en sí, sino que tienen que enlazar con la belleza o simplicidad de otros niveles, para dar a luz a la verdad.

El único de los trascendentales platónicos que Penrose no pone explícitamente en contacto con los demás es el bien. Sin embargo, la misma actividad científica reclama una bondad en su ejecución y apunta a la bondad del conocimiento científico. Dicha bondad se observa también en la honestidad y claridad con que el científico debe presentar sus resultados, aunque necesariamente se encuentren en el contexto de una interpretación. Así sucede por ejemplo en los libros de Penrose, donde a la vez que se resaltan los aspectos más relevantes para mostrar su visión de conjunto, también se reconocen algunos puntos dudosos.

Esta interpretación de los datos científicos es más relevante de lo que pudiera parecer a primera vista, ya que con los mismos datos se puede llegar a conclusiones radicalmente distintas. Así por ejemplo, algunos autores sugieren que la mecánica cuántica, tal y como la conocemos, pone en duda tres intuiciones de sentido común: dar por sentado que dos objetos no pueden estar a la vez en el mismo sitio; que lo que sucede aquí no puede estar afectando *simultáneamente* en algún lugar muy lejano; y que hay un mundo real *abí fuera*, con independencia de que lo contemplemos o no. Según la mecánica cuántica parecería que la observación *crea* la realidad física observada.

Estas conclusiones, que se deducen en el primer caso de la superposición cuántica, en el segundo de los efectos EPR y en el tercero de la influencia del observador sobre lo observado, son para muchos científicos prácticamente canónicas; y, sin embargo, distan bastante de las sostenidas por Penrose u otros autores. Con frecuencia lo que se presenta como conclusiones de las teorías científicas, responde simplemente a marcos filosóficos previos.

Penrose hace una lectura distinta de cada uno de esos fenómenos y considera que el problema está en la insuficiencia del conocimiento físico de dichas realidades. Para nuestro autor, si algo se aleja mucho del sentido común

y científico conviene replantearse su interpretación actual. Aunque algunos fenómenos se conozcan con mucha precisión técnica, si no se conocen las causas que lo explican entonces uno no puede dejarse guiar solo por los datos experimentales y las teorías consolidadas. La tarea científica requiere por tanto de cierto *instinto* científico, del sentido común, de la sencillez de los planteamientos, de la objetividad en las conclusiones, de la armonía con otros ámbitos del saber, de la humildad para reconocer que se puede estar en el error y de la existencia de una realidad *ahí fuera*. Si no existiese un mundo real no se podría hacer ciencia y el hecho de que se pueda hacer ciencia es una manifestación de que existe una realidad que se puede conocer objetivamente.

En resumen, los límites del método científico determinan los límites de la imagen del mundo que la ciencia puede aportar. Sin embargo, la imagen que los científicos tienen del mundo trasciende los límites de la ciencia actual. Hay preguntas que no se pueden responder dentro de las ciencias y dan lugar a interpretaciones. De este modo la interpretación de la ciencia, o el conjunto de interpretaciones, acaba conformando el panorama del mundo y permea la cultura de una determinada época, alcanzando el trasfondo filosófico. La ciencia, por ende, se constituye en un valor específicamente humano y, por eso, resulta legítimo aventurarse con nuevas interpretaciones que catalicen la investigación. Sin embargo, para que ese proceso sea eficaz y fructífero se requiere de una reflexión que acote las interpretaciones.

Con este propósito se ha expuesto y analizado en este trabajo el pensamiento de Roger Penrose, cuyo nervio central está constituido por el *nuevo argumento de Penrose*. Aquel que, partiendo de las matemáticas, le lleva a afirmar que en el substrato físico sobre el que tiene lugar la consciencia –y el pensamiento matemático– debe haber un elemento no computable.

Al analizar dicho argumento se llegó a la conclusión de que puede ser matemáticamente correcto y filosóficamente válido, aunque no concluya porque no es demostrable la premisa matemática de consistencia. En teoría no habría motivos suficientes para dudar de dicha consistencia y, sin embargo, resulta imposible probarla desde dentro de las matemáticas computables. Demostrar la consistencia del algoritmo remite a unas matemáticas más amplias, cuya consistencia se debe probar. A su vez, estas remiten a otras, hasta que se llega al conjunto de las matemáticas, cuya consistencia y fundamentación última resulta intrínsecamente indemostrable. Aunque no hay motivos de peso para dudar de la consistencia de las matemáticas, sin embargo, tampoco se pueden demostrar matemáticamente. Por tanto, a título personal, concluiría que las

matemáticas no constituyen un sistema perfectamente cerrado, sino abierto, en el que la carga de la prueba estaría a favor de la consistencia.

Todo esto me lleva a compartir las conclusiones filosóficas a las que llega Penrose de que las máquinas nunca podrán llegar a ser conscientes como lo es el hombre. Su argumento es sugerente y debe ser tenido en cuenta como argumento filosófico, pero su pretensión de hacer un argumento inatacable es errónea porque pivota sobre la premisa indemostrable de consistencia matemática. En última instancia, dicha consistencia es asumible como verdadera pero no demostrable, lo que lleva a muchos matemáticos realistas, como Penrose o Gödel, a buscar un fundamento platónico para las matemáticas. A mi parecer se trata de una conclusión filosófica equivocada, en la que se da un salto ilícito en un esfuerzo por fundamentar el realismo matemático.

Como se ha argumentado en este trabajo, bastaría con confirmar que las matemáticas requieren de un mínimo de realismo por su relación con los modos de conocer del ser humano y de la realidad. Sería algo análogo a lo que sucede con la física u otras ciencias, cuya actividad requiere de algún tipo de realidad, aunque carece de un lecho de roca firme – los consabidos hechos experimentales– que la fundamenten. En última instancia las matemáticas se justificarían tanto internamente, mediante nuevos constructos y demostraciones que la estructuran, como externamente, mediante la realidad y la actividad de conocimiento humano.

No sería necesario recurrir ni al platonismo matemático ni a su método de conocimiento por intuición directa, ya que ambos conllevan más dificultades ontológicas y epistemológicas que las que resuelven. El fundamento externo de las matemáticas se adquiriría en su relación con otros saberes, con la realidad conocida y con el sujeto cognoscente. De este modo se redimensionaría el *misterio* que une el mundo platónico de las matemáticas con el mundo físico y a la vez se rompería el esquema de los tres mundos de Penrose.

Dicho esquema, que expresaba los prejuicios de Penrose, resulta filosóficamente inviable porque muestra, sin justificar, un entrelazamiento de tres mundos en una especie de *causa sui* retardada¹³². En este sistema, aunque los

¹³² En Penrose 1999a, 111, se pone en paralelo el triángulo de Penrose con su esquema de los tres mundos para afirmar que no es casualidad que se parezcan tanto. Personalmente tomaría pie de este paralelismo para afirmar que, al igual que el triángulo de Penrose no puede ser real aunque localmente lo parezca, tampoco su esquema global de los tres mundos es real, aunque los *misterios* locales sigan existiendo.

misterios trascienden cada uno de los mundos, quedan encerrados en el conjunto global. Parece como si, en su afán por fundamentar la existencia objetiva de la realidad, Penrose llegase a un esquema cerrado que no respeta el *misterio* global y del que no sabe cómo salir aunque reconozca la incoherencia lógica. Se podría decir que se aproxima a la realidad con una mentalidad objetivo-deductiva que desea controlar, mientras que quizá se requiere una mentalidad contemplativa que desea descubrir los *misterios*.

A mi parecer ese enfoque objetivo-deductivo no solo se observa en sus argumentaciones matemáticas en relación al *nuevo argumento*, como ya se ha visto, sino también en sus argumentaciones físicas. Penrose acepta un marco matemático y físico, dando por hecho la consistencia de dicho marco y por tanto de las argumentaciones realizadas dentro de él. Acepta implícitamente el marco legaliforme de la física y de las matemáticas, pero el problema reside en que, como su esquema, se queda encerrado en esos presupuestos. Da por hecho que las respuestas a los grandes problemas se pueden encontrar en ese marco. Así, presupone por ejemplo que las matemáticas y las ciencias pueden dar razón del fenómeno de consciencia. Lo que, en última instancia, constituye un *acto de fe* en el poder de las matemáticas y en la física para alcanzar la verdad por sí mismas, porque ninguna de ellas tendría un fundamento intrínseco propio.

En resumen, el esquema de los tres mundos de Penrose resulta poco convincente por las restricciones de causalidad que impone entre los mundos. Con su deseo de justificar el conjunto, no solo limita la riqueza del *misterio*, sino que lo hace más difícil de entender. La realidad física parece más rica que las matemáticas que subyacen en ella, de igual modo que la realidad mental parece mucho más rica que la física o las matemáticas que puedan subyacer en ella. Con esto no pretendo afirmar que esos marcos no sean válidos sino que en el esquema de Penrose son demasiado estrechos.

Como alternativa se podría presentar un esquema donde los tres mundos se entrelazasen entre sí mediante *misterios* difusos y bidireccionales. Las relaciones entre el mundo mental y las matemáticas, por ejemplo, serían recíprocas: las matemáticas en parte se descubrirían y en parte se construirían; al igual que la relación entre el mundo mental y el mundo físico: gracias al mundo físico podrían existir seres humanos conscientes que conocen y actúan sobre el mundo físico, creando nuevas estructuras o realidades que antes no existían. Por otro lado, estos *misterios* bidireccionales tendrían una fuerte componente holística que desdibujaría los elementos necesarios para justificar todas las

relaciones. Además, cada uno de esos mundos y el conjunto global gozaría de una permeabilidad y apertura intrínsecas; lo que a su vez resultaría más acorde y respetuoso con las tareas del científico, del matemático y del filósofo. De este modo emergería con más claridad un nuevo *misterio* que se encontraba oculto por el esquema de Penrose y que a lo largo de este trabajo se ha ido desvelando como conclusión filosófica derivada de las matemáticas, de la física y de la consciencia: la apertura intrínseca del conjunto global, no solo de cada uno de los mundos de Penrose.

Por tanto, parecería necesario afirmar que el conjunto de las matemáticas son consistentes; de igual modo parece necesario sostener –desde la física– que existe una realidad cognoscible *abí afuera* y –desde la persona– que existen sujetos conscientes que conocen y reflexionan sobre las matemáticas, sobre la realidad y sobre sí mismos. No hay motivos suficientemente fundados para refutar ninguno de estos supuestos, así que la carga de la prueba estaría en demostrar las hipótesis contrarias. Detrás de esas afirmaciones subyacen no solo la consistencia matemática sino lo que podríamos denominar una *consistencia física* de la realidad, así como una *consistencia personal* del sujeto que hace matemáticas y física. Consistencias que a su vez se fundamentan recíprocamente.

Una vez asumido esto, tiene sentido llegar a la conclusión que llega Penrose, según la cual en el pensamiento matemático debe haber algo no algorítmico o, lo que es lo mismo para él, no computacional. Esa conclusión particular, generalizada para el pensamiento y extendida a la realidad material, se encuentra en las matemáticas, en la física y en la consciencia. Si, como concluye Penrose, ni siquiera las matemáticas se pueden reducir a algoritmos, cuánto menos se podrán reducir a algoritmos la realidad física o la consciencia. El nuevo reto que se plantea ahora es buscar esa realidad no algorítmica.

Sin embargo, desde mi punto de vista hay un error de planteamiento en esa búsqueda. El punto de partida implícito de Penrose para su razonamiento es suponer que toda la realidad es computable; y cuando descubre que no lo es, busca lo no computable. Pero la conclusión no debería ser, solo, que debe haber algo no computable en la realidad, lo cual es cierto, sino que la misma premisa de computabilidad es inaceptable. La cuestión es ¿por qué Penrose no se da cuenta de esto?

A mi parecer, intuye por su sentido común que no todo es computable pero, a la hora de afrontar el dilema, desea realizar una contra-argumentación desde las matemáticas similar a la que Gödel hizo con el formalismo de Hilbert. Por tanto imbuido en un reductivismo matemático se aproxima al pro-

blema de argumentar contra el mecanicismo computacional. Al final se llega a lo que hemos afirmado sobre la consistencia y la imposibilidad de cerrar completamente el argumento, aunque este sea verdadero. Siempre existe un espacio desde el que desarrollar una contra-argumentación a favor de la computabilidad, lo que no significa que en la realidad sea posible.

Dicho esto, me parece que el punto de partida correcto sería afirmar la existencia de una realidad en sí a la que llegamos por inducción de nuestra experiencia. Posteriormente esa realidad, se estudia, se observa o se analiza desde un punto de vista científico, matemático o computacional. Pero al observarla así se deja fuera todo lo que es no-científico, no-matemático o no-computacional. Luego cabría afirmar que siempre habrá mucha realidad con esas «no-características» porque la realidad trasciende la aproximación reductiva para mostrarse abierta. *Ante una aproximación reductiva, la realidad acaba mostrando su apertura.*

Después de esta digresión volvamos sobre el hilo conductor de las conclusiones de Penrose. Una vez descubierta la dimensión no algorítmica de las matemáticas, de la física y de la consciencia, la nueva pregunta que se suscita es: ¿dónde se encuentra ese algo no algorítmico? Entre los *lugares*, nuestro autor señala la intuición matemática, la paradoja de la medida o los juicios del pensamiento. De esas tres localizaciones, todas ellas interesantes, Penrose se centra especialmente en la paradoja de la medida, pero sin aclarar en qué consiste ese algo no algorítmico. Aun así, de lo que hemos podido descifrar a lo largo de este trabajo, se puede concluir que *lo no algorítmico* es aquello que un ordenador no puede realizar porque *requiere una visión de conjunto que trasciende los aspectos locales*. Lo algorítmico o computacional carecería de una radical apertura a la novedad, a lo que no está previsto: solo admitiría las novedades consideradas en su programación, sin poder trascender el aspecto local de aquello para lo que ha sido programado.

Localmente se aprecian cosas que están determinadas y que son computables y otras que parecen indeterminadas porque en ellas tienen lugar unas *decisiones* o *novedades* que no son computables. Sin embargo, la *decisión* o *novedad* que aparece en cierto nivel podría estar perfectamente determinada por alguna ley de un nivel superior. Esa ley sería la responsable de impedir que las *escaleras de Penrose* existan en la realidad, aunque localmente parezcan posibles, o de permitir que los cuasi-cristales aperiódicos sean configuraciones físicas reales a pesar que no tienen un patrón local que las estructure. Por lo tanto, en el nivel local todo estaría determinado pero no todo sería computable, porque su deter-

minación proviene de un nivel superior o global. Es decir, la no computabilidad de Penrose se manifiesta localmente, pero remite a un elemento no-local.

Un paso ulterior, dado por nuestro autor, es admitir la posibilidad de que existan varios niveles de determinación. De este modo sugiere que en el nivel superior se podrían encontrar tanto la ley que rige el universo como la consciencia que actúa libremente. Desde esos dos niveles superiores de ley universal y libertad personal se determinarían los acontecimientos en los niveles inferiores.

La ley que gobierna el universo, sería temporalmente asimétrica y surgiría de una unificación de la relatividad general con la mecánica cuántica en una nueva teoría de gravedad cuántica. Con esta nueva teoría se resolvería la aparente aleatoriedad de los saltos cuánticos y se explicarían los fenómenos de entrelazamiento cuántico y de no localidad de los efectos EPR. La nueva gravedad cuántica regiría el comportamiento de las partículas en los niveles cuánticos conforme a una dirección. Las aparentes aleatoriedades fundamentales responderían a algunas motivaciones globales del universo a distintos niveles, como el hecho de que en el mundo haya seres sentientes y seres conscientes que hacen ciencia.

Aunque sea llamativo, Penrose parece defender a la vez la selección natural y cierto principio antrópico débil, en una especie de «selección natural dirigida». En este caso, la selección natural sería la causa local pero insuficiente de que aparezca la consciencia porque pasaría por alto el propósito teleológico de la existencia de consciencia. Las *decisiones* selectivas estarían dirigidas desde un nivel superior, sin que coincida aunque se aproxime al concepto de principio antrópico:

«(...) el argumento del principio antrópico débil *podría* proporcionar (al menos) una razón para que la consciencia exista sin que tenga que ser favorecida por la selección natural. Pese a ello, yo no puedo creer que el argumento antrópico sea la razón auténtica (o la única razón) para la evolución de la consciencia. Hay evidencia suficiente procedente de otras direcciones para convencerme de que la consciencia *tiene* una poderosa ventaja selectiva, y no creo que se necesite el principio antrópico».

La aparente contradicción de la cita anterior adquiere nueva luz con la siguiente:

«una característica notable de las estructuras teselantes cuasi-cristalinas que he estado describiendo es que su ensamblaje es necesariamente no-local. Es

decir, al ensamblar las estructuras es necesario, de cuando en cuando, examinar el estado de la estructura a muchos y muchos ‘átomos’ de distancia del punto de ensamblaje si queremos estar seguros de no cometer graves errores cuando juntemos las piezas. (Esto es quizá parecido al aparente ‘andar a tientas inteligente’ al que me referí con la selección natural)».

Ese *andar a tientas* vendría dado por la *selección natural* mientras que la nota de *inteligencia* vendría dada por las decisiones *dirigidas* desde el nivel no-local que examina desde la distancia.

Con todo lo expuesto se llega a la conclusión de que para Penrose la ley que gobierna el universo es determinista y es la causante de que existan seres conscientes. Ahora bien, con esta conclusión, se suscitan dos nuevas preguntas: cómo es posible que una consciencia pueda emerger de la materialidad y cómo es posible que dicha consciencia mediante la voluntad influya realmente en la realidad.

Como se ha visto, Penrose asocia el fenómeno de la consciencia a una acción coordinada sobre una infinidad de neuronas del cerebro, que estaría provocada por el fenómeno de reducción orquestada del vector de estado (*Orch OR*) en los microtúbulos neuronales. Como los microtúbulos se encuentran en muchas estructuras de los seres vivos, Penrose llega a sostener que incluso los paramecios realizarían cierta actividad consciente, porque tienen un citoesqueleto formado por microtúbulos. En función de la complejidad de las estructuras se darían grados de consciencia, más elevados en el caso de los mamíferos, y muy especiales en el caso de los hombres.

Por tanto, para Penrose la consciencia *emerge* de lo material y señala a los microtúbulos como las estructuras en las que se podrían dar las condiciones de posibilidad de las acciones conscientes debido a los efectos de la gravedad cuántica durante el colapso de la función de onda.

En la naturaleza, el fenómeno el colapso de la función de onda se produciría de modo natural por la acción de la gravedad cuántica, después de transcurrido un tiempo durante el que la partícula está superpuesta. Si esa partícula estuviese entrelazada dicho colapso se podría producir antes por un fenómeno de decoherencia cuántica. Además, en lugar de una partícula se podría tener un grupo de partículas que son coherentes entre sí y, por lo tanto, se superponen y colapsan conjuntamente. Esto sucedería constantemente en los microtúbulos. Por último, en entidades muy complejas como los cerebros humanos, podrían existir masas superpuestas coherentes y entrelazadas cuyo

colapso se podría producir de modo orquestado en distintas zonas del cerebro (implicando a una infinidad de microtúbulos neuronales) y asociarse a una acción consciente. Al producirse este colapso orquestado, la acción consciente, que estaría en un nivel superior, actuaría en la configuración local del patrón de tubulinas que surge en los microtúbulos.

Por tanto la paradoja de la medida se resolvería de tres modos: los dos primeros para cualquier partícula observada localmente y el tercero solo en las acciones conscientes que engloban a muchas partículas simultáneamente y que se dan solo en algunas estructuras y condiciones físicas especiales. El modo ordinario de colapso de la función de onda sería por decoherencia con el entorno, ya que las partículas individuales estarían normalmente entrelazadas. Sin embargo, este colapso escondería el proceso natural de colapso de una partícula, por efecto de la gravedad cuántica, cuando esa partícula se mantiene sin influencias externas (cuando es coherente). Y el tercer modo se daría por una acción coordinada de la consciencia en los cerebros sobre una enorme cantidad de partículas entrelazadas que se disponen a lo largo de muchas zonas del cerebro.

La consciencia en su parte pasiva sería excitada por la superposición cuántica y el momento volitivo tendría lugar en el colapso que todas las partículas entrelazadas al escoger un patrón. Visto localmente esta decoherencia sería un efecto gravitatorio, mientras que visto globalmente sería el fenómeno de la consciencia. También globalmente sería fruto de la ley que gobierna el universo.

Casi ningún autor apoya esta teoría por la cantidad de cabos sueltos que deja. Tampoco a mí me convence. Aun así, en lugar de rechazar de plano la propuesta, he preferido centrarme en dos puntos filosóficos interesantes que no están muy claros en Penrose: cómo se relaciona la libertad (parte activa de la consciencia) con la ley que rige el universo; y si concibe la libertad como auténtica capacidad de autodeterminación o como una simple apariencia fruto de una ley superior. En este terreno caben múltiples lecturas filosóficas –desde un espíritu hegeliano hasta una libertad que configura el mundo–, cada una de las cuales suscita nuevas y controvertidas preguntas.

La que parecería a primera vista más canónica conforme a los planteamientos de Penrose sería la que considera que existe una ley que gobierna el universo físico, junto con una libertad de autodeterminación de cada persona. La primera sería científicamente alcanzable y estaría presente en toda la realidad, mientras que la segunda sería la dimensión activa del fenómeno de la

consciencia. Dicha consciencia *emergería* como un fruto de esa ley universal que gobierna el mundo.

Desde sus intuiciones científicas y personales Penrose defiende tanto la realidad de la experiencia de la libertad como el determinismo universal a gran escala. Aun así, él mismo reconoce que ni el fenómeno de la libertad¹³³, ni la cuestión del determinismo a gran escala son tan claros como a él le gustaría que fuesen. En sus concepciones, el libre albedrío no se opone al determinismo sino a la computabilidad. Eso hace, que según sus categorías pueda existir un mundo perfectamente determinado pero no computable en el que quepa la libertad. Decir más, nos llevaría a no respetar la postura de Penrose, aun así conviene aclarar dos puntos que quedan implícitos en este enfoque.

El primero es que sus conclusiones, que intentan mantenerse en un equilibrio fisicista, difícilmente impiden desembocar en alguno de los dos extremos rechazados por Penrose: o un materialismo donde la libertad es solo aparente o un dualismo mentalista científicamente indemostrable. No sabría decir con claridad en cual desemboca. Desde el punto de vista computacional parecería caer en la segunda mientras que desde el punto de vista físico se inclinaría a la primera.

El segundo punto que deseo señalar es que en su deseo por encontrar las condiciones de posibilidad de la acción libre humana en la física entra de lleno en un el debate de gran solera como es el del determinismo y la libertad. Y aunque no consigue llegar a conclusiones muy convincentes, su distinción entre los niveles locales y globales merecerían una ulterior investigación, quizá priorizando la libertad humana sobre la ley del universo. Personalmente considero que precisamente en esos dos misterios de la ley universal y la libertad humana es donde mejor se manifiesta la condición de apertura de la ciencia. Me parece que en la ciencia no se puede encontrar la respuesta a estas dos preguntas: porqué el mundo es y es así; y porqué en él hay seres libres, para cuya existencia el universo parece tan preparado.

En resumen, me parece que el problema central de Penrose está en el punto de partida cuando compara el pensamiento humano con un algoritmo. Una cosa es que dicho pensamiento se pueda volcar en algoritmos, como también se pueden volcar en música, en palabras o en poesía y otra es reconocer que el pensamiento es mucho más rico. Por lo tanto, la cuestión no es que

¹³³ En Penrose 1999b, 535 escribe: «nuestra (¿ilusión de?) libre albedrío».

exista algo no-algorítmico. Eso viene dado. Lo no algorítmico es todo aquello de nuestro pensamiento que no se puede volcar en algoritmos. A mi modo de ver, la mayor parte.

Por eso Penrose, por el modo de argumentar y por lo que desea conseguir, acaba siendo presa tanto dentro del paradigma computacional – de computación cuántica, si se prefiere –, como del paradigma científico actual. No hay motivos para dudar de que el método científico sea consistente; sin embargo, sí los hay para dudar de que algunas preguntas formuladas por Penrose con relación al fenómeno de la consciencia, se puedan formular y responder dentro de ese paradigma.

Otros autores también resultaron presa del paradigma orgánico o del paradigma mecanicista de su época y es obvio que siempre nos acercaremos a la realidad con una pre-comprensión, pero también es lógico que aprendamos a reconocer los límites de nuestras aproximaciones. Aun así, sería injusto no reconocer el gran mérito de Penrose por intentar hacer una filosofía de la naturaleza y por hacer accesible el saber físico-matemático.

Índice del Excerptum

PRESENTACIÓN	253
ÍNDICE DE LA TESIS	261
BIBLIOGRAFÍA DE LA TESIS	265
LA PROPUESTA HEURÍSTICA DE ROGER PENROSE. EN BUSCA DE LAS BASES DE LA CONSCIENCIA	275
I. LÍNEAS BÁSICAS DE SU FILOSOFÍA	275
1. El método científico	276
1.1. La posibilidad del conocimiento	279
1.2. El punto de partida	280
1.3. El trabajo científico	282
1.4. Sentido común y apertura a la filosofía	286
1.5. Límites del método científico	293
2. La importancia de las matemáticas	294
2.1. El formalismo frente al realismo	295
2.2. La intuición directa como método de descubrimiento	300
II. NECESIDAD DE UNA NUEVA FÍSICA PARA ENTENDER LA MENTE	303
1. El estatuto de las matemáticas	304
1.1. Tipos de realismo matemático	307
1.2. Una alternativa al realismo de Penrose	309
2. Relación entre computación y consciencia	311
2.1. Cuatro perspectivas: A B C D	312
2.2. El argumento de John Searle	314
2.3. El argumento de David Chalmers	315
2.4. El argumento «científico»	316
3. Procedimientos de computación	318
3.1. Top-down y bottom-up	319
3.2. Caos	320
3.3. Aleatoriedad	321
4. No-computabilidad en el pensamiento matemático	322
4.1. Gödel y el programa de Hilbert	323
4.2. Turing y el problema de decisión	324

4.3. El nuevo argumento de Penrose	325
4.4. Alcance del argumento	328
4.5. Conclusiones	330
4.6. Necesidad de un elemento no-algorítmico	335
III. UNA APROXIMACIÓN HEURÍSTICA PARA UN NUEVO PARADIGMA CIENTÍFICO	337
1. Hacia la gravitación cuántica	338
1.1. Ecuaciones dinámicas y condiciones de contorno	339
1.2. Determinismo y probabilismo	345
1.3. Gravedad, reducción del estado cuántico y asimetría temporal	351
1.4. Elementos no algorítmicos	355
2. El fenómeno de la consciencia	358
3. Hacia la base física de la consciencia	363
4. Dificultades de la propuesta	366
CONCLUSIONES	368
ÍNDICE DEL EXCERPTUM	381

