UNIVERSIDAD DE SEVILLA

TESIS DOCTORAL

¿Triunfo del indeterminismo?

La apuesta científico-filosófica de Roger Penrose acerca de la naturaleza

Autor: Daniel Heredia González

Directores de tesis: Juan Arana Cañedo-Argüelles y Francisco Soler Gil

2020

Declaración de autoría

Yo, Daniel Heredia González, declaro que esta tesis doctoral titulada « ¿Triunfo del indeterminismo?: La apuesta científica-filosófica de Roger Penrose acerca de la naturaleza» es de mi autoría.

Confirmo que:

- Cuando he consultado trabajos previamente publicados por otros, se encuentra claramente señalado.
- Cuando he citado trabajos previamente publicados por otros, siempre es señalada la fuente. Con la excepción de dichas citas, esta tesis doctoral es enteramente fruto de mi propio trabajo.
- Señalo y reconozco todas las fuentes de ayuda que han sido utilizadas para la realización de esta tesis doctoral.

Firmado: Daniel Heredia González

Fecha: 2020

Índice

Capítulo 3

Argumentos físicos

Agradecimientos

Siempre he tenido la sensación de que mis palabras de agradecimiento suelen quedarse cortas, ya que me resulta imposible expresar lo que verdaderamente siento. En fin, cosas de quien no se gana la vida escribiendo. Una vez escuché que muy probablemente alguien haya dicho mejor lo que pretende decir uno mismo, y precisamente por ello es mejor aprovechar para citarle, de este modo te libras de expresarte pobremente. Es lo que haré para agradecer a la gente que citaré más adelante. Las palabras, como muchas(os) sabrán, son de Boccaccio:

> Toda fatiga tiene su castigo, y ahora prefiero gozar de cuanto apetecible hay en el reposo. Mas aunque haya cesado mi pena, no por ello ha desaparecido el recuerdo de los beneficios recibidos de aquellos que, por su benevolencia para conmigo, sufrieron con mis fatigas. Ese recuerdo no se borrará jamás, si no es con la muerte.

Son muchas las personas a las que me gustaría dedicarles estas palabras y muy seguramente se me olviden algunas, por ello pido que se me disculpe. Todo el mundo sabe que la memoria no es infalible.

Empezando por las personas que me ha regalado el mundo académico, no puedo dejar de mencionar a mis directores de tesis, Juan Arana y Paco Soler. Sin intentar molestar lo más mínimo a Paco Soler (por otro lado sé que no lo haré), quiero hacer una mención especial al Profesor Arana. Con él he tenido la oportunidad de crecer intelectualmente desde el Grado, pasando por el Máster hasta esta tesis. Aparte de haber sido un padre intelectual, también ha mostrado su atención y preocupación en lo personal, y por ello no tengo más que palabras de agradecimiento a su figura. Sinceramente, gracias.

Otro de los profesores a los que no quiero olvidar es Paco Valls, quien, sin duda, dota a la Facultad de Filosofía de una calidad humana innegable. Y para acabar con el profesorado, quiero acordarme de María de Paz, José Ferreirós y Jesús Navarro, con quienes he aprendido que nuestra facultad goza de una muy buena salud investigadora.

En lo que respecta a compañeros(as) de clase, es para mí una obligación nombrar a Daniel Pino y Noelia Domínguez, quienes hicieron mucho más fácil mi andadura en el Grado. También a Miguel Palomo y a quienes me acompañaron en el Máster, con especial cariño a Josepe y Ramón.

Del pueblo, a mis fieles amigos de siempre: José Diego, Fernando, Joselu, Dioni, Juanma, Juanjo, Antonio, Pepe, Pablo, José Carlos y Ricardo. A quienes aparecieron en el camino para quedarse: Paco, Nazaret, María, Ana Rosa, Marta, Mari, Pedro, Antonios, Dani, Fabi…

A quienes Sevilla cruzó en mi vida: Clara, Ane, Fran, Borre, Diego, Julia, Edu, Jesús, Santi, Juan, Josele, Paula, Vicky. Sin intentar molestar de nuevo,

debo decir que entre esta gente se alzan cuatro nombres en especial: Agathe, Claudia, Andrea y Carmen.

Por supuesto, a mi familia. Empezaré por la familia que he podido descubrir en los últimos años, la de Rajamanta. Gracias a Gregorio, Lola, Antonio, África, Quico y Mari. A mi familia alemana, Gloria, Sara, Bego y Manolo. Pero hablando de la familia a la que me une la sangre más directa, una persona es a quien más tengo que agradecer, y probablemente seguir agradeciendo, y ella es mi madre, Victoria. Verdaderamente no sé qué podría haber hecho sin ella. A mi hermano y mis hermanas, Violeta, Sandra, "Chico" y Beatriz. A los nenes de la familia, Youna, Nelo y Saulo. Y a mis abuelos Juana y José.

Por último, a Ana Salguero y Marta Cumplido, ellas saben muy bien el porqué.

Índice de abreviaturas y figuras

Abreviaturas:

La nueva mente del emperador	NME
Las sombras de la mente	SM
Moda, fe y fantasía en la nueva física del universo	MFF
El camino a la realidad	EcalR
Lo grande, lo pequeño y la mente humana	GPMH

Figuras:

Prólogo

Recibidor de un hotel que tiene una escalinata. Al pie de esta, junto a una estatua de Apolo, Einstein y Bohr intentan acabar una discusión que ya les tuvo ocupados hacía un rato.

Invitándole a que se sentase en el sofá del recibidor, Einstein retomó la palabra y le dijo a Bohr:

- Voy a recapitular por si me he dejado algo en el tintero o por si he entendido de manera errónea algún punto de tu defensa. Dices que los procesos naturales son indeterministas. Que obviamente este indeterminismo no es pleno porque, de serlo, no podríamos explicar por qué no vivimos en un completo y absoluto caos. Con esto, ¿quieres decir que en algún sentido el determinismo no desaparece definitivamente de la imagen de la Naturaleza?

- Tus apuntes son correctos, pero tu pregunta está viciada. Lo que yo defiendo es que una vez el indeterminismo, sea en el grado que sea, entra en la imagen, por seguir con tu manera de decirlo, el determinismo en su sentido estricto queda anulado. El determinismo, como sabes, nos dice que los procesos están enlazados de forma necesaria. Es decir, que lo que ocurre está condicionado inamoviblemente por lo ocurrido y que, a su vez, condiciona también inamoviblemente lo que ocurrirá. No queda espacio para los procesos que no responden a esa necesidad. Ante tal panorama, ¿qué grado de verdad, y con esta expresión estoy intentando ser polémico adrede, podemos concederle a los resultados de la física que se está abriendo paso con un éxito explicativo abrumador, y que nos dicen precisamente que esa necesidad no es, valga la expresión, necesaria? Creo, querido amigo, que debes rendirte a la evidencia.

- Siento que hayas interpretado que intentaba viciar mi pregunta. Como te dije, hacía una recapitulación por si había entendido bien tu postura o no. Con lo que me respondes ya veo que no acerté del todo. Sin embargo, ahora que todas las cartas están encima de la mesa he de decirte que aún sigo sin estar de acuerdo contigo. Me toca preguntarte, ¿piensas que la física que se está desarrollando es definitiva?

- En estos instantes mucho me temo que la respuesta a esa pregunta es que no, si a lo que te refieres es que dicha física está completamente acabada. Por otra parte, y creo que en esto tampoco estaremos de acuerdo, sí que pienso que las respuestas que nos ofrece la nueva física, es decir, allí adonde aspira a llegar esta, sí que constituirá el final de un camino. Ese camino es el emprendido para describir fielmente el comportamiento de la Naturaleza. Lo que quiero decir, por tanto, es que el indeterminismo ha venido para quedarse.

- El indeterminismo que nos asola en estos momentos, querido Bohr, no es más que la falta de un conocimiento más completo, algo que al fin y al cabo obliga al indeterminismo a ser pasajero. Y no me entiendas mal, reconozco la precisión y el gran alcance de tu física indeterminista, pero lo que no puedo reconocer es que el final del camino esté lo cerca que vaticinan tus deseos. Y sí, digo tus deseos porque son estos los que dominan tu discurso más que el sentido común.

- El sentido común de los científicos que nos rodean se está viendo obligado a seguir aquello que defiendo. Así que si hay alguien que está yendo a contracorriente, estimado amigo, ese eres tú.

- Puede que la tendencia que le aguarda a la física en los años venideros será aquella que dices, y puede que la gente que piensa como yo se encuentre cada vez más sola. No obstante, la esperanza de que se retome el camino que considero correcto y ponga las cosas en claro no desaparezca. Por otro lado, no me preocupa demasiado tal desaparición, porque la Verdad no puede hacerlo. Ahora bien, es imprescindible que toda la tarea para la recuperación del determinismo no recaiga en mis manos, porque por mí solo, sin indicios suficientes y sin apoyo, no podría llevar lejos mi investigación…

Introducción

El presente trabajo intenta sintetizar las ideas filosóficas de uno de los científicos más importantes de los últimos tiempos: Roger Penrose.

Penrose se constituye como uno de los pensadores actuales con mayor carácter multidisciplinar y sus planteamientos no dejan indiferente en ninguno de los campos en los que interviene.

Este trabajo intenta recoger las principales sugerencias de Penrose en los distintos ámbitos del conocimiento. La estructura sigue dos criterios: i) el cambio de terreno en el que se expresa Penrose (ya sean las matemáticas, o la física, etc.) y ii) el orden en el que, considero, Penrose expone sus argumentos. El contenido de cada capítulo es el siguiente.

En el Capítulo 1 veremos el contexto en el que Penrose intenta exponer sus ideas filosóficas. Dicho contexto está situado en el debate de la posibilidad de una inteligencia artificial. Es bien conocida la postura contraria de nuestro autor a tal posibilidad y en este capítulo veremos desarrollados sus argumentos más destacados por sus críticos(as) y aquellos en los que Penrose se recrea de un modo más profundo. También podremos ver la clasificación de los diferentes puntos de vista que, considera Penrose, participan en dicho debate.

Tener en cuenta sólo la perspectiva de Penrose en materia de Inteligencia Artificial me pareció poco adecuado y por ello decidí hacer un breve repaso a la historia de esta corriente para, así, tener una panorámica más amplia. Una vez hecho este repaso resulta más cómodo tratar las perspectivas denominadas Inteligencia Artificial *fuerte* y la Inteligencia Artificial *débil*, que serán explicadas en el grueso del trabajo. El capítulo concluye con cuestiones que siguen vigentes hoy en día en el ámbito de la Inteligencia Artificial, las cuales serán respondidas según la opinión de Penrose.

En el Capítulo 2 pasaremos a ver de manera exclusiva un campo que, según Penrose, subyace a todos los demás, este es, el matemático. Empezaremos viendo qué papel juegan las matemáticas y cuál es su alcance. Más tarde, veremos cómo nuestro autor considera que, volviendo al tema anterior, el debate de la Inteligencia Artificial en realidad pertenece al debate matemático, en el cual se intenta dar respuesta a sus fundamentos. Para ello, haremos un repaso a las distintas perspectivas que participan en tal debate, siguiendo la clasificación que Penrose hace de ellas, la cual no coincide plenamente con la aceptada por norma general.

De un modo más independiente, aunque no del todo, veremos la influencia del trabajo de Kurt Gödel en el pensamiento de Penrose, sobre todo en lo que concierne al teorema de incompletitud. Nuestro autor defiende que el teorema de Gödel no sólo sirve para dar respuesta dentro del debate de los fundamentos de las matemáticas, sino que también puede llegar a tener competencias, y muy importantes, dentro del debate de la Inteligencia Artificial. Esta defensa, como podremos ver, es controvertida a la par que seria.

Una vez expuesto el teorema de Gödel, pasamos a las matemáticas puras. Penrose entiende que las matemáticas son fáciles de percibir en el mundo que nos rodea. Esta idea no suscita excesivos debates, ya que las perspectivas que creen que las matemáticas son un constructo humano pueden llegar reconocer una noción de este tipo. Otro asunto es cuando extendemos tal perspectiva a los números complejos, números más abstractos. Por su parte, Penrose cree que los números complejos ya no sólo también *son reales*, sino que les otorga un grado de realidad superior. Este tipo de ideas Penrose las sustenta en una postura filosófica a la que se adhiere de un modo peculiar, aunque por otro lado también tajantemente: el platonismo. Sobre el platonismo y cómo lo entiende Penrose habrá desarrollado un apartado completo, en el que podremos estudiar la noción que maneja nuestro autor y las críticas a las que esta se ve sometida. El capítulo concluye con la cuestión de en qué medida las matemáticas constituyen la consciencia humana.

El Capítulo 3 intenta poner en orden la reforma que Penrose quiere llevar a cabo en la física para que, de esta manera, se resuelvan los problemas que se plantean en los capítulos anteriores. Para hacer efectiva esa reforma, Penrose cree conveniente que sepamos cuán importante es la física, haciendo un repaso de los distintos ámbitos que pueden proporcionarnos conocimiento útil sobre la naturaleza. En un principio veremos qué elementos de la física clásica podemos seguir usando, aunque veremos destacadas las limitaciones de esta. Algo parecido sucede con la física moderna. La parte que corresponde a la exposición de la física moderna está compuesta de una descripción de los eventos que hicieron posible su aparición, con las diferentes aportaciones de distintas figuras significativas de la primera mitad del siglo XX. El motivo de todo ello tiene que ver con la importancia de los debates que Einstein mantuvo con Bohr, que son aquellos que Penrose quiere, en última instancia, retomar al exponer su pensamiento. Para nuestro autor, la derrota de Einstein en aquellos debates no fue definitiva, sobre todo si contemplamos la posibilidad de una reforma en la física actual, que es aquello que él mismo propone.

El modo en el que Penrose intenta hacer manifiesta la reapertura de los debates entre el determinismo einsteniano y el indeterminismo bohriano es construir argumentos en base a dos ideas que, considera, siguen manteniendo de alguna forma en vigencia la perspectiva determinista. Tales ideas son las implicaciones del experimento EPR y la función de ondas de Schrödinger.

El capítulo concluye preguntando hasta qué punto es posible una reforma en física, según los criterios que defiende Penrose.

El Capítulo 4 recoge lo tratado en los anteriores, pero desde un enfoque filosófico más directo. En la primera parte de este capítulo veremos dos nociones que constituyen la columna vertebral del pensamiento penroseano, estas son, la consciencia y el determinismo.

Con respecto a la consciencia tendremos la ocasión de tratar las características que destaca Penrose, las cuales hacen posible una definición si bien no definitiva, sí aclaratoria del tipo de consciencia al que se ha referido nuestro autor en todo momento. Como una constante en su pensamiento, la perspectiva de Penrose no deja de ser marcadamente particular. Prueba de ello es el modo en el que nuestro autor relaciona la existencia de la consciencia con la selección natural. Es cierto que no entra en detalles profundos o muy concretos de biología o de la teoría darwiniana, pero su aportación es innegablemente lúcida. La parte de la consciencia acaba con la defensa de Penrose de que la naturaleza de esta es necesariamente no-algorítmica.

Por su parte, del determinismo veremos una serie de definiciones para intentar comprender de un modo más esclarecedor con qué tipo de determinismo concreto nos hemos estado encontrando a lo largo del trabajo. Generalmente es conveniente que en primer lugar se defina el concepto para que su comprensión sea más fluida. Pero podremos ver que en este caso no resulta un problema el haber ignorado una definición previa antes de tratar el concepto, ya que las definiciones no añaden nada nuevo. En su lugar, estas sirven para poder tener una idea más esquemática del determinismo en cuestión. También serán tratados dos conceptos que generalmente suelen estar involucrados en el debate entre determinismo e indeterminismo, estos son, el principio antrópico y la entropía, a los cuales Penrose dedica parte de sus estudios.

Aparte de ser un capítulo que contenga temas marcadamente filosóficos, la ciencia no dejará de estar presente. Las investigaciones de Penrose en neurociencia también estarán incluidas, ya que estas incluyen aportaciones teóricas del propio Penrose, junto a Stuart Hameroff, como son la naturaleza de los microtúbulos. En este apartado en concreto nuestro autor vuelve a la defensa de la consciencia como un proceso no-algorítmico, pudiendo ser intuido a través de aquello que expone sobre los microtúbulos.

El capítulo acaba retomando una noción que es de una importancia capital en el pensamiento de Penrose: la realidad. Esta es tratada de manera muy concreta a través de la teoría penroseana de los tres mundos, tanto la original como la reformulación que el mismo Penrose hace de ella. No obstante, el capítulo terminará con la cuestión que, considero, nuestro autor pretende hacer frente en último término.

Como puede verse en la descripción del trabajo, he intentado ser fiel a lo que Penrose intenta decir con sus planteamientos, pero ello no impide que mi perspectiva influya a la hora de exponerlos, algo natural si lo que se pretende es hacer una tesis.

Decir también que he intentado no abusar de la utilización de fórmulas para la exposición de los planteamientos de Penrose. Simplemente he añadido aquellos que considero que son imprescindibles.

Espero que con este trabajo se susciten nuevas preguntas acerca de este gran pensador de nuestra época, siempre teniendo en cuenta lo ambiciosa que puede llegar a ser tal esperanza.

Capítulo 1

Escenario en el que Penrose expone su pensamiento: debate de la IA

1. La posibilidad de la computabilidad de la mente y la consciencia humana

1.1. Planteamiento [contextual] del problema

« ¿Llegarán los ordenadores a poseer inteligencia, a experimentar situaciones similares a las de los humanos?». Esta doble pregunta aparece en la portada del libro de Roger Penrose *La nueva mente del emperador* (en la edición en castellano de 1991), y es justo decir que sobre ella gira el grueso de la obra. La respuesta en particular del matemático inglés es que no, pero con unos matices que son importantes a tener en consideración. Que un ordenador adquiera cualidades que le hagan asemejarse al ser humano y que ello le haga responder de un modo determinado debido a la situación en la que está inmerso no es algo que Penrose niegue. Es más, lo llega a reconocer de muy buena gana (Penrose, 2012: 60-61), ya que los avances que la Inteligencia Artificial nos regala cada x tiempo nos permiten observar que estos son un hecho.

Pero la cuestión de fondo es otra. ¿Y cuál es esta? Lejos de querer hacer una filosofía del lenguaje o analítica, Penrose encuentra en las expresiones «inteligencia» y «experimentar» de la pregunta inicial el *quid* del asunto. Para poder entablar el debate acerca de la posibilidad de una Inteligencia Artificial o negarla es preciso tener claro qué se entiende por inteligencia. ¿Es

inteligente un ordenador capaz de jugar al ajedrez al nivel de un ajedrecista profesional o de hacer cálculos matemáticos a una velocidad que no está alcance de ningún ser humano? La victoria de *Deep Blue* sobre Gary Kaspárov en 1997, o los posibles cálculos que puede hacer cualquier ordenador ante problemas de cierta complejidad muestran capacidades que son dignas de ser consideradas como inteligentes, pero parece que el concepto «inteligencia» lleva forzosamente a ir más allá.

Existen otros dos conceptos que conviene tener claros para situarnos de manera adecuada en el pensamiento de Penrose. Estos dos conceptos son «mente» y «consciencia». Entendamos que la mente se refiere a las capacidades cognitivas que tenemos los seres humanos, mientras que la consciencia es la manera en la que el ser humano se conoce a sí mismo y las cosas que aprende.

El término «inteligencia» es más problemático y requiere de varias aclaraciones. Es por ello que el siguiente punto estará centrado en ofrecer una perspectiva amplia de este concepto para, luego, ver cómo este encaja en el esquema de los partidarios de la Inteligencia Artificial y, finalmente, en el pensamiento de Penrose.

1.2. La inteligencia, sus tipos y la posibilidad de una inteligencia artificial

> *«Pensar, analizar, inventar (me escribió también) no son actos anómalos, son la normal respiración de la inteligencia. Glorificar el ocasional cumplimiento de esa función, atesorar antiguos y ajenos pensamientos, recordar con incrédulo estupor lo que el* doctor universalis *pensó, es confesar nuestra languidez o nuestra barbarie. Todo hombre debe ser capaz de todas las ideas y entiendo que en el porvenir lo será.» (Borges, 1974:450)*

La inteligencia[1] es comúnmente definida como la capacidad que tiene la mente para pensar, aprender y tomar decisiones. En primera instancia parece que la inteligencia puede ser perfectamente esto que se nos dice con dicha definición. Pero un segundo análisis nos alerta de que a esta concepción le faltan algunos ingredientes más. Ejemplos de ellos son la capacidad lógica, creativa, de orientación, de autoconciencia, de entender los sentimientos, de memoria o incluso también la capacidad de enseñar. Todas estas

[1] Es digno de ser comentado que en la gran mayoría de los casos de mi búsqueda en enciclopedias, libros, artículos y trabajos de investigación acerca del concepto inteligencia, los resultados obtenidos estuvieran relacionados o directamente enfocados a la Inteligencia Artificial. De esta forma es fácil entender en qué estado se encuentra el interés de la investigación acerca de este concepto y bajo qué criterios se suele construir. No obstante, he pensado que lo mejor es dar una definición general (o más bien una definición de la inteligencia humana) para luego ver cómo encaja con la corriente que la domina casi por completo.

capacidades[2], sin duda, *requieren* de inteligencia o, mejor dicho, *son* constitutivamente inteligentes. Pero, ¿cuál de ellas define verdaderamente a la inteligencia? ¿De veras existe entre estas capacidades una jerarquía? En la actualidad existen tendencias en las que plantear este tipo de cuestiones carece de sentido.

Una de las más importantes es aquella que sigue a la teoría de las *inteligencias múltiples*, propuesta por el psicólogo estadounidense Howard Gardner. Según Gardner, la inteligencia no debe entenderse como el desarrollo equilibrado entre los distintos tipos de inteligencia. ¿Y cuáles son esos tipos de inteligencia? Gardner las clasifica del siguiente modo: inteligencia musical, inteligencia kinestésica-corporal, inteligencia lógico-matemática, inteligencia lingüística, inteligencia espacial, inteligencia interpersonal, inteligencia intrapersonal y la inteligencia naturalista[3].

Según Gardner, los distintos tipos de inteligencia deben entenderse como independientes. Por ejemplo, que alguien no sea muy hábil socialmente no le exime de ser inteligente si posee una gran capacidad para la lógica-matemática[4]. Es más, lo común es destacar en una inteligencia, en lugar de tener una gran capacidad en todas ellas. A pesar de todo, la independencia de las distintas inteligencias no puede ser plena. No suelen darse casos en los que sólo se posea una inteligencia o que las distintas inteligencias no tengan ningún tipo de relación. Hay inteligencias que se retroalimentan. Gardner llama a esta retroalimentación *combinación*:

---

[2] Nótese que con la primera definición se puede incluir sin ningún tipo de inconveniente al resto de animales, al menos a aquellos que tienen un cerebro desarrollado. Sin embargo, una vez se le añaden las demás características, esta inclusión se hace cada vez más compleja. Por ello vale aclarar que en todo momento estaré tratando sólo y exclusivamente al ser humano y sus capacidades.

[3] Podemos ver que los siete primeros tipos son fácilmente identificables, no haciendo falta una explicación acerca de en qué consiste cada una. No pasa lo mismo con la inteligencia naturalista. Probablemente la razón por la que fue añadida posteriormente se debiera a que no es tan intuitiva como las otras. Gardner explica por qué aumentó el número de inteligencias en los siguientes términos:

Para los ocho posibles criterios de una inteligencia, la inteligencia naturalista tiene un buen encaje. En este tipo de inteligencia, existe la capacidad central de reconocer particularidades como miembros de una especie. Por otra parte existe también la historia evolutiva de la supervivencia, que a menudo depende del reconocimiento de las características específicas y de evitar a los depredadores. Los niños pequeños hacen distinciones fácilmente en el mundo naturalista; de hecho, algunos niños de cinco años son mejores que sus padres o abuelos para distinguir entre las especies de dinosaurios.

Examinar la inteligencia naturalista a través de las lentes culturales o cerebrales pone el foco en algunos fenómenos interesantes. Hoy en día, pocas personas en el mundo desarrollado dependen directamente de la inteligencia naturalista. Simplemente vamos a la tienda de comestibles o pedimos comestibles por teléfono o Internet. Y, sin embargo, sugiero, que toda nuestra cultura de consumo se basa en la inteligencia naturalista. Esta incluye las capacidades que implementamos cuando nos sentimos atraídos por un automóvil en lugar de otro, o cuando seleccionamos un par de zapatillas o guantes en lugar de otro (Gardner, 2006: 19).

Es decir, que la inteligencia naturalista se caracteriza por la capacidad que se tiene al saberse dentro del entorno natural (el mundo) y de reconocerse como especie, sabiendo la responsabilidad que ello conlleva. Gardner reconoce que recibió muchas propuestas de otros tipos de inteligencias para añadirlas a su lista, pero sólo la naturalista fue la que le convenció, debido a que era la única que se constituía como independiente.

[4] La teoría de Gardner lleva a admitir que en este mismo caso de una persona con poca capacidad de relacionarse socialmente no deja de ser inteligente si posee una gran inteligencia intrapersonal. Este caso es más conflictivo, pero la teoría lo contempla como plenamente legítimo.

> Sin embargo, casi todos los roles culturales con cualquier grado de sofisticación requieren una combinación de inteligencias. Por lo tanto, incluso un papel aparentemente exclusivo, como tocar el violín, trasciende la dependencia de la inteligencia musical. Para convertirse en un violinista exitoso se requiere de una destreza kinestésica-corporal, también de habilidades interpersonales, como la que se debe tener para relacionarse con una audiencia y, de una manera diferente, para elegir un manager; de alguna forma este papel también implicaría una inteligencia intrapersonal. Bailar, por ejemplo, requiere habilidades en inteligencias kinestésica-corporal, musical, interpersonal, y espacial, en grados diferenciados. La política requiere una habilidad interpersonal, una facilidad lingüística, y quizás alguna aptitud lógica (Gardner, 2006: 22).

Esta, sin duda, es la característica principal de la teoría propuesta por Gardner, o, al menos, la que más polémica levantó y sigue levantando en los debates concernientes a la inteligencia. Pero antes de pasar a ver el porqué de esta polémica veamos los otros dos aspectos que destaca como fundamentales, ya que estos también contribuyen a una definición más extensa del concepto principal.

La primera de las características es la *singularidad* de los perfiles intelectuales (*intellectual profiles*), debido a su composición material. Dos personas (o tres o cuatro…) no pueden poseer las mismas cualidades intelectuales, ya que ello depende en última instancia del material que compone sus cuerpos (¡que siempre es diferente!). Gardner utiliza el ejemplo de los gemelos idénticos. Si bien el aspecto de tales individuos es innegablemente semejante, la materia de la que están hechos es distinta y particular. Por ello es plenamente comprensible que puedan tener capacidades intelectuales totalmente opuestas.

La segunda tiene que ver con la importancia que Gardner le da a la capacidad de elegir, o más bien a la de elegir bien (o como hemos estado viéndolo hasta ahora, la capacidad de resolver problemas). Este aspecto tiene una importancia capital dentro del esquema de Gardner, hasta el punto de que en todos los tipos de inteligencia la capacidad de resolver problemas está implícito en diferentes grados.

Resulta llamativo que estos dos aspectos no hayan creado una polémica mínimamente parecida a la que suscitó la independencia (con colaboración) de los distintos tipos de inteligencia. Poner encima de la mesa temas como la relación de la materia con la inteligencia o entender como rasgo común la toma de decisiones, en condiciones normales, significa entrar en un terreno tremendamente escurridizo. A pesar de que no han pasado del todo desapercibidos sí que es notable la diferencia con respecto a la primera característica. Dejando a un lado este posible capricho circunstancial pasamos a ver qué críticas recibe la teoría de las inteligencias múltiples. Ello nos ayudará a comprender mejor cómo se usa actualmente el término inteligencia, que es aquello que nos conviene tener claro.

Los psicólogos opuestos a Gardner están de acuerdo en que la inteligencia se compone de diferentes facetas. Pero que estas facetas sean independientes de tal modo que alguien pueda poseer en alto grado una de ellas y ser nulo en otra, todo ello sin perder la condición de ser inteligente, es necesariamente incorrecto. Si la inteligencia se caracteriza por algo es por poseer un equilibrio entre las distintas capacidades. Por ello es tan importante para este bando del debate la información que nos ofrecen los test de inteligencia. Pero, ¿hasta

qué punto es conveniente poner en manos de estos test la posesión de la última palabra en un tema tan complejo como lo es el definir la inteligencia? Ni los mismos partidarios del desarrollo y perfeccionamiento de los test de inteligencia reclaman tan atrevido fin. Aquello de lo que sí parecen no estar dispuestos a renunciar es a los datos experimentales y a las pruebas empíricas, que es precisamente de lo que carecen las teorías de inteligencias múltiples[5].

La cuestión de fondo es si la inteligencia depende de un equilibrio entre sus distintas facetas, teniendo que estar todas estrechamente relacionadas, o si, por el contrario, hay independencia entre ellas de tal modo que podemos hablar de inteligencias múltiples. La tendencia actual (y entendamos que un consenso real como tal no existe[6]) es entender la inteligencia como ese conjunto de distintas capacidades que guardan relación entre ellas y que ayudan a resolver problemas. A cuanta más inteligencia, tanta más facilidad para resolver problemas. Es decir, que una especie de combinación entre las dos posturas vistas son las que generalmente se acepta. No obstante, parecen que siguen faltando elementos, al menos conceptuales. En ninguna de las dos teorías se contemplan conceptos tan importantes a la hora de hablar de inteligencia, como lo son comprensión o consciencia. Y no estoy insinuando que los ignoren. En mi opinión cometen un error más grave, este es, dar tales conceptos por aceptados o entendidos. Más adelante[7] veremos qué lugar ocupan dichas concepciones con respecto a la de inteligencia. Como por ahora no resulta conflictivo entender inteligencia al modo de Gardner y sus detractores seguimos con la exposición.

Una vez allanado ese terreno veremos cómo podemos relacionar tan difícil concepto con la posibilidad de una inteligencia artificial.

Visto lo visto nos podemos preguntar: ¿es posible una inteligencia artificial? Como personas afines al campo de la filosofía podemos hacernos las preguntas que queramos (¡faltaría más!). Otro asunto es obtener respuesta y más teniendo en cuenta que este es sólo el inicio de la exposición del trabajo. Lo conveniente es seguir aclarando los distintos conceptos que se irán viendo a lo largo del mismo y con la mayor de las cautelas.

---

[5] Esta es la principal acusación de los detractores de la teoría de las inteligencias múltiples. Unos de los mayores exponentes de esta postura fue el psicólogo Hans Eysenck, quien criticó ferozmente a Gardner particularmente.

Eysenck aducía que la teoría de las inteligencias múltiples no respondía claramente a las preguntas fundamentales acerca de la inteligencia. Es más, dicha teoría es tan vaga, en cuanto a aportaciones empíricas se refiere, que permite cualquier tipo de respuesta (Eysenck, 1998: 108). Otro de los reproches que le dedica Eysenck a Gardner tiene que ver con la lucha que el psicólogo estadounidense mantiene contra quienes defienden la completa dependencia de una inteligencia general. Para Eysenck esto no tiene sentido, ya que esta postura está prácticamente obsoleta desde hace tiempo y no cuenta con representantes serios en la actualidad (Eysenck, 1998: 109).

[6] La situación de desacuerdo es de la mayor actualidad. Max Tegmark, en su obra *Vida 3.0* (2017), cuenta una anécdota relacionada con este fenómeno que dice así: Mi mujer y yo tuvimos la buena fortuna de atender recientemente a un simposio sobre inteligencia artificial organizada por la Fundación Nobel de Suecia, y cuando se les pidió a un panel de principales expertos investigadores en IA que definieran inteligencia, discutieron largamente sin llegar a un consenso. Encontramos esta situación muy divertida: no hay acuerdo sobre qué es la inteligencia ni siquiera entre inteligentes investigadores de inteligencia. Vemos entonces que no hay una definición indiscutiblemente «correcta» de inteligencia (Tegmark, 2017: 71)

[7] Véase nota 12.

Hasta ahora hemos visto en varias ocasiones la expresión Inteligencia Artificial. Es el momento de determinar qué definición le corresponde para saber exactamente con qué tipo de Inteligencia Artificial nos encontraremos de aquí en adelante.

La Inteligencia Artificial es aquella corriente filosófica que defiende la posibilidad de crear un ente inteligente a través de diferentes procesos y asentado en un aparato[8]. Según Russell y Norvig (2003), la ambición de los partidarios de esta corriente está dividida en dos vertientes:

1) Los que pretenden alcanzar el modo de actuar y de pensar de los seres humanos, por lo que es estrictamente necesario llegar a conocer nuestra naturaleza.

2) Los que pretenden alcanzar el modo de actuar y de pensar propio de la racionalidad, haciendo falta para ello tener conocimiento del perfecto funcionamiento de la razón, con la ambición de superar a la humana.

A partir de ahora me referiré, para distinguirlas de un modo más esquemático, a la primera vertiente como IA-1 y a la segunda como IA-2.

La vertiente IA-2 no tiene necesariamente que aspirar a superar al ser humano a través de conocer su naturaleza. De hecho, para ser considerada una corriente distinta a IA-1 debe entenderse de modo tal que esta pretenda alcanzar una inteligencia diferente a la del ser humano. En IA-1 se da por supuesto que justo en el instante en el que las máquinas lleguen al nivel del pensamiento humano estas serán superiores. El principal problema es si ese alcance por parte de las máquinas tendrá lugar o no. En realidad este no es un problema para los partidarios de IA-1, que tienen la convicción de que sí tendrá lugar.

La diferencia de IA-1 con respecto a IA-2 reside en que para IA-2 no es necesario llegar al nivel del pensamiento humano. Se debe ir más allá de este, perfeccionando el camino hacia una racionalidad correcta. Cuando IA-2 defiende la obtención de dicha racionalidad no quiere decir que considere al ser humano como carente de razón. No se trata de arrebatar al ser humano una característica que siempre le fue asociada. La exposición de IA-2 tiene más que ver con que nuestra razón no es perfecta y para una máquina no debería resultar imposible poder superar al ser humano en esta faceta. Por otra parte, esta vertiente también es consciente de la dificultad del problema. Pero, al igual que en IA-1, la meta es considerada alcanzable.

Otro rasgo común que existe entre IA-1 e IA-2 es que ambos pretenden llegar a sus respectivas metas (que, como hemos visto, en realidad es la misma) a través de la computación[9]. Pero, ¿qué es la computación? Generalmente se suele entender a la acción de introducir en una máquina un conjunto de conocimientos y métodos científicos para que esta pueda manejarlos de manera automática[10]. Tal y como Penrose aclara (Penrose, 2012: 32-34), la computación consiste, en mayor medida, de dos tipos de procedimientos:

---

[8] Generalmente dicho «aparato» es electrónico.

[9] De hecho, hay quienes entienden la Inteligencia Artificial como Inteligencia Computacional. Véase, por ejemplo Poole, Mackworth y Goebel (1998).

[10] Aparatos y procedimientos computacionales han existido desde casi siempre. Pero como estamos situados en el contexto de la Inteligencia Artificial entendamos que las máquinas a las que me referiré serán computadoras (u ordenadores, o como quiera llamárseles, ya que nombraré a estas de modo indiferenciado).

i) de-arriba-abajo, que son aquellos cuya estructura es fija, permitiendo una solución única y, por tanto, precisa.

Y por el otro lado tenemos los procedimientos:

ii) de-abajo-arriba, que son entendidos de manera tal que sus reglas pueden sufrir variaciones. La solución también puede ser precisa, pero no constituye un paso necesario dentro de su estructura[11].

Pero a pesar de que la computación es un aspecto común en ambas vertientes, en ella también se encuentra la diferencia fundamental anteriormente vista. Mientras que en IA-1 la computación debe tener como base la forma en la que un humano piensa y actúa, para IA-2 ello puede resultarle incluso irrelevante. Este aspecto es más conflictivo para IA-1 que para IA-2, en el sentido de que IA-1 está suponiendo que la inteligencia humana responde en última instancia a una actividad computacional. Y esto no es, ni mucho menos, evidente. De hecho, como hemos visto más arriba, el debate que surge a partir de esta idea será aquel en el que estará centrado este trabajo.

¿Dónde queda la postura de Penrose en todo este asunto? Como sucede con casi todo su pensamiento, su defensa constituye un peculiar punto de vista. Nuestro autor entiende el concepto inteligencia tal y como lo hemos visto, pero con la particularidad de que la inteligencia está subordinada a la consciencia[12]. ¿Quiere decir que la inteligencia es irrelevante o que, al menos, tiene una importancia mínima? Ni mucho menos. Como su nombre indica, el debate que hemos estado viendo está situado en la posibilidad de una inteligencia artificial. Por supuesto que para Penrose es importante dicha concepción. El asunto es que nuestro autor considera que es imprescindible

---

[11] Penrose lo expone desde un plano neutral (o matemático). Yo sigo situado dentro del ámbito de la Inteligencia Artificial. Visto desde esta perspectiva es más fácil detectar la diferencia entre estos dos tipos de procedimientos. Dicha diferencia reside en que aquellas máquinas que responden al primer procedimiento no necesitan de un aprendizaje, ya que tienen todas las herramientas necesarias desde el momento en el que son programadas. Mientras, las que derivan sus respuestas acorde al segundo procedimiento, al ser programadas de manera *incompleta*, sí que requieren dicho aprendizaje.

[12] Penrose encuentra necesario entender de manera adecuada cuatro conceptos fundamentales en este debate. Estos son «conocimiento», «comprensión», «consciencia» e «inteligencia». Para nuestro autor, y nosotros lo entenderemos del mismo modo de aquí en adelante (aunque sin ignorar lo ya visto acerca de la inteligencia), existe una relación entre estos cuatro conceptos. Esa relación de es de subordinación o requerimiento. Penrose lo explica tal que:

a) «inteligencia» *requiere* «comprensión»

y

b) «comprensión» *requiere* «conocimiento» (Penrose, 2012:54).

«Consciencia», por su parte, sería aquello que estaría *detrás* de todas ellas. Penrose reconoce el carácter tanto activo como pasivo de la consciencia, pero él prefiere no hacer esa distinción y decide entenderla en su forma general (Penrose, 2012: 55).

Cabe destacar un matiz sobre los conceptos que queda más claro si atendemos a la versión inglesa. Para referirse a «conocimiento», Penrose habla de «awareness» (que al castellano se traduce como «conciencia») y no de «knowledge» (que es como normalmente se traduce «conocimiento»). El motivo por el que el traductor lo realiza de este modo es para hacer patente el carácter *pasivo* del conocimiento (del *awareness*, la *conciencia*) para lograr distinguirlo de la consciencia («conciousness»).

Entonces, cuando nos topemos más adelante con cualquiera de estos cuatro conceptos se deberá tener en cuenta la relación existente entre ellos, esta es, el conocimiento como ingrediente esencial de la comprensión, la comprensión como parte de cualquier inteligencia auténtica (Penrose, 2012: 56) y que una inteligencia auténtica no puede darse sin consciencia.

entender esta –llamémosla así- jerarquía entre la inteligencia y la consciencia, ya que en ella se encuentra la clave de todo. Para Penrose es imposible que las máquinas tengan una inteligencia al modo humano porque ellas no pueden llegar a tener consciencia. Él lo expresa del siguiente modo:

> En mi propia forma de ver las cosas, la cuestión de la inteligencia es subsidiaria de la de la consciencia. Me parece poco concebible que la verdadera inteligencia pudiera estar presente a menos que estuviera acompañada de la consciencia. Por el contrario, si resultara que la gente de la IA fuera finalmente capaz de simular inteligencia sin que la consciencia esté presente, entonces podría considerarse insatisfactorio definir el término «inteligencia» sin incluir esta inteligencia simulada. En ese caso, el tema de la «inteligencia» no me interesaría realmente para este punto. Estoy interesado principalmente en la «consciencia».
> Cuando afirmo mi propia creencia en que la verdadera inteligencia requiere consciencia, estoy sugiriendo implícitamente (puesto que yo no creo en la tesis de la IA *fuerte* de que la simple activación de un algoritmo produciría la consciencia) que la inteligencia no puede simularse adecuadamente mediante procedimientos algorítmicos, es decir, mediante una computadora, en el sentido en que hoy utilizamos el término. […] En efecto, argumentaré con fuerza dentro de un momento […] que debe haber un ingrediente esencialmente *no-algorítmico* en la actuación de la consciencia (Penrose, 1991: 505).

Es decir, que para Penrose la posibilidad de que las máquinas lleguen a poseer una inteligencia no es imposible. De hecho, lo puede llegar a contemplar como una posibilidad más que probable. Pero todo cambia cuando la meta es llegar a la inteligencia humana, algo que nuestro autor niega rotundamente. En un principio, entonces, Penrose no lanza su crítica contra aquellos que siguen la vertiente IA-2, sino contra los partidarios de IA-1.

Esta última idea es de una gran importancia en el pensamiento de Penrose. Pero ella no nos ofrece un plano general del esquema penroseano. Este propósito lo he reservado para el punto que viene a continuación.

1.3. Panorámica de la visión de Penrose

Que las máquinas sean capaces de *mostrar* algunas de las aptitudes humanas e incluso varias de ellas a la vez no significa que puedan *poseer* la naturaleza de nuestra especie, ya que esta es demasiado compleja como para que se reduzca a tal concepción. Esta es la respuesta de Penrose a la pregunta de si es posible una Inteligencia Artificial al modo humano. Pero, ¿es realmente inalcanzable nuestra naturaleza? ¿Qué nos hace tan únicos que es imposible hacer computable nuestras capacidades? ¿Acaso, precisamente, la no-computabilidad? Nuestro autor cree que en la respuesta a esta última pregunta está la clave del asunto. En la naturaleza humana existen rasgos no computables, rasgos que nos caracterizan con respecto a las máquinas más sofisticadas que el ser humano pueda crear. Este es un argumento demostrable a través de la ciencia.

A pesar de ofrecer una respuesta tajante, lo cierto es que nuestro autor apela a una respuesta con carácter de proyecto. En la actualidad no puede obtenerse un veredicto definitivo para el debate de la computabilidad o no

computabilidad de la mente y la consciencia humana. Pero en un futuro nada es descartable[13]. Sobre todo si la ciencia toma el camino que «debe» tomar.

Penrose postula que es imprescindible entender la mecánica cuántica para poder arrojar luz sobre el verdadero funcionamiento de las cosas (entre estas, nuestros cerebros) y poder observar que es imposible la computación de nuestra consciencia. Pero esto, obviamente, no es una tarea fácil. El mismo Penrose tiene la sospecha de que esto es así. Está proponiendo nada más y nada menos que cambiar los fundamentos de la teoría física más prolífica del último siglo. La ventaja con la que cuenta Penrose es que es un gran conocedor de la mecánica cuántica, así que conoce de sobras la magnitud de aquello que está proponiendo, es decir, que no está dando palos de ciego. Su proyecto puede ser considerado, paradójicamente, tanto vago como firme. Vago porque la reforma que plantea Penrose está lejos ya no de ser lograda, sino planteada. Y firme porque nuestro autor está plenamente convencido de que dicha reforma es condición necesaria para poder alcanzar respuestas más contundentes.

Aun teniendo en cuenta estos aspectos cabe preguntarse, ¿pretende Penrose dar una respuesta definitiva? Sabiendo de la elocuencia y capacidad crítica que posee el matemático inglés es fácil darse cuenta que a lo que realmente aspira es a seguir planteando problemas más que a dar soluciones que no admitan discusión:

> Yo no puedo proporcionar tales respuestas: nadie puede, aunque algunos puedan tratar de impresionarnos con sus conjeturas. Mis propias conjeturas jugarán un papel importante en lo que sigue, pero trataré de distinguir claramente tales especulaciones de los hechos científicos brutos, y trataré también de dejar claras razones que subyacen a mis especulaciones. No obstante, mi principal propósito aquí no es el de conjeturar respuestas sino el de plantear algunos temas aparentemente nuevos concernientes a la estructura de la ley física, la naturaleza de las matemáticas y del pensamiento consciente, y de presentar un punto de vista que no he visto expresado hasta ahora (Penrose, 1991: 24).

Si bien los temas que trata Penrose en sus obras no son genuinos (a estas alturas pedir genuinidad a quien se atreve a poner encima de la mesa problemas filosóficos resulta inapropiado), su forma de tratarlos sí que constituye un punto y aparte en los debates en los que se adentra. Plantear una reforma de la mecánica cuántica desde el debate de la Inteligencia Artificial supone un riesgo del que él mismo es (valga la palabra) consciente. Pero aun así decide entrar, con las consecuencias que ello conlleva.

Entonces, ¿existe la posibilidad de computar la mente y la consciencia humana de manera que podamos introducirla en una máquina para que esta pudiera pensar y actuar como nosotros? Penrose sentencia que no y he de

[13] Cuando Penrose trata este tema lo hace en términos muy filosóficos, como lo es, por ejemplo, la realidad. En NME dice: […] No obstante, algún día la ciencia podrá darnos acceso a una comprensión *más* profunda de la que nos proporciona por ahora la teoría cuántica. Mi opinión personal es que incluso la teoría cuántica es insuficiente e inadecuada para ofrecernos hoy una imagen del mundo en el que realmente vivimos. Pero que esto no nos sirva de excusa. Es preciso que comprendamos la imagen del mundo según la teoría cuántica existente.
Por desgracia, los teóricos tienden a tener enfoques muy diferentes (aunque desde puntos de observación similares) sobre la *realidad* de esta imagen (Penrose, 1991: 287-288)

reconocer que yo mismo me siento inclinando a apoyarle, aunque con algunas diferencias que intentaré dejar claro a lo largo de este trabajo.

Las ideas desarrolladas siguen la senda de las obras de Penrose que más han trascendido en el ámbito de la filosofía. Estas son *La nueva mente del emperador* (1989) y *Las sombras de la mente* (1994). A pesar de ello, no he descuidado otras obras importantes como *Camino a la realidad* (2004), *Lo pequeño, lo grande y la mente humana* (1997) o *Moda, fe y fantasía en la nueva física del universo* (2016), entre otras.

En NME, por ejemplo, el interés del autor se centra en mostrar las peculiaridades del pensamiento matemático, sobre todo en la primera parte, para ver que en estas se encuentran una de las claves para demostrar la no-computabilidad de la mente y la consciencia humana. Mientras, en SM, también en la primera parte, podemos observar que su discurso se inclina más hacia una aclaración de lo relevante que puede ser el lenguaje formal para refutar la postura de la Inteligencia Artificial[14].

Si bien esta diferencia existe entre una obra y otra, aquello que se mantiene en ambas es la confianza que Penrose deposita en el quehacer científico para dar cuenta de sus argumentos. Nuestro autor entiende que si la ciencia es aquella que nos da las respuestas más fructíferas acerca de la naturaleza, y nosotros (con nuestra mente y nuestra consciencia) pertenecemos a la naturaleza, lo conveniente es echar mano de la ciencia:

> Una visión científica del mundo que no trate de entender en profundidad el problema de la mente consciente no puede tener pretensiones serias de compleción. La consciencia es parte de nuestro universo, de modo que cualquier teoría física que no le conceda un lugar apropiado se queda muy lejos de proporcionar una descripción auténtica del mundo. Mantendré que todavía no existe ninguna teoría física, biológica o computacional que esté cerca de explicar nuestra consciencia e inteligencia consiguiente, pero eso no debería detenernos en nuestro intento de búsqueda de una. […] Quizá algún día se formularán todas las ideas a este respecto. Si es así, es casi seguro que nuestra perspectiva filosófica quedará profundamente alterada. No obstante, todo conocimiento científico es un arma de dos filos. Lo que realmente *hacemos* con nuestro conocimiento científico es otra cuestión. Tratemos de ver dónde pueden llevarnos nuestras visiones de la ciencia y la mente (Penrose, 2012: 22).

Aparte de hablar de lo adecuado de la ciencia como sustento de sus argumentos, Penrose pone encima de la mesa un tema que en el caso de los partidarios de la Inteligencia Artificial parece pasar desapercibido: la ciencia puede ser un arma de doble filo. El entusiasmo por obtener los resultados esperados provoca que se pierdan de vista las posibles consecuencias de dichos resultados. ¿Acaso es tan imposible que la situación se vuelva cada vez más compleja con el avance de la tecnología? Penrose tiene claro que no y por ello piensa que es estrictamente necesario ser responsables. Pero si las máquinas llegasen a evolucionar hasta el punto de ser independientes, ¿a quién (o quiénes) pediríamos dicha responsabilidad? Con la expresión «hacemos» de la cita vemos que Penrose entiende que la responsabilidad de

[14] Cabe aclarar que en NME también se encuentra la crítica al lenguaje formal y que en SM, del mismo modo, aparece destacada la particularidad del pensamiento matemático. Sólo que en NME (en la parte que he destacado) su crítica gira entorno a los argumentos en contra del test y la máquina de Turing (para hacer patente las características del pensamiento matemático), mientras que en SM el teorema de Gödel tiene un protagonismo más que palpable, sirviendo ello de acicate a su postura.

todo lo que pueda ocurrir debe recaer en los seres humanos. Los posibles avances dependen de la consciencia humana y no de la de los ordenadores (¡estos no llegarán a tenerla nunca!).

Su postura es clara. Pero en un debate es necesario conocer los demás puntos de vista para, así, poder argumentar y contraargumentar en base a lo que defiende y rechaza cada uno de ellos. La clasificación de la que se sirve Penrose es la que veremos en el punto que sigue.

2. El A, B, C, (y D) del debate de la IA

2.1. Diferentes puntos de vista que Penrose tiene en cuenta

El plano general de las características del pensamiento de Penrose ha sido tratado en gran medida. Aun así, este no nos permite entender dicha postura dentro de un marco concreto. El mismo Penrose se percata de ello y encuentra necesario hacer una clasificación entre los distintos puntos de vista del debate para, de esa forma, no dispersarse en sus críticas. Esta es la clasificación que Penrose ofrece en *Las sombras de la mente*:

> A. Todo pensamiento es computación; en particular, las sensaciones de conocimiento consciente son provocadas simplemente por la ejecución de computaciones apropiadas.
> B. El conocimiento es un aspecto de la acción física del cerebro; y si bien cualquier acción física puede ser simulada computacionalmente, la simulación computacional no puede por sí misma provocar conocimiento.
> C. La acción física apropiada del cerebro provoca conocimiento, pero esta acción física nunca puede ser simulada adecuadamente de forma computacional.
> D. El conocimiento no puede explicarse en términos físicos, computacionales o cualesquiera otros términos científicos (Penrose, 2012: 26)[15].

Tenemos que A pertenece a la denominada IA *fuerte* (o dura). Este punto de vista defiende que los misterios que guardan la mente y la consciencia humana [sin excepción] son susceptibles de ser conocidos. Es decir, que podemos tener pleno poder tanto de la mente como de la consciencia, hasta el punto de poder introducirlas en una máquina y que adquiera dichas capacidades. En un principio puede parecer que esta corriente aboga por una explicación materialista de la consciencia y la mente. Sólo haría falta encontrar los dispositivos adecuados para llevar a cabo la labor pretendida.

[15] Considero preciso aclarar que dentro de los diferentes puntos de vista que nos ofrece Penrose existen conflictos, hasta el punto que corrientes que en esta clasificación entran en el mismo bloque de pensamiento son consideradas dispares. Pienso que es de rigor tenerlo en cuenta, pero es más práctico (probablemente Penrose también lo hiciera por este motivo) no comenzar un ejercicio de división y subdivisión entre los puntos de vista según qué tipo de matices diferencian a uno de otros, ya que ello sería contraproducente para el fin de la exposición pretendida aquí.

No obstante, Penrose señala, acertadamente, que en A se intercede a favor del papel de la información[16] más que de la materia[17]:

> Supongo que este punto de vista –el que afirma que los sistemas físicos deben ser considerados como entidades meramente computacionales- es consecuencia en parte del papel creciente y poderoso que juegan las simulaciones por ordenador en la ciencia moderna del siglo XX, y también de la creencia de que los objetos físicos son meramente «estructuras de información», en cierto sentido, que están sujetas a leyes matemáticas computacionales. Después de todo, la mayor parte de la materia de nuestros cuerpos y nuestros cerebros está siendo reemplazada continuamente, y es sólo su *estructura* lo que persiste (Penrose, 2012: 28).

Es decir, que lo material puede pasar a ser un elemento secundario, una mera «estructura de información» que simplemente responde a lo que le dicta su programación matemática. El punto de vista A será sobre el cual Penrose dirige la mayor parte de su crítica. De todos modos ello no impide que nuestro autor reconozca que siente admiración por ella, ya que la meta principal de esta corriente es sacar todo el jugo posible del quehacer científico. Toda corriente que otorgue a la ciencia la última palabra será del agrado de Penrose, al menos en esa parte. Ello no impide que Penrose se encuentre en las antípodas en los demás aspectos de este punto de vista en concreto.

El modelo B se corresponde con la comúnmente conocida IA *débil*. En B se defiende que el comportamiento de los objetos físicos puede responder a una operación computacional, al igual que sucede en A. La diferencia entre estos dos puntos de vista reside en que para A un artilugio cuando responde a su programa computacional lo está haciendo «conscientemente», mientras que en B no lo es, o al menos no está claro que lo sea. La principal razón sobre la que se sostiene este matiz es que la composición física (es decir, material) del cerebro no puede extrapolarse a la que la máquina[18] tiene. Por tanto, podemos ver que el papel secundario que tenía la materialidad en A pasa a tener uno principal en B, provocando ello la diferencia más notable entre ambos puntos de vista. Por otra parte, aunque pertenece al mismo aspecto, tenemos también que la computación, esencial en A, ahora en B cobra un papel más bien fútil (Penrose, 2012: 29). La computación sirve para *simular* consciencia y la materia es lo que *permite* tener consciencia.

El punto de vista C no tiene un nombre identificativo al modo en el que lo tienen los dos anteriores puntos de vista. C es el nombre que Penrose da a su propia postura y así es conocida. Como este punto de vista será el más tratado de aquí en adelante, por ahora nos ahorraremos dar pinceladas acerca de él.

Por último tenemos D. En D encontramos una postura que rechaza toda respuesta proveniente del ámbito científico, condición que le hace estar relacionada con la mística (Penrose, 2012: 26). Al defender que es imposible para la ciencia resolver los misterios de la mente y la consciencia humana, lo mejor que esta puede hacer es guardar silencio. Aunque en principio esta

[16] Entre las diferentes acepciones que tiene el concepto «información», aquí lo debemos entender como la serie de conocimientos que se introducen en la máquina para que estos constituyan la estructura de su conciencia. Como podemos observar, este tema no está exento de polémica, ya que se usan conceptos que entrañan muchas aclaraciones. Por ello creo que lo más aconsejable es entenderlo en su forma más general.

[17] Obviamente no se elimina la materialidad (tal y como puede advertirse en autores como Edelman, Tononi o Tegmark –véase Soler, 2017: 299-310), sólo que la información cobra una importancia crucial.

[18] Entendamos que dicha máquina se comporta como lo haría un ser humano.

visión choca de frente con el modelo planteado por Penrose (en el que el papel de la ciencia es crucial) ello no es óbice para que ambas posturas se encuentren en un punto común. En C existe un desencanto con la ciencia actual. Tal es así que ello le lleva a renegar de los resultados obtenidos por dicha ciencia. ¿Existe la misma desconfianza con respecto a la ciencia? Pues habría que responder con cierta trampa: sí y no. Sí, en el sentido de que es un hecho que la ciencia en la actualidad no permite respuestas contundentes y ello la mantiene impotente en este debate en concreto. Y no, en cuanto que Penrose no piensa abandonar la ciencia, sino que cree que esta necesita sin paliativos una remodelación. Penrose aclara dicho matiz del siguiente modo:

> […] Según C, el problema del conocimiento consciente es realmente un problema científico, incluso si por el momento no se dispone de la ciencia adecuada. Yo apoyo firmemente este punto de vista; creo que es con los métodos de la ciencia –aunque ampliada adecuadamente de formas que quizá sólo pueden ser vislumbradas en el momento presente- como debemos buscar nuestras respuestas. Esta es la diferencia clave entre C y D, por mucho que puedan aparecer posibles similitudes en las correspondientes opiniones respecto a lo que la ciencia *actual* es capaz de conseguir (Penrose, 2012: 30).

Definitivamente, D se desmarca tanto de C como de A en este apartado, pero no, así, tanto de B. En realidad, el punto de vista D es el único que permite adaptarse a todos. Es obvio el porqué: este punto de vista es el menos atrevido. Al no tomar una postura firme y preferir guardar silencio, alguien puede optar por D sin impedirle ello poder especular e incluso tener la convicción de la posible veracidad de los restantes puntos de vista. Pero como estos no permiten dar con la solución al debate lo mejor es reinventar el proverbio árabe: si lo que vas a decir no es más correcto que el silencio, no digas nada.

Sigamos viendo aspectos comunes y diferentes entre estos puntos de vista.

2.2. Algunas características más de los diferentes puntos de vista

Para empezar, C comparte con B la idea de que lo físico (o al menos en la mayoría de los casos) resulta un territorio inexpugnable para los programas de las computadoras. ¿Pero en qué se basa la duda de Penrose acerca de que las ciencias son incapaces de explicar la no computabilidad de la consciencia? ¿Es acaso la composición de las mismas, la forma de ser entendidas, no quedando claro de esta forma las metas a alcanzar? Y si fuera así, ¿no se estaría intentado dar un vuelco a los planteamientos de la ciencia? Ya vimos más arriba que esta es precisamente la postura de nuestro autor. Es conveniente dejar claro que esta visión del modelo C comprende la postura de Penrose. Existen partidarios de este punto de vista que entienden que la física actual es apropiada para hacer frente al problema que se debate. Sólo haría falta un nuevo enfoque de los distintos problemas no resueltos, pero siempre dentro del marco de la física que tenemos. Penrose, como hemos visto, se sitúa en un punto más radical, en el que la reforma de la física que plantea haría temblar sus cimientos. Con esta declaración C se proclama como el punto de vista más atrevido y ambicioso. Sin embargo, ello tampoco le libra de ser el más nebuloso, ya que al fin y al cabo está dejándolo todo en

manos de una reforma que hoy en día no está ni tan siquiera planteada de manera formal. Si bien quienes defienden A no pretenden cambiar la ciencia, sí que comparten la ambición de poder dar una respuesta definitiva a través de ella. Por su parte, el hecho de que B sea una postura más cauta en este aspecto hace que se la considere como el punto de vista de «sentido común científico» (Penrose, 2012: 29), ya que no considera a la ciencia incompleta o inadecuada, ni tampoco la pone en el aprieto de tener que dar respuestas definitivas.

Un factor estrechamente vinculado a lo anterior es la relación que cada uno de los puntos de vista guarda con respecto al fisicalismo y al mentalismo. ¿Por qué cree Penrose que es importante hacer una aclaración sobre este tema en particular? Por norma general estas relaciones se malinterpretan y dan lugar a confusiones acerca de lo que defiende y rechaza cada uno de los modelos, algo inadecuado si lo que se pretende es entablar un debate serio.

En primer lugar, el punto de vista A es considerado normalmente como perteneciente al fisicalismo. No obstante, esta consideración es problemática. El motivo de la problematicidad de relacionar este modelo concreto con el fisicalismo reside en su relación con el materialismo. Aquello que se defiende en A entra en claro conflicto con aquello que defiende el materialismo.

Antes de seguir con este asunto concreto es conveniente hacer una aclaración previa, esta es, intentar definir los rasgos generales del materialismo y el fisicalismo. Recordemos, en primer lugar, el materialismo. El *materialismo* es una corriente filosófica monista, para la cual todo es material o físico. Todo aquello que en un principio no parezca ser constitutivamente material, como lo puede ser la mente, el materialismo lo sigue entendiendo como un fenómeno de lo material. A lo largo de la historia el materialismo ha abarcado los más diversos debates filosóficos y ha conseguido mantenerse en escena sin descanso. La importancia de esta corriente es debida al concepto que le da su nombre: materia.

Lejos de hacer un estudio profundo acerca de la materia, sí que es pertinente entender algunos puntos sobre esta[19], ya que ello nos ayudará a entender qué tipo de materialismo es aquel que maneja Penrose en su discurso.

Como apunte general, es bien sabido que la materia suele entenderse dentro del plano concreto de la naturaleza, siendo las expresiones «forma» y «espíritu» aquellas que de manera más común se contraponen a ella (Arana: 714). La materia es cambiante mientras que la forma o el espíritu se presumen constantes. Es ampliamente conocido que la decisión de darle primacía a lo material o a lo formal (o a lo espiritual) ha sido objeto de debate en innumerables ocasiones, sobre todo en el plano filosófico.

Otro aspecto importante de las características de la materia es, en primer lugar, la extensión, la cual implica «una multiplicidad de partes unidas y separadas por relaciones de coexistencia y exterioridad» (Arana: 714). En segundo término la materia también cuenta con la cohesión y la impenetrabilidad, las cuales suponen «fuerzas de atracción y repulsión, capacidad de ejercer acciones causales para mantener unidas las partes que

[19] Para la siguiente aproximación al concepto de materia me he servido de una comunicación personal con Juan Arana. Los números de páginas citados pertenecen al que le corresponde en los escritos enviados, pero al no haber sido publicados no puedo añadir la fecha ni será incluido en la bibliografía general.

integran un determinado cuerpo o impedir que una injerencia extraña las separe» (Arana: 714).

Volviendo al papel de la materia en el terreno de la filosofía y teniendo en cuenta lo visto hasta ahora, se sabe que este concepto estuvo ya presente desde los presocráticos (manifiesto en el *arjé* de Tales, Anaxímenes, Empédocles o los atomistas Demócrito y Leucipo). No obstante, no fue hasta Aristóteles cuando este concepto adquirió una notoriedad mayor (Arana: 715). El estagirita fue el encargado de establecer la teoría hilemórfica, la cual establece una relación de dependencia entre materia y forma. Habiendo dicho que la materia y la forma suelen entenderse de manera antagónica, dicha relación resulta cuanto menos particular. A pesar de ello, esta encaja sin problemas en el esquema aristotélico:

> […] Al negar la existencia independiente de la forma, y por simetría también de la materia, Aristóteles convierte una y otra en coprincipios indisolublemente unidos de la realidad sensible. El único modo de separar una forma de la materia es transformar esta última (dotarla de otra forma). La materia da la cifra de la variabilidad de las cosas, su aptitud para ser modificadas, para alcanzar un estatuto diferente. En sí misma nada es, pero sin ella nada puede ser. Lo propio de la materia es la potencia, como lo que caracteriza a la forma es el acto (Arana: 715).

Si bien esta relación materia-forma puede tornarse compleja, lo cierto es que en ella se acude a principios que traducen de manera fiel la realidad que observamos y pensamos. Ese entrelazamiento entre su complejidad y su fidelidad a la hora de describir la naturaleza logró que la polémica sin acuerdo estuviera servida durante muchos años.

Con el surgimiento de la ciencia natural (no reconocida como tal por entonces) la situación de desacuerdo sufriría un cambio considerable. Por supuesto que persistirían los debates, pero un hecho hizo que el asunto adquiriese un enfoque diferente, este fue la matematización de la materia, teniendo en Descartes uno de los mayores representantes de este movimiento:

> […] Cansados de distingos y aporías, los estudiosos volvieron la vista hacia la matemática, disciplina que había enseñado a los astrónomos el modo de resolver de una vez por todas enigmas seculares. Cundió la desconfianza frente a la argumentación lógica y el concepto aristotélico de materia quedó irreversiblemente erosionado. La primera alternativa seria que se propuso para suplantarlo fue la cartesiana. Descartes sigue apegado al criterio de que es preciso evitar las ambigüedades e imprecisiones del lenguaje ordinario. Quiere definir con exactitud qué tipo de entidad posee lo que genéricamente llamamos materia y cuál es su esencia. Recurre para ello a la idea clara y distinta de sustancia extensa, y toma sobre sus hombros la ingente responsabilidad de explicar el mundo material a partir de una única forma y sus variaciones (Arana: 716).

La idea clara y distinta de la sustancia extensa propuesta por Descartes tiene como consecuencia necesaria tomar el concepto materia de un modo *cerrado*. Este es uno de los modos en los que el concepto materia puede abordarse. Por otro lado tenemos la concepción *abierta* de la materia.

Es importante tener en cuenta la diferencia entre el concepto cartesiano (*cerrado*) y el newtoniano (*abierto*) de materia, porque esta daría lugar a gran parte de los debates posteriores (Arana: 717). Descartes defiende que para alcanzar una ciencia rigurosa es imprescindible que sus conceptos sean claros y distintos, que no den lugar a la imprecisión. La apuesta de Newton con respecto a la rigurosidad de la ciencia no se basa en manejar conceptos

volátiles. Sin embargo, sí que entiende que se debe ser más flexible a la hora de utilizar ciertos conceptos fundamentales. La ciencia posterior tomaría en mayor medida el camino de Newton. El proyecto newtoniano apartaba de las manos de una sola perspectiva la tarea de definir el concepto materia, haciéndolo depender de un desarrollo colectivo a través de la historia (Arana: 717). Sin duda, Newton supuso un antes y un después en el devenir de la ciencia occidental y su brazo se extendió de manera que nadie puede negarlo, pero también es cierto que existieron alternativas. Se suele distinguir tres alternativas principales, que son las siguientes:

> a) Imitar su actitud circunspecta y seguir estudiando la forma de matematizar aspectos de la materia, sin pretender agotar el conocimiento de la realidad que designamos con ese término, ni reducir su contenido ontológico al de un concepto perfectamente acabado y definido […].
> b) Estimar que no hizo más que empezar el trabajo, porque la tarea del físico sólo estará completa cuando haya sido definido un concepto de materia capaz de explicar del todo el comportamiento de las realidades a las que dicho concepto se atribuye. A tal propósito se barajaron las propiedades mecánicas (posición, impenetrabilidad, inercia y leyes del choque) contempladas por Descartes, aun sin pretender reducirlas a una sola idea clara y distinta. Éste es el punto de vista del mecanicismo moderno […].
> c) Intentar llegar a una definición completa, pero no a partir de las propiedades mecánicas, sino de las dinámicas. La materia estaría constituida por una red de centros inextensos unidos por fuerzas de atracción o repulsión. Dichas fuerzas se regirían por leyes matemáticas que determinarían el comportamiento del todo resultante (Arana: 718).

Podemos ver que en las alternativas el concepto materia sigue siendo abierto, y precisamente es en esa dirección del materialismo en la que decide moverse la ciencia de hoy en día, como por ejemplo lo hace Penrose. Si bien nuestro autor es partidario de concretizar en los asuntos que trata, también hemos observado que en lo referente al tratamiento de conceptos se muestra muy cauto. Es por ello que cuando Penrose menciona el materialismo es coherente pensar que lo hace acerca de aquel en el que su concepto de materia es abierto. Con respecto a quienes defienden el punto de vista A, no es descabellado pensar que el concepto de materia que les gustaría manejar es el cerrado, pero, considero, que aquel que usan (siempre dentro del debate tratado por Penrose) es el abierto.

Una vez vistas algunas de las características principales del concepto materia para, así, determinar a qué tipo de materialismo se refiere Penrose en su exposición, volvamos ahora a la diferenciación entre el materialismo y el *fisicalismo*.

El *fisicalismo* es una corriente menos amplia [y por tanto menos comprometida] que el materialismo, ya que ha desarrollado sus debates mayormente dentro de la filosofía de la mente. El fisicalismo defiende que aquello cuanto existe u ocurre está constituido por entidades físicas.

También es conocido como fisicalismo la doctrina que defiende que cuanto existe u ocurre puede ser descrito plenamente a través de la física. Es innegablemente manifiesta la relación que existe entre materialismo y fisicalismo (soliéndose entender, incluso, al fisicalismo como un tipo de materialismo). Pero el matiz de la amplitud del materialismo con respecto al fisicalismo logra distinguirlas.

Ambas acepciones del fisicalismo ayudan a colocar las piezas del problema de A con el materialismo. Regresemos, por tanto, al problema mencionado al inicio de este punto.

El modelo A estipula que la composición material del artefacto puede llegar a ser irrelevante a la hora de incorporar los elementos necesarios para que este obtenga consciencia. Lo verdaderamente importante es el programa computacional. Tenemos por un lado que el fisicalismo está estrechamente relacionado con el materialismo. Por otro lado que el materialismo no encaja con lo defendido en A con respecto a la materia. Por lo tanto, relacionar el punto de vista A con el fisicalismo no resulta ni cómodo ni evidente.

Por su parte, B y C son claramente fisicalistas. Tanto B como C están en consonancia completa con el fisicalismo, incluida su relación con el materialismo.

En el otro lado tenemos el mentalismo, que es aquella corriente que defiende la existencia de «lo mental», pero sin descartar la dependencia con respecto a la materia y lo físico[20]. En lo que respecta a esta corriente debemos saber que D es el único punto de vista indiscutiblemente afín a ella. A, B y C no contemplan el mentalismo, al menos de manera explícita.

Por las razones expuestas cuando más adelante me refiera a corrientes *mentalistas* sólo incluiré a D, mientras que para las *fisicalistas* estaré mencionando a las tres restantes, siempre teniendo en cuenta la particularidad de A que fue señalada.

2.3. Algunos puntos de vista más

A Penrose se le ha reprochado que en un debate de tal envergadura el pensador inglés contemple de manera inadecuada los diferentes puntos de vista involucrados en este debate. Una de las críticas más destacadas es la llevada a cabo por el filósofo Aaron Sloman en su artículo "The Emperor´s Real Mind"[21] ("La verdadera mente del emperador", en clara referencia a *La nueva mente del emperador* de Penrose, ya que es un trabajo de revisión de dicha obra).

Dicho artículo contiene dos críticas principales: i) la poca precisión de Penrose a la hora de hablar de los partidarios de la Inteligencia Artificial, ii) el replanteamiento de la utilización penroseana del teorema de Gödel. En este punto sólo veremos la primera crítica, ya que es aquella que está relacionada con lo que venimos viendo. La segunda parte la encontraremos desarrollada en el capítulo 2.

Sloman considera que el modo en el que Penrose se refiere a los partidarios de la IA es claramente deficiente, ya que incluye diferentes puntos de vista

[20] Algunas teorías actuales comprenden al mentalismo dentro del materialismo. En un principio estas no las tendré en cuenta, pero en el caso de aparecer sería convenientemente aclarado.

[21] Aunque el artículo original es de 1992 y está publicado en *Artificial Intelligence*, yo he utilizado la versión electrónica con lista de contenidos y un nuevo post scriptum publicado en 2018, que está disponible en la siguiente dirección URL: https://www.cs.bham.ac.uk/research/projects/cogaff/sloman-penrose-aij-review.html

como uno solo, sin tener en cuenta los matices que diferencian a unos de otros:

> Frecuentemente, Penrose se refiere a la «gente de la IA» como si hubiese un acuerdo ortodoxo en el campo. Por supuesto que todo campo tiene defensores ingenuos, malinformados o sobreentusiasmados, y la IA no es una excepción, habiendo atraído a muchos graduados en ciencia computacional, física o matemáticas sin preparación en filosofía o ciencia cognitiva. Sin embargo, el campo tiene adherentes más sofisticados […] (Sloman, 2018: 4).

Los defensores de la Inteligencia Artificial, por tanto, no pueden ser acotados todos en un mismo grupo. Por ello mismo Sloman ofrece una nueva y más extensa clasificación de los diferentes modos de defender la posibilidad de una inteligencia artificial, todas ellas incluidas dentro de la IA *fuerte*. Dicha clasificación consiste en nueve teorías diferentes, que denomina como T1, T1a, T2, T3, T4…

Algunas de ellas, defiende, son tratadas por Penrose, aunque también de manera inadecuada. Aquellas que son desarrolladas por el matemático inglés son las más extremas y Sloman las denomina como T1 y T1a. Estas dos teorías defienden que la clave de todo se encuentra en un único algoritmo que aún no ha sido descubierto. Este algoritmo sería de una gran sutileza y también susceptible de ser conocido. Pues bien, Sloman cree que tratar de un modo serio a este tipo de teorías es una tarea inútil, ya que la defensa de tal principio es absurda. Y el rechazo no está basado en una opinión personal (que también) sino a que prácticamente nadie que defiende la posibilidad de una inteligencia artificial se basa sostiene sus ideas en este principio en particular.

Luego está T2. Esta teoría está relacionada con el enactivismo. Debido al poco desarrollo que esta corriente tenía por entonces[22] y a que no deja claro que todo dependa de un algoritmo, Sloman cree que es demasiado pronto como para someterla a un análisis profundo, ya que se podrían discutir de aspectos de esta que aún no han sido pulidos. No obstante, Sloman deja entrever que esta teoría tiene mucho más potencial que las dos anteriores, a pesar de que llegue a contemplar en algún momento la importancia de un algoritmo con una singularidad concreta.

Todo esto cambia con T3. Esta teoría, si bien tampoco niega la existencia de tal algoritmo desconocido y «especial», se diferencia de las demás al defender que aparte de tal algoritmo también existen múltiples procesos de interacciones computacionales (*multiple interacting computational processes*). Es decir, que no descarta la composición de diversos algoritmos.

La composición múltiple de algoritmos de T3 deja patente un aspecto: los sistemas simples (de un solo algoritmo) dan paso a los sistemas complejos. T4 es otro ejemplo de ello. Esta teoría establece que los estados mentales se producen a partir de colecciones de procesos computacionales llevados a cabo en colecciones distribuidas de procesadores[23] (*distributed collection of processors*). Ya no sólo es importante entender que los estados mentales son provocados por complejos sistemas algorítmicos, sino que también en el medio en el que estos se dan también cobran notoriedad. No quiere referirse con ello a la materialidad, sino a la forma de la composición que permite la

[22] En la actualidad sigue siendo una corriente muy actual y poco tratada.
[23] Entendamos «procesadores» como circuitos electrónicos.

circulación de las colecciones de procesos computacionales. Por su parte, T5 es muy similar a T4, con la particularidad de que permitiría que el diseño de un agente inteligente requiriese de elementos no necesariamente computacionales. Según Sloman, esta última teoría resulta tan vaga al no declarar qué tipo de elementos no-computacionales serían necesarios que corre el riesgo de no ser realmente interesante como teoría a tener en cuenta (Sloman, 2018: 19). De todos modos no es declarada como tal.

T6 es una teoría que pone encima de la mesa temas hasta ahora no tratados por las anteriores teorías. El más importante de ellos es la relación que pueden tener los estados mentales con el medio. Esta teoría maneja la posibilidad de simular el mundo físico en una computadora. El fin de esto es ver si una mente creada a partir procesos computacionales que, defienden, son como los humanos, actuaría del mismo modo que los humanos. El problema surge cuando se intenta simular el medio físico en términos computacionales, ya que puede darse el caso de que existan rasgos en el mundo físico que se escapen de procesos computables. En el caso de encontrarse con tales rasgos no computables la simulación no sería posible. Aparte de este inconveniente, T6 asume que en la interacción del ser humano con el medio no existen procesos esencialmente continuos (*essentially continuous processes*). En este caso, entonces, es difícil poder determinar hasta qué punto tales procesos son caóticos. El supuesto de que los procesos finalmente fueran caóticos sólo haría evidente la imposibilidad de simular computacionalmente el mundo físico (Sloman, 2018: 32).

La situación de T6 lleva de nuevo a T5, en el sentido de que es necesario encontrar procesos no-computacionales que consigan dar con la clave del asunto. Pero, como vimos más arriba, T5 no ofrece la alternativa. Algo similar a presentar tal alternativa será lo que hace T7. Esta teoría establece que es posible introducir en una computadora los conjuntos de procesos mentales siempre y cuando se haga en *tiempo compartido*. El *tiempo compartido* es un proceso que permite que varios usuarios ejecuten varios programas en una sola computadora al mismo tiempo, lo que facilita la amplitud del campo de actividad de los procesos computacionales. Si bien con T7 no se da lugar a procesos no-computacionales, no menos cierto es que los procesos que plantea añaden un plus de complejidad más propia de los procesos mentales. Vemos que en las últimas teorías dejan prácticamente de lado la idea de la búsqueda de un único algoritmo para la explicación de los procesos mentales para dar lugar a la idea de conjuntos e interacciones de procesos computacionales. Lo mismo sucederá con la última tesis que analiza Sloman, T8.

T8 es más radical a la hora de negar que un solo algoritmo sea el responsable de los procesos mentales. Estos son demasiado complejos como para que su actividad dependa del quehacer de un algoritmo, por muy sutil que este sea. En su lugar plantea que esta tarea podría ser encomendada a la implementación de una red de computadoras (*implementation on a network of computers*).

El modo que tienen estas teorías de la Inteligencia Artificial de intentar acercarse a los procesos mentales es añadir elementos que contribuyen a su complejidad y, por tanto, a su similitud.

Pero, ¿realmente aportan algo nuevo que invalide los argumentos de Penrose? Es cierto que Penrose comete el error de identificar a todos los

partidarios de la IA con aquellos que buscan un único algoritmo, ignorando los puntos de vista pertenecientes a la IA que descartan dicha búsqueda. Pero este desajuste sólo pertenecerá a NME. De hecho, con motivo de paliar tal error, en SM nuestro autor hace una mención explícita al artículo de Sloman (Penrose, 2012: 27).

Con respecto a NME, la crítica de SM es más amplia y menos específica. En ella no sólo sigue criticando la defensa de la búsqueda de ese algoritmo «especial», sino que también añade argumentos en contra de los conjuntos de algoritmos.

No obstante, Penrose en SM sigue considerando en el mismo punto de vista (el A) a todos los partidarios de los distintos tipos de IA *fuerte*. ¿Sigue Penrose cometiendo el mismo error? Personalmente pienso que no. Podemos partir de la idea de que la clasificación de Sloman es necesaria tenerla en cuenta. Pero no deja de ser cierto que las nueve teorías comparten una respuesta final en términos computacionales, ya sean estos más simples o ya sean más complejos. También es digno de mención que las teorías más complejas son menos consistentes, por lo que llevar a cabo un debate aparte con ellas carece de sentido. Además, cuando Penrose critica a alguna en especial hace mención a ella nombrando aquello que la caracteriza[24].

Siendo nueve, cuatro o uno, lo cierto es que el punto de vista defensor de la posibilidad de una inteligencia artificial tiene una relevancia innegable. Pero, ¿cuándo adquirió dicha importancia esta idea? ¿Cómo fue su origen? ¿Siempre fue contemplada del mismo modo? Para situar adecuadamente cómo entró en escena, en el siguiente punto veremos un repaso a la historia de la Inteligencia Artificial.

3. Breve historia de la Inteligencia Artificial

Querida Prenda
Tú eres mi ávido sentimiento amigo.
Mi afecto pende curiosamente
de tu deseo apasionado. Tú
eres mi triste simpatía: mi
tierno cariño.
Tuyo rendidamente,
(Computador de la Universidad de Manchester[25])

11 de mayo de 1997, Equitable Center, Séptima Avenida, Manhattan, Nueva York. Un hecho histórico ha sucedido. La computadora diseñada por IBM, *Deep Blue*, ha derrotado, en lo que suponía la revancha, en ajedrez al gran campeón ruso Gary Kaspárov, con un resultado de 2 partidas a favor del ordenador, 1 a favor del humano y 3 empates. Una máquina había conseguido superar a una de las personas que mejor ha jugado al ajedrez en la historia, y

[24] Es cierto que no se refiere a ellas de manera específica al modo de Sloman (T1, T1a,…), pero sí que se percibe dicho cambio con respecto a NME en este apartado.

[25] Carta de amor realizada por la máquina Mark I. Fue Alan Turing quien introdujo las cartas de amor mediante el generador de números aleatorios de la máquina (Copeland, 1996: 22).

además en un período de preparación relativamente corto. Un primer impulso lleva a concluir que las máquinas están muy cerca de asemejarse al ser humano, en lo referente a capacidad intelectual. Hay quienes más atrevidamente se lanzan a afirmar que la igualdad no existe: las máquinas nos han superado y es cuestión de tiempo que lo hagan definitivamente. Pero un segundo impulso lleva a la pregunta de si de verdad esa posible igualdad o superioridad se ha dado realmente a consecuencia de un acontecimiento de este calibre.

En este apartado veremos, en primer lugar, el camino que han recorrido las máquinas hasta la actualidad. En dicho repaso podremos ver el desarrollo técnico-científico de las computadoras y el asentamiento de la Inteligencia Artificial como debate filosófico. Aparte conoceremos cuáles son los fundamentos en los que se asienta esta teoría y las principales personalidades que contribuyeron a su desarrollo. Para acabar, volveremos a la pregunta de si es cierta la igualdad o superioridad de las máquinas.

Con respecto al estudio historiográfico de este apartado, podremos ver que estará dividido en dos partes:

1) El desarrollo técnico-científico de artilugios con capacidad computacional, donde veremos algunos de los modelos de computadoras más relevantes y también el contexto en el que dio lugar la aparición de dichos modelos.

2) El nacimiento del debate filosófico acerca de la Inteligencia Artificial, donde veremos la acuñación de este término y cómo este debate se ha ido asentando poco a poco como uno de los temas principales en el ámbito de la filosofía.

3.1. Los primeros pasos de la Inteligencia Artificial: desarrollo técnico-científico

Empezamos[26] por el año 1941. Fue entonces cuando un joven alemán llamado Konrad Zuse (1910-1995) culminó, en lo que sería su tercer intento, el proyecto de crear un computador programable de propósito general. A pesar de la hazaña este logro no pudo obtener el éxito ni el reconocimiento merecido. Los motivos por los que generalmente la aportación de Zuse no tenga una consideración mayor realmente se desconocen. Hay quienes achacan este hecho al desinterés del ingeniero alemán por el desarrollo comercial del computador, aunque también suele atribuirse a posibles restricciones impuestas por el régimen germano del momento, situación que no se resolvería hasta 1951 (Copeland, 1996: 21).

Dando un breve salto en el tiempo y en el espacio nos situamos ahora en Gran Bretaña, concretamente en Buckinghamshire, en la instalación militar inglesa de Bletchley Park. En dicho emplazamiento el computador también aparecería, pero a partir de una necesidad diferente. El propósito era conseguir un aparato que descifrara los códigos de las fuerzas armadas de la

[26] Para la línea histórica que sigue me he servido de los trabajos de Jack Copeland, *Inteligencia artificial* (1993) y la obra conjunta de Stuart J. Russell y Peter Norvig, *Inteligencia Artificial: un enfoque moderno* (2003).

Alemania nazi en la guerra. Por este motivo particular se dio lugar al Colossus, un computador electrónico[27], que haría las delicias de esta difícil empresa. El Colossus comenzó a operar en 1943 y logró los resultados esperados de él. Pero «sólo» consiguió cumplir con el rendimiento de esa tarea concreta. Los ingenieros responsables de la construcción del Colossus pusieron a prueba al aparato, intentado que hiciera otros cometidos (largas multiplicaciones, por ejemplo), pero dicho fin no fue conseguido. La máquina de Zuse, pese a ser anterior al Colossus, podía hacer frente a distintos quehaceres. La diferencia entre la computadora de Zuse y el Colossus es que la primera fue diseñada con propósito general, mientras que la segunda cumplía un propósito especial (Copeland, 1996: 21).

De todos modos Gran Bretaña no fracasaría en su intento por conseguir una computadora con propósito general, ya que cinco años después del nacimiento del Colossus, un equipo de Manchester, liderado por Freddie Williams (1911-1977) y Tom Kilburn (1921-2001), logró crear el Mark I (dando con el producto final un año después). El proyecto de diseño y construcción del Mark I tenía una importancia notoria, hasta el punto de que los mejores científicos del país llegaron a estar involucrados de alguna forma. Alan Turing, al cual veremos más adelante, llegó a formar parte del equipo de investigación encargado de elaborar el Mark I, aunque por distintos motivos abandonó el proyecto. Para entender la magnitud del Mark I cabe destacar que se convirtió en el primer computador electrónico de programa almacenado que se manufacturó de manera comercializada (Copeland, 1996: 22). Podría decirse entonces que la tendencia a la utilización y, por lo tanto, extensión de este tipo de aparatos serían un hecho a partir del Mark I.

Sin embargo, no sería hasta la intervención de Estados Unidos en esta empresa cuando el diseño y construcción de máquinas cobrarían una relevancia internacional.

Al igual que en Alemania, en el continente americano los artífices de la primera máquina han pasado a la historia (si es que lo han hecho) por pasar al ostracismo, más que por sus logros (que, por otra parte, no fueron pocos). El físico John Mauchly (1907-1980) y el ingeniero John Presper Eckert (1919-1995), decidieron crear un aparato llamado ENIAC. El ENIAC no resultó demasiado práctico, ya que para poder programarlo hacía falta montar y desmontar la máquina, lo cual suponía una pérdida de tiempo y un esfuerzo considerable. Pero ello no cejaría en el empeño de ambos y más tarde se dio paso al BINAC (una máquina de programa almacenado). Si bien estas máquinas fueron importantes y realmente útiles, el gran acontecimiento llegó cuando construyeron el UNIVAC, que supuso la primera oferta en el mercado de la industria informática norteamericana (Copeland, 1996: 23). Si bien el

[27] El mismo Copeland no deja pasar la oportunidad de explicar este aparato, ya que entiende que es un concepto que no debe ser pasado por alto. Por la misma razón he decidido citar su explicación: «Un computador electrónico es el que está construido con componentes electrónicos. En los primeros tiempos eran tubos de vacío y ahora semiconductores. […] Zuse empleaba relés. Estos son pequeños interruptores mecánicos accionados eléctricamente. Son mucho más lentos que los tubos de vacío. Los tubos deben su velocidad a que carecen de partes móviles, salvo el flujo de electrones (de ahí el término). En las calculadoras de antes de la guerra se usaban profusamente los relés, y, en fecha tan tardía como 1948, la IBM sacó un computador basado en esta vieja tecnología (llevaba 21.400 relés y 13.500 tubos de vacío). La máquina, llamada SSEC, estaba anticuada antes de ejecutar su primer programa» (Copeland, 1996: 21).

gran acontecimiento positivo pudo llegar a ser satisfactorio para los propósitos de estos pioneros, el destino les tenía reservado un revés no muy fácil de encajar.

Mauchly y Eckert se vieron envueltos en una polémica relacionada con la originalidad de su ENIAC. Al parecer, entre 1936 y 1942, el Doctor John Vincent Atanasoff (1903-1995) ya habría intentado construir una computadora, aunque nunca llegó a conseguir que fuese completamente funcional, ya que tuvo problemas con el almacenamiento de información en tarjetas perforadas (Copeland, 1996: 24). Esto podría no tener importancia en principio. Pero el asunto cambió porque Mauchly había visitado el laboratorio de Atanasoff antes de que el ENIAC saliera a la luz. A consecuencia de este hecho, las partes encontraron necesario entrar en liza. La sentencia del juez acabó dictada a favor de Atanasoff, ya que si bien no demostraba que Mauchly robara la idea a Atanasoff, la probabilidad de la influencia del trabajo de este último en el ENIAC era muy alta.

Por si esto pudiera parecer de poca importancia, Mauchly y Eckert de nuevo tuvieron que hacer frente a otro acontecimiento poco favorable para ellos, relacionado con el ilustre matemático húngaro-estadounidense, John Von Neumann (1903-1957):

> […] Von Neumann oyó hablar de ENIAC durante un encuentro casual en una estación de ferrocarril. En aquel entonces estaba trabajando en el Proyecto Manhattan en Los Alamos, donde aplicaba su gran genio a siniestros problemas tales como calcular la altura exacta a la que debe estallar una bomba atómica para infligir la mayor destrucción. En seguida vio las implicaciones de una máquina como ENIAC («trabajos en obuses, bombas y cohetes… progresos en el campo de los propulsores y explosivos… problemas de aerodinámica y de ondas de choque…»). Se ofreció como asesor en el proyecto de Eckert y Mauchly, y rápidamente se constituyó en portavoz nacional de la nueva tecnología de los computadores (Copeland, 1996: 24-25).

En principio, contar con una autoridad como Von Neumann para el proyecto no puede ser otra cosa que muy positivo para el mismo, ya que la importancia que se le daría cobraría un peso diferente. Pero la historia demostró una vez más que si bien todo tiene su cara buena, también es mejor hacerse a la idea de encontrarse la amarga. En este caso fue que el reconocimiento de la labor del célebre matemático relegó a un segundo plano tanto a Mauchly como a Eckert. Prueba de ello es que en cualquier libro sobre computadoras, el nombre Von Neumann sale a la palestra, mientras que los de Mauchly y Eckert sólo lo hacen de manera anecdótica[28]. Por otro lado, el «olvido» de estos dos investigadores no habría sido del todo gratuito, ya que el quehacer de Von Neumann (quien participó en el diseño de un ordenador muy sofisticado) influyó de manera innegable en los primeros pasos en programación (Copeland, 1996: 25-26). De hecho, una de las computadoras más prolíficas en la historia de estos aparatos, ya que marcaría el camino a seguir de los dispositivos ulteriores, fue llamada JOHNNIAC (debido a su nombre, John, y las siglas se corresponden a las iniciales de *numerical integrator* and *automatic computer*). El JOHNNIAC hizo acto de presencia

[28] Ello no significó que su trabajo fuera olvidado por completo, ya que ambos fueron reconocidos con diferentes distinciones, algunas de ellas entre las más importantes dentro del campo de la ciencia.

en 1953 y estaría activo hasta 1966, período de operatividad nunca antes conseguido por una máquina de sus características.

Fue precisamente en el JOHNNIAC donde se incorporó el programa que marcaría un antes y un después dentro de lo que se denominará Inteligencia Artificial, el Lógico Teórico. Pero aún no hemos llegado a ese punto de la historia.

3.2. La Inteligencia Artificial como debate filosófico

El espacio de la Inteligencia Artificial como objeto de debate filosófico pertenece por derecho propio al anteriormente citado Alan Turing. Esta es una opinión reconocidamente extendida como puede verse, por ejemplo, en la siguiente cita:

> Stanley Frankel, uno de los primeros científicos que emplearon el ENIAC, escribe: «Von Neumann conocía sobradamente la importancia fundamental del artículo de 1936[29] de Turing […] que describe los principios del Computador Universal del cual cada computador moderno […] es una realización […] Muchos han aclamado a Von Neumann como «padre del computador», pero estoy seguro de que él nunca habría cometido ese mismo error. Bien se le podría haber llamado la comadrona quizá, pero para mí, y estoy seguro de que también ante otros, insistía en que la concepción fundamental pertenecía a Turing, puesto que ni Babbage, Lovelace ni otros la habían anticipado» (Randell, cit. por Copeland, 1996: 31).

Tenemos entonces que Turing fue el genio que inició el camino de la Inteligencia Artificial. Sus ideas principales se encuentran en el artículo citado de 1936, "Sobre los números computables, con una aplicación al *Entscheidungsproblem*", y en el publicado en 1950, "Maquinaria computacional e Inteligencia".

En el primero de ellos, Turing nos presenta un concepto fundamental, como lo es la máquina que lleva su apellido. ¿En qué consiste dicha máquina y por qué es tan importante? Aquello que Turing expuso con su máquina fue que esta es capaz (una vez se le hayan introducido las instrucciones necesarias, la cuales no tienen por qué constituirse en un número muy amplio. Es más, sería conveniente introducir cuantas menos mejor, ya que ello probaría la inteligencia de la misma) de hacer frente a cualquier tipo de cálculo (matemático). En este artículo, y relacionada con la máquina, Turing da respuesta a aquello que se conoce como el problema de la parada[30].

En el segundo, surge la idea de poner a prueba a las máquinas, pero de otra forma distinta. Ello se daría a través de un test (que también lleva su apellido), que consistiría en que la máquina pudiera entablar una conversación (generalmente basadas en preguntas y respuestas) con una persona. La máquina pasaría el test si su interlocutor humano no fuera capaz de discernir si la conversación que está manteniendo es completamente humana o si existe un intruso (artificial) que intenta hacerse pasar por humano. Este tema, sin

[29] "Los números computables, con una aplicación al Entscheidungsproblem". En él, Turing presenta su máquina ideal y también da respuesta al problema de la parada (ambas cosas las veremos más adelante en este mismo apartado).

[30] Sobre este concepto veremos algún detalle en el capítulo 2 (concretamente en §2.4).

duda, será uno de los más importantes, sino el que más, dentro del debate de la Inteligencia Artificial.

Reconocido el papel de Turing, no hay que dejar pasar la ocasión de nombrar a otros actores (no menos importantes, dicho sea de paso) que también hicieron posible el desarrollo de la Inteligencia Artificial.

Quienes tienen el privilegio de ser reconocidos como los primeros autores de un trabajo relacionado con la Inteligencia Artificial son Warren McCulloch (1898-1969) y Walter Pitts (1923-1969), concretamente en 1943 (Russell, Norvig, 2004: 19). McCulloch y Pitts elaboraron un modelo matemático de una red neuronal artificial que dio grandes resultados. McCulloch tenía formación en el campo de la filosofía y la psicología, lo cual hace entender mejor el interés del trabajo por simular el comportamiento de las neuronas en el cerebro[31]. Por su parte, Walter Pitts no gozó de una educación al uso. Autodidacta, de pequeño aprendió lógica de forma notable, hasta el punto de que dominaba los problemas de los *Principia Mathematica* de Russell y Whitehead. Tal era su capacidad que acabó trabajando en la universidad, donde entablaría relación con McCulloch. Entre ambos conformaban un tándem muy valioso, ya que el conocimiento del cerebro que poseía McCulloch, junto a la destreza en lógica de Pitts sirvió para sentar las bases del modelo de redes neuronales que hoy en día lleva sus nombres y que tanto aportó al ulterior desarrollo en el campo de la Inteligencia Artificial.

No tardarían mucho en adoptar las ideas de McCulloch y Pitts en el plano técnico de la Inteligencia Artificial, de hecho sólo pasaron ocho años (1951). Dos jóvenes de Princeton, Marvin Minsky (1927-2016) y Dean Edmonds crearon una computadora que funcionaba siguiendo el comportamiento de las redes neuronales. Esta computadora es considerada como el primer aparato que marcó el devenir de la Inteligencia Artificial y prueba de ello es el prestigio que les reportó a sus creadores, sobre todo a Minsky, quien será considerado como una de las figuras más importantes de la Inteligencia Artificial.

Hasta ahora he estado aludiendo al concepto «Inteligencia Artificial» no haciendo hincapié en que en esos momentos de la historia tratados dicho concepto no se empleaba entonces. No sería hasta 1956 cuando el término fue acuñado por el informático John McCarthy (1927-2011). Sería con la ocasión de una reunión de investigadores que el mismo McCarthy organizó, con la ayuda de Minsky, Nathaniel Rochester (1919-2001) y Claude Shannon (1916-2001), y que denominó *The Dartmouth Summer Research Project on Artificial Intelligence* (o lo que es lo mismo, El proyecto de investigación de verano en Dartmouth sobre Inteligencia Artificial). Este taller tuvo un formato poco convencional, ya que la idea era que durara dos meses. En esos dos meses deberían fluir las ideas de los presentes (en forma de *brainstorming*) y ello proporcionaría una rica gama de perspectivas y contribuciones que, sin duda, marcarían el paso de la evolución de la Inteligencia Artificial. Pero el plan tendría sus grietas. La reunión era demasiado larga e intensa y ello provocaba que los investigadores optaran por estancias cortas, lo que también ocasionaba una irregularidad que poco

[31] De un tiempo hasta ese momento ya se había consolidado definitivamente la idea de localizar la actividad mental en el cerebro, debido, sobre todo, a los estudios realizados en el campo de la neurociencia.

beneficiaba a la fluidez del taller. Esto sería motivo de frustración para los organizadores (Copeland, 1996: 29).

Sin embargo, no todo sería negativo, ni mucho menos, en esta reunión. Sin ir más lejos es de destacar que el taller fue el escenario en el que tuvo lugar la prolífica propuesta de tres investigadores. Estos fueron Allen Newell (1927-1992), John Cliff Shaw (1922-1991) y Herbert Simon (1919-2001). Con estos tres investigadores retomamos la parte de la historia del JOHNNIAC que dejamos aparcada más arriba: la introducción en este del programa Lógico Teórico. La importancia de la introducción de dicho programa en el JOHNNIAC residía en la capacidad que adquiría la computadora de resolver problemas lógicos. En un principio el programa habría tenido otro cometido que no pudo conseguir, este era resolver problemas geométricos. Pero ello no fue un obstáculo para que Newell, Shaw y Simon decidieran intentar cambiar el propósito del programa. El propósito sin duda era intentar que la computadora razonara del mismo modo en el que lo hacemos los humanos. Obviamente la capacidad lógica es un rasgo caracterizador de los seres humanos y que una máquina pudiera llegar a hacerlo la acercaría inevitablemente. Y así fue. El programa Lógico Teórico dio con los resultados esperados. Por supuesto que hacía frente a problemas lógicos simples, y además podía resolver muchos de los planteados en la obra insigne de la lógica: *Principia Mathematica* de Russell y Whitehead. Pero eso no era todo. Algunas de las soluciones propuestas por el Lógico Teórico eran incluso más elegantes y cortas que las ofrecidas en la citada obra, despertando la admiración del mismísimo Bertrand Russell (Russell, Norvig, 2004: 20). Tal era el grado de convencimiento de los creadores del programa de que su programa permitía a la computadora razonar al modo humano que no dudaron en añadir el nombre Lógico Teórico al artículo que escribieron conjuntamente sobre las demostraciones del programa. No obstante, en la edición final de *The Journal of Symbolic Logic* el nombre del programa no aparecería.

A pesar de que los resultados obtenidos no causaban tanta impresión fuera de la comunidad encargada del desarrollo de la Inteligencia Artificial, ello no impidió que a partir del Lógico Teórico se viviera en dicha comunidad una etapa de grandes expectativas. Y no era para menos. Hasta el momento, el quehacer de las máquinas era bastante limitado, pero los nuevos programas y las nuevas computadoras lograban traspasar esos límites. Sólo un año después de que el Lógico teórico fuera presentado en Dartmouth (1957), Newell, Shaw y Simon crearon el sistema de resolución general de problemas (SRGP). Este sistema no sólo se ceñía a resolver problemas de lógica, sino que estaba preparado para hacer frente, en principio, a cualquier problema simbólico formal. Pero eso no era todo. El SRGP no operaba en sus razonamientos al modo en el que lo habían hecho los sistemas anteriores, sino que mostraba grandes semejanzas al modo de pensar de los humanos. Por ello, este programa llegaría a ser considerado por algunos como el primero que incorporaría el enfoque de «pensar como un ser humano» (Russell, Norvig, 2004: 21). Por supuesto que esto respondía al optimismo que vivían en esa época más que a resultados verdaderamente concluyentes, ya que el SRGP sólo hacía frente a problemas muy específicos y con una dificultad bastante asequible[32].

[32] Hablando en términos de las capacidades humanas.

En 1959, Herbert Gelernter (1929-2015) consiguió llegar allí donde el Lógico Teórico de Newell, Shaw y Simon no pudo. Gelernter creó un programa que podía resolver teoremas de geometría. Y no sólo podía resolverlos al modo en el que había sido programado para ello, sino que el mismo programa buscaba soluciones propias a los distintos teoremas. Es cierto que el programa no fue puesto a prueba con problemas geométricos de una dificultad extrema, pero es de destacar la capacidad que el programa tuvo de buscar más allá de aquello que le fue introducido.

La aportación de Gelernter a la Inteligencia Artificial tendría otra vertiente si cabe más importante que la de la máquina geométrica. Dos años antes de crear su programa (1957), Gelernter, junto al anteriormente citado Nathaniel Rochester, había estado involucrado en el proyecto de crear un lenguaje computacional. Dicho lenguaje finalmente fue el denominado FORTRAN. Tal fue el éxito de dicho lenguaje que se siguieron haciendo versiones de este, llegándose incluso a utilizar en la actualidad.

Otro de los lenguajes que surgieron en esa época que hoy en día sigue vigente es el creado por McCarthy, el conocido como LISP. Este lenguaje apareció en un año relevante para McCarthy en particular y para la Inteligencia Artificial en general, 1958. Fue en este año cuando McCarthy elaboró del *Generador de Consejos*, un programa que tendría la capacidad de solucionar problemas de carácter general (incluso problemas de la vida cotidiana humana). McCarthy estaba firmemente convencido de que la clave para entender el pensamiento humano pasaba por la lógica. Y esta convicción cobraría más fuerza cuando el filósofo-matemático John Alan Robinson (1930-2016) descubrió y añadió al *Generador de Consejos* el método de resolución, que consiste en un algoritmo que permite la demostración de teoremas de lógica de primer orden. Era el año 1965 y la Inteligencia Artificial aun gozaba de muy buena salud. Pero esto no sería así siempre.

Llegaron las vacas flacas y todo tenía un porqué. Los frutos de la Inteligencia Artificial eran innegables. Pero también era evidente que el progreso de la nueva inteligencia era cada vez más paulatino. Esto provocó irremediablemente que la actitud con respecto al alcance de los programas y las máquinas cambiara de manera notable. Las fuerzas económicas que se encargaban de invertir en investigación para el desarrollo de Inteligencia Artificial ya no veían con tan buenos ojos seguir empeñándose del mismo modo.

Algunos de los desencantos más significativos provenían precisamente de los mayores logros conseguidos. Los métodos de resolución de problemas prometían mucho porque su capacidad de acción podía extenderse hacia diferentes ámbitos (desde problemas cotidianos como ir al aeropuerto, como problemas muy específicos como lo es la estrategia militar). Pero desafortunadamente dejaban de ser útiles cuando se les añadía un número determinado de parámetros a tener en cuenta. Simplemente no eran capaces de afrontar los mismos problemas que los humanos.

Es de justicia añadir que la Inteligencia Artificial vio perjudicada su imagen no sólo porque en la práctica mostrara carencias, sino también porque los trabajos en esta época ponían el foco en las limitaciones que le pertenecían de por sí. En un trabajo de una autoridad dentro de la Inteligencia Artificial

como lo era Marvin Minsky, *Perceptrons*[33], destacaba el gran alcance de ciertas redes neuronales (perceptrones). Pero en dicha obra también se daba cuenta precisamente de las limitaciones inherentes en tales redes (Russell, Norvig, 2004: 26).

El camino de la inteligencia Artificial estaba marcado por una condición que parecía pasar desapercibida: si los programas y las computadoras se daban de frente con sus limitaciones era debido a que sus propósitos eran de carácter general. Una vez los investigadores decidieron volver a conseguir propósitos de carácter particular la Inteligencia Artificial experimentaría un resurgimiento.

El renacimiento de la Inteligencia Artificial se dio en el campo de la biología, de la mano de Edward Feigenbaum (1936), quien en 1965 se embarcó[34] en el proyecto de crear un programa que interpretara la estructura molecular. Dicho proyecto perduró durante 10 años y acabó con el resultado del programa conocido como DENDRAL, que daría grandes resultados. Basado en el DENDRAL, aunque sólo al principio, estaría el programa MYCIN. Este programa fue desarrollado por Edward Shortliffe (1947) y también contó con la colaboración del mismo Feigenbaum. Su propósito era diagnosticar enfermedades a través de un procedimiento muy similar a como lo hacían los humanos. De hecho, sus diagnósticos no diferían en exceso del dado por expertos y era superior al de los estudiantes de medicina (Russell, Norvig, 2004: 27). También se crearon programas fructíferos en el ámbito de la comprensión del lenguaje. Parecía, efectivamente, que la Inteligencia Artificial había vuelto con paso firme y de manera perentoria.

Ya en la década de los 1980, concretamente a partir de 1986, la tendencia de la investigación de la Inteligencia Artificial se volcó hacia una vuelta al estudio de las redes neuronales al modo de McCulloch y Pitts. Si bien el «abandono» de este estudio no se dio en todos los campos (en física siguió siendo importante), lo cierto es que en el terreno de la Inteligencia Artificial había sido apartado desde finales de los años 1960. Este hecho, junto a la renovada fuerza que está adquiriendo la Inteligencia Artificial, ha provocado que los investigadores se propusieran de nuevo elaborar programas con propósitos de carácter general, aunque teniendo en cuenta los errores del pasado. Y los proyectos de última generación en este campo obtienen grandes resultados, pero la pregunta aún sigue en el aire…

3.3. ¿Hasta dónde es correcto seguir afirmando una igualdad o una superioridad de las máquinas con respecto a los seres humanos?

Una vez hecho este repaso y ver a grandes rasgos cuál es la situación actual de la Inteligencia Artificial, ahora volvemos a la cuestión inicial de este apartado: en qué estado se encuentra la Inteligencia Artificial. ¿Ha igualado al ser humano? ¿Lo ha superado?

[33] Dicha obra la escribió junto al matemático Seymour Papert (1928-2016).
[34] Junto a Bruce G. Buchanan (1940) y Joshua Lederberg (1925-2008).

Desde una perspectiva neutral, es decir, fuera de los puntos de vista A, B, C y D, la respuesta es que no se ha igualado al ser humano, y mucho menos lo ha superado. Esta perspectiva nos permite determinar que el asunto es tremendamente complejo como para intentar dar una respuesta definitiva o simplista del asunto en términos de sí o no. Pero una vez nos situamos en los puntos de vista actores en el debate las respuestas son necesariamente diferentes. No es que esté tildando de simplistas a estos [filosóficamente] legítimos puntos de vista. Mi observación tiene más que ver con el hecho de creer que es prácticamente inevitable caer en reduccionismos a la hora de tomar parte en un debate. Esto, por otra parte, no tiene que ser necesariamente malo. De hecho, la manera de obtener argumentos (más allá de la obviedad de la complejidad del asunto) es tomar parte en la discusión. En los apartados que siguen veremos cómo las dos posturas de la Inteligencia Artificial toman parte en el debate y cómo Penrose intenta rebatirlas desde su punto de vista.

4. La Inteligencia Artificial *fuerte* y un muro llamado Penrose

4.1. La postura de la IA *fuerte*

Sin lugar a dudas, la Inteligencia Artificial vino para quedarse en los debates de filosofía. Aunque esto, desde luego, no ha sido del agrado de todos. No pocas veces habremos podido escuchar que seguir debatiendo acerca de la posibilidad de una inteligencia artificial resulta, cuanto menos, ignominioso para el quehacer filosófico. Desde su origen, la filosofía ha tratado los temas más fundamentales del ser humano. ¿Cómo es posible que ahora se pierda el tiempo en pensar en cachivaches? Esto sólo puede tener como resultado el acabar dando la razón a aquellos que tildan a la filosofía de una actividad prácticamente inútil. Pero, ¿realmente es así? ¿Barajar un escenario en el que las máquinas lleguen a pensar como los seres humanos es simplemente una quimera superflua? No hay que negar lo fácil que resulta dejar volar la imaginación ante una propuesta de tal naturaleza, pero tampoco es conveniente creer que el tema carece de profundidad. En mi opinión, la clave se encuentra en la siguiente pregunta: cuando nos preguntamos acerca de si una máquina puede pensar tal y como lo hacemos los seres humanos, ¿estamos preguntado por la naturaleza del pensamiento de la máquina o, en cambio, por la del ser humano? La respuesta es menos obvia de lo que parece. Aquellos que se niegan a ver el asunto de la posibilidad de una inteligencia artificial como algo que incumba a la filosofía deberían responder que se pregunta por el pensamiento de la máquina. Y si esa es la respuesta reconozco que debo darles la razón. Ese tema es superfluo y de ningún interés filosófico. No obstante, esto no les libraría de estar en un error. La filosofía no se ha preguntado acerca de la naturaleza del pensamiento de la máquina. Si la cuestión de la inteligencia artificial ha ocupado un lugar importante en los debates filosóficos es porque estos han estado centrados en saber cuál es en última instancia la naturaleza del pensamiento humano. Querer implantar en una máquina una inteligencia equivalente a la humana implica el conocimiento de la segunda, sino ¿cómo es posible hacerlo o tan siquiera especular con ello? Simplemente no se puede.

Una vez se tiene claro este punto llega lo que realmente complica las cosas: saber qué constituye la inteligencia humana. Muchas propuestas han aparecido a lo largo de la historia de la filosofía en general, y no menos han surgido en el debate de la Inteligencia Artificial en particular. Lo que sigue en el resto de este punto será la respuesta de los partidarios de la IA *fuerte* con respecto a esta cuestión.

¿Qué es aquello de lo que está constituida la inteligencia humana? ¿Realmente existe algo que permita explicarla? Y en el caso de que lo hubiese, ¿es susceptible de ser conocida por el ser humano? Los defensores de la IA *fuerte* ofrecen respuestas a estas preguntas. Con respecto a la primera y la segunda –que están íntimamente relacionadas-, existen varias respuestas. La más extendida, y que es aquella contra la que Penrose dirige la mayor parte de su crítica, defiende que la inteligencia humana consiste básicamente en un algoritmo. La pregunta pertinente ahora es: ¿qué es un algoritmo[35]? Para seguir situados en el debate que nos concierne, veamos la definición de algoritmo que se da dentro del ámbito de la informática, más concretamente en el plano de la programación. En este terreno se entiende algoritmo como *la descripción de una sucesión finita de acciones que permite transformar el entorno del estado inicial dado en el final deseado* (Álvarez et al., 2005: 6). Tenemos, pues, que el algoritmo es el conjunto de reglas que posibilita la ejecución del programa. Una vez vista esta definición, ¿podemos ver en qué se diferencian algoritmo y computación (vista en §1.2 y §3.2)? Si la respuesta que encuentra la lectora (o lector) es que no, ello no debería preocuparle. En este trabajo entenderemos los términos algoritmo y computación como sinónimos, tal y como los entiende Penrose (Penrose, 2012: 32).

Entonces, ¿la inteligencia humana está basada en la simple acción de un algoritmo? La respuesta de la IA *fuerte* es que sí. Pero dicha respuesta guarda más complejidad de lo que parece en principio. Que los partidarios de este modelo defiendan tal condición no les exime de tener que recular a la hora de dar explicaciones acerca de dicho algoritmo (o algoritmos, porque reducirlo a uno o varios es indiferente para esta postura). El algoritmo (o algoritmos) que constituye(n) la inteligencia humana no está disponible para los seres humanos. El motivo por el que no podemos conocerlo reside en la sutileza de la computación que lleva a cabo la inteligencia. Los algoritmos que conocemos (y ni tan siquiera la unión de muchos de ellos) no permite(n) explicar la inteligencia. No obstante, la IA *fuerte* no basa su defensa en una simple especulación. Existen algoritmos que permiten, al menos, reproducir facetas de la inteligencia humana. Por ello, piensan, es cuestión de tiempo que se acabe desarrollando el algoritmo que dé con la naturaleza última de nuestra inteligencia.

Estas son las ideas principales que Turing[36] puso encima de la mesa hace ya más de medio siglo. Ahora bien, es preciso tener en cuenta algunos detalles acerca de tales ideas. Por ejemplo, con respecto a esta última que se

[35] Anteriormente (§1.2 y §3.2) se ha hecho mención a este concepto, pero se hacía de tal manera que no requería una explicación. Como de aquí en adelante esta noción tendrá una importancia central lo mejor es ofrecer dicha descripción.

[36] A pesar de que esta es una consideración aceptada extendidamente (y aquí se entenderá de tal forma) ello no significa que sea consensuada de forma plena. Para un ejemplo paradigmático de este tipo de postura –digámoslo así- alternativa véase Juliet Floyd (2017) en *Philosophical Explorations of the Legacy of Alan Turing*.

caracteriza por el optimismo de encontrar la pieza definitiva que permita dar la respuesta final al debate. Indudablemente la IA *fuerte* confía en dar con dicha respuesta, pero ello no es óbice para que la cautela también entre en escena. Por supuesto que los partidarios creen (y quieren) que el deseado momento llegue más temprano que tarde, aunque saben de las limitaciones que tiene una meta tan ambiciosa. Como vimos en el repaso histórico, hubo un momento de –digámoslo así- optimismo descontrolado por parte de los defensores de la IA. Por otro lado, no era para menos. Los buenos resultados se daban con la presteza esperada (en algunos casos incluso se superaban) y además se daban de manera incesante. Tal actitud, empero, no constituye un rasgo característico de la postura de la IA *fuerte*, al menos de la defendida por Turing. Su postura se caracteriza, al contrario de lo que se suele pensar, más bien, por llamar a la cautela[37]. Es cierto que Turing formuló algunas predicciones y que estas no se cumplieron en su mayor medida. Por otra parte, la finalidad de tales predicciones no era pronosticar el futuro de manera exacta e inequívoca. Su intención tenía que ver más con el apoyo a la idea de que las máquinas llegarán a pensar como los seres humanos, ya fuera más tarde o más temprano de lo que él mismo había dicho.

Como consecuencia de no tener en cuenta el «error de cálculo» de Turing los partidarios de la IA *fuerte* siguieron la senda marcada y el cometido encomendado sigue vigente incluso en la actualidad. No pocos son los seguidores de las ideas principales de Turing a día de hoy. No sólo en el mundo de la informática. En el ámbito filosofía también cuenta con una serie de adeptos de una seriedad y renombre intelectual considerable. Daniel Dennett, filósofo contemporáneo y de gran influencia, hace una defensa de esta postura desde la perspectiva de la filosofía de la ciencia. A lo largo de su trabajo es común encontrarse con discursos con un claro carácter optimista con respecto al progreso de la Inteligencia Artificial, tales como, por ejemplo, estos en *Dulces sueños*:

> […] Después de todo, ya contamos con excelentes explicaciones mecanicistas del metabolismo, el crecimiento, la autorreparación y la reproducción, que hasta no hace mucho también parecían procesos demasiado complejos para el lenguaje existente.

[37] Jack Copeland, experto en la obra de Turing, hace hincapié sobre este asunto en *The Turing guide* (2017). En uno de sus capítulos en dicha obra (concretamente el 25, titulado "Intelligent Machinery") advierte de las malinterpretaciones que ha sufrido la defensa de Turing con motivo de sucesos puntuales. En concreto habla de una noticia de 2014, relacionada con un programa llamado Eugene Goostman. Según las fuentes (fuentes con gran impacto de divulgación, como son *Washington Post* y la BBC), este programa consiguió superar, haciéndose pasar por un niño ruso de 13 años, el test de Turing. Copeland destaca la poca fidelidad que guardaba lo que se decía en tal noticia con respecto a lo que predijo Turing:

Obviamente, Turing no podría haber considerado que engañar al 30% de los jueces durante una serie de conversaciones de 5 minutos equivalía a pasar la prueba, ya que predijo que esta tasa de éxito del 30% se lograría «en unos cincuenta años», pero también dijo que pasarían «al menos 100 años» antes de que una computadora *pasara* su prueba. La afirmación errónea […] de que Eugene Goostman pasó la prueba de Turing se basa en haber relacionado la predicción del 30% de Turing con una especificación de lo que cuenta como pasar la prueba. Simplemente ignoraron la cuidadosa especificación de Turing, en términos de otro juego en el que un hombre imita las respuestas de una mujer, de lo que en realidad se consideraría como pasar la prueba. (Copeland et al., 2017: 272-273).

> Si adoptamos esta perspectiva optimista, la conciencia[38] es algo maravilloso, pero no tan maravilloso; es decir, no tan maravilloso que no pueda explicarse con los mismos conceptos y teorías que han funcionado a la perfección para las demás áreas de la biología (Dennett, 2006: 20).

> […] Tenemos la seguridad de que una explicación naturalista y mecanicista de la consciencia no sólo es posible sino que está haciéndose realidad a toda velocidad. Lo único que necesitamos es mucho trabajo, del estilo del que se hizo en biología a lo largo del siglo XX y en ciencia cognitiva en la segunda mitad (Dennett, 2006: 22).

A pesar de todo, Dennett en concreto no es un gran representante de la postura que equipara la consciencia y la inteligencia humana con la acción algorítmica o computacional[39]. Como hemos podido ver, la pertenencia del filósofo estadounidense a las ideas de la IA *fuerte* tiene que ver con dos rasgos esenciales de esta postura. En primer lugar, el optimismo de dar una respuesta afirmativa ante la cuestión de si es posible llegar a conocer aquello que hace posible la inteligencia y la consciencia humana; y en segundo lugar, la apuesta por la ciencia para aportar las pruebas de la declaración anterior.

La elección de las citas en concreto deben su explicación a que ponen el foco en un tema que es ubicuo en la perspectiva de aquellos que defienden el modelo A y que vimos más arriba: la curiosa relación de la IA *fuerte* con el materialismo. Dennett habla concretamente de la necesidad del mecanicismo, el cual sabemos que guarda una relación íntima con el materialismo. Esta, sin embargo, no parece ser precisamente la relación a la que se refiera el filósofo estadounidense, al menos en grado sumo. Dennett admite el carácter físico de la consciencia. ¿Esto lo aleja de la IA *fuerte*? No necesariamente, ya que su postura más que responder al materialismo más puro, parece estar comprometida con el fisicalismo[40]:

> Muchas personas opinan que la conciencia es un misterio, el espectáculo de magia más maravilloso que se pueda imaginar, una serie interminable de efectos especiales que desafían toda explicación racional. Para mí, están equivocadas: la conciencia es un fenómeno físico, biológico, como el metabolismo, la reproducción o la autorreparación, de un ingenio exquisito en su funcionamiento, pero no milagroso, ni siquiera misterioso.
> Parte de la dificultad para explicar la conciencia radica en la existencia de fuerzas poderosas que nos hacen creer que es más maravillosa de lo que en realidad es […] en ese sentido la conciencia se parece a la magia que se hace sobre un *escenario*; un conjunto de fenómenos que explotan nuestra credulidad y hasta nuestro deseo de que nos engañen, nos engatusen, nos dejen con la boca abierta. Explicar un espectáculo de magia es en cierto sentido una tarea ingrata; a quien revela los secretos de un truco se lo mira con malos ojos, se lo considera un aguafiestas. A veces, tengo la impresión de que mis intentos de explicar aspectos de la conciencia

[38] He decidido conservar la palabra conciencia en lugar de escribir consciencia con la intención de no alterar la traducción que he utilizado. Pero vale tener en cuenta que cuando haga mención a la conciencia en palabras de Dennett debemos entenderla al modo de Penrose, este es, como consciencia.

[39] Hasta donde yo sé (que, por otro lado, no dejo de ser una fuente poco fiable en este aspecto), no se manifiesta en contra explícitamente. No obstante, sí que puede detectarse cierto escepticismo, como, por ejemplo, en *La conciencia explicada* (1995: 238).

[40] Dennett no es partidario del fisicalismo que defiende que todo puede tener una explicación a través de la física. El filósofo estadounidense, como vimos en la primera cita, está convencido de que las respuestas pueden llegar a través, sobre todo, aunque no exclusivamente, de la biología.

> generan la misma resistencia. ¿No es más lindo que nos dejen regodearnos en el mágico misterio del asunto? O dicen esto: si logramos explicar la conciencia, los seres humanos quedaremos disminuidos, reducidos a meros robots proteicos, meros *objetos* (Dennett, 2006: 75).

Para Dennett, entonces, el velo que cubre la consciencia puede apartarse *físicamente* y es a la ciencia a la que otorga el privilegio de quitarlo, permitiéndonos salir de la ignorancia en la que estamos sumidos. Podemos suponer sin temor a equivocarnos, por tanto, que su pertenencia a la IA *fuerte* es clara.

Dejando a un lado, aunque no del todo, el punto de vista particular de Dennett[41], en el punto que sigue volveremos a retomar el problema que suscita para la IA *fuerte* el materialismo.

4.2. IA *fuerte* y su complicada relación con el materialismo

Que la inteligencia se intente definir en términos de materialidad o de otra cosa que esté más allá de ella es volver, intencionadamente o no, al problema del dualismo. ¿Tan imposible nos resulta salir de dicha discusión que esta, incluso, invade los nuevos términos que hoy en día usamos? Efectivamente, eso es lo que parece. En el ámbito de la Inteligencia Artificial y la ciencia cognitiva, sin duda, este problema tiene una vigencia indiscutible. Un ejemplo lo encontramos en el tema que trata la célebre ecuación hardware/software = cerebro/mente. Para saber hasta qué punto está relacionado este tema con el problema del dualismo veamos algunos de los conceptos principales.

Entre los conceptos básicos dentro del campo de la computación están los denominados *hardware* y *software*. Cuando hablamos de *hardware* nos referimos a aquello que compone físicamente a la computadora (tales como los cables, chips, placas, etc.). Por su parte, el *software* se entiende como aquello que permite realizar determinadas tareas a tal dispositivo, es decir, su programa. El programa consiste en la introducción y ejecución de instrucciones (conocidas como *inputs*), las cuales producen respuestas específicas (conocidas como *outputs*). La pregunta pertinente ahora es: ¿qué es lo importante para un ordenador, aquello que lo compone físicamente o el programa? El programa marca su modo de actuar. Una máquina repleta de chips funcionales, con una interconexión de cables impecable y con todo el aparataje a disposición plena, no ofrece respuesta alguna si no hay un programa que le dicte cómo debe actuar. Pero un programa perfectamente acabado sin composición material ¡no garantiza que las respuestas para las que está dispuesto tengan lugar! La ecuación es fácilmente deducible y la relación con el debate del dualismo igual de sencilla.

La parte que le corresponde al punto de vista A en este asunto también es evidente: el *software* es indispensable para la consciencia de la computadora.

¿Qué sucede, entonces, con el *hardware*? ¿Acaso se puede prescindir de ello? Siendo fidedignos a lo que defienden los partidarios de la IA *fuerte*, la respuesta debe ser necesariamente que no. Sin embargo, también hay que

[41] Sobre todo porque constituye una visión actual del punto de vista A.

hacerse cargo de que estos se apresurarán en aclarar que la importancia del *hardware* con respecto al *software* es mucho menor. No se renuncia a la materialidad, pero la reconocen como secundaria con respecto a la consciencia.

El tema parece estancarse una y otra vez en el mismo lugar. Pero, ¿de veras seguimos dando vueltas sobre los mismos puntos exactamente? En mi opinión, esta no es la situación en las que nos encontramos. Pienso que los debates de la ciencia cognitiva y las teorías sobre la computación nos permiten llegar a terrenos sin explorar, los cuales nos han abierto nuevos caminos por recorrer. Y no se trata de un optimismo infundado. Las aportaciones de Daniel Dennett, Douglas Hofstadter, Jack Copeland, John Searle, Ned Block, el matrimonio Churchland o el mismo Penrose (por nombrar sólo unos pocos) están lejos de ser estudios superfluos o vanos. Asuntos de una importancia indiscutible a nivel filosófico se siguen debatiendo y rebatiendo.

Un ejemplo, y con ello seguimos con el problema del dualismo tratado desde la perspectiva de A, es la explicación de la individualidad que ofrece este modelo. Penrose se percata de la problematicidad del asunto y plantea un supuesto en el que la IA *fuerte*, según nuestro autor, vería complicado dar cuenta de él. Comienza explicando la postura de A:

> Aceptemos que la individualidad de una persona no tiene nada que ver con la individualidad que pudiéramos atribuir a sus constituyentes biológicos. Más bien está relacionada con la configuración, en cierto sentido, de dichos constituyentes, digamos la configuración espacial o espacio-temporal […] Pero los defensores de la IA *fuerte* van mucho más lejos. Ellos dirán que si la información contenida en tal configuración puede ser transferida a otra forma desde la cual puede ser recuperada, entonces la individualidad de la persona debe permanecer intacta […]. Incluso parecen querer decir que la consciencia de una persona persistirá aunque su «información» esté en esta otra forma. Desde este punto de vista, una «consciencia humana» debe considerarse, en realidad, como un elemento de software, y su manifestación particular como ser humano material debe considerarse como la ejecución de ese software por el hardware de su cerebro y su cuerpo (Penrose, 1991: 51-52).

El supuesto de Penrose consiste en imaginarse un caso en el que la teletransportación fuera posible. Tenemos una máquina teletransportadora. Supongamos que esta se compone de manera *convencional* (con respecto a las estructuras en los relatos de ficción): dos plataformas y una compleja (normalmente indescriptible) estructura maquinaria. Su funcionamiento también sería el *tradicional*: cuando se activa la máquina, un objeto colocado en una de las plataformas pasaría a la otra de manera instantánea. ¿Cuál es el proceso por el que pasa el cuerpo teletransportado? La máquina explora de arriba abajo dicho cuerpo, registrando con todo detalle la localización exacta y la especificación completa de cada átomo y cada electrón de su cuerpo.

En opinión de Penrose, resulta llamativo que cuando se ha especulado acerca de las máquinas teletransportadoras la discusión haya girado en torno a la posibilidad o imposibilidad de su creación. Para nuestro autor, lo verdaderamente importante es tener en cuenta las implicaciones que una máquina como esta tendría en un individuo (Penrose, 1991: 53). El porqué del planteamiento de Penrose es palmario: quiere saber qué papel juega en este asunto lo material.

Al hacer que un individuo pase de una plataforma a la otra, ¿estamos teletransportando a ese individuo o estamos realizando un duplicado del mismo? En el caso de estar haciendo un duplicado, ¿podemos seguir hablando, así, de que el individuo que aparece en la plataforma de destino *es* el mismo que el que *se encontraba* en la plataforma de partida?

La respuesta no es sencilla. De nuevo hay que hacer el ejercicio de intentar dar una respuesta que case con lo defendido por los partidarios de la IA *fuerte*. Estos responderán que lo importante es teletransportar el programa (*software*) de manera exacta para que el individuo siga siendo el mismo. El cuerpo podría ser duplicado o lo que se quiera, ya que el *software* se encargará de que la consciencia siga siendo la misma. Hemos visto que la IA *fuerte* intenta no abandonar la parte fisicalista, pero ante un supuesto como este parece que en última instancia se ve forzada a hacerlo.

Para Penrose una respuesta de este calibre no se corresponde con la realidad. Suponiendo que podemos trasladar una consciencia a otro cuerpo, incluso a cuerpo no biológico, estamos pasando por alto aquello que también nos constituye como seres humanos:

> Veamos ahora qué relación guarda el punto de vista de la IA *fuerte* con la cuestión de la teleportación. Supongamos que en algún lugar entre los dos planetas hay una estación repetidora en la que se almacena temporalmente la información antes de ser retransmitida a su destino final. Por conveniencia, esta información no es almacenada en forma humana sino en algún dispositivo magnético o electrónico. ¿Estaría presente la «consciencia» del viajero en este dispositivo? Los defensores de la IA *fuerte* tendrán que hacernos creer que así debe ser. Después de todo, dicen ellos, cualquier pregunta que decidiésemos plantear al viajero podría ser respondida en principio por el dispositivo, estableciendo «simplemente» una simulación apropiada de la actividad de su cerebro. El dispositivo contendría toda la información necesaria; el resto sería sólo un asunto de computación. Puesto que el dispositivo respondería a todas las preguntas exactamente como si fuera el viajero, entonces (por la prueba de Turing) sería el viajero.
>
> Esto nos lleva de nuevo a la concepción de la IA *fuerte* según la cual el hardware real no es importante en los fenómenos mentales. Esta opinión me parece injustificada. Se basa en la presunción de que el cerebro (o la mente) es, en efecto, una computadora digital. Supone que cuando pensamos no está en juego ningún fenómeno físico concreto que requiera estructuras físicas concretas (biológicas, químicas) como las que tienen realmente los cerebros (Penrose, 1991: 54-55).

Sin embargo, la respuesta de la IA *fuerte* a este supuesto no es unidimensional. Vimos más arriba que un férreo defensor de este punto de vista, como lo es Dennett, otorga al componente físico una importancia mucho mayor de la que se ha visto en el supuesto que Penrose trae a colación. El motivo de esta asimetría es que cuando Penrose habla de IA *fuerte* se refiere exclusivamente a aquellos que siguen las pautas del pensamiento de Turing. Pero el pensamiento de Turing tiene más de medio siglo, ¿se queda la crítica de Penrose anquilosada en un debate que ya está obsoleto? Precisamente esta es una de las acusaciones más recurrentes de los críticos de los planteamientos de Penrose. Pero sobre ello volveremos más adelante. Ahora nuestra atención se dirigirá a otras ideas que Turing tenía acerca de la consciencia de las computadoras para, así, acabar de ilustrar el modelo A.

4.3. Las convicciones de Turing

En su célebre artículo de 1950, Turing puso encima de la mesa la posibilidad de crear una máquina que pensara. Una de las maneras de poder dar respuesta a este planteamiento sería a través de lo que él denominó «juego de la imitación» y que posteriormente pasaría a conocerse como «test de Turing». Aquella máquina que superara dicho juego podría ser considerada como «máquina que piensa». Ahora bien, ¿cuáles son las posibilidades *reales* de que una máquina llegue a superar satisfactoriamente dicho juego? El mismo Turing creía que dichas posibilidades podrían ser una realidad con el paso de los años. Es cierto que tenía la certeza de que tal escenario tardaría en darse, pero su convicción sobre las «máquinas pensadoras» también era inamovible.

Estas son las características que generalmente suelen resaltarse del pensamiento de Alan Turing. Lo interesante es intentar comprender por qué el matemático inglés llegó a este tipo de conclusiones. Obviamente jamás podremos llegar al fondo de sus justificaciones. A lo que podemos aspirar es a conocer parte de su biografía y contexto para ver de qué manera unos factores u otros influyeron en su pensamiento[42].

Definitivamente la manera en la que Turing vivió la Segunda Guerra Mundial pasó factura en el desarrollo de sus ideas con respecto a este tema. Bien conocida es la labor de Turing descifrando códigos de estrategia del bando nazi. Pero no tan ínclita es la influencia que este hecho ejerció en el matemático inglés.

La Alemania nazi tenía a su disposición una máquina, *Enigma*, que cifraba sus códigos de estrategia[43]. *Enigma* era un verdadero quebradero de cabeza para el bando Aliado. Tanto era así que varios proyectos fueron llevados a cabo con el único fin de intentar descifrar los códigos formulados por esta máquina. Dos fueron los proyectos más relevantes. Uno lo desarrolló un equipo polaco y fue denominado *bomba*. Y poco más tarde un equipo británico dio paso al proyecto que se llamaría *bombe* (en clara alusión al seguimiento del trabajo realizado por el equipo polaco). Ambos proyectos consistían en la creación de una máquina que permitiera dar con el modo de cifrar códigos de *Enigma*. En definitiva, todo seguía siendo un pulso por el desarrollo tecnológico y este lo estaba ganando Alemania.

Si bien el proyecto *bomba* pareció encauzar el camino correcto hacia el desciframiento de *Enigma*, llegando a conocer mensajes del bando alemán desde principios de los años 30 (Batey, 2017: 98), la realidad fue que no conseguían dar con la clave. Ello estaría reservado para el equipo británico. La participación de personalidades como Turing y Dilly Knox (1884-1943) fueron decisivas para que el proyecto *bombe* pudiera llegar adonde no pudieron llegar los polacos.

[42] Este punto no estará dedicado a hacer un repaso biográfico o histórico, sino que destacaré hechos puntuales que, pienso, influyeron de manera definitiva en los planteamientos de Turing.

[43] Aunque también hacía las veces de descifrador de códigos.

Knox ya era conocido por su papel en la Primera Guerra Mundial como descifrador de códigos. Habiendo recalado del mundo de la papirología, este lingüista consiguió divisar el sistema de cableado de *Enigma* a través de una versión comercial de esta máquina. Esto, sin duda, allanó el camino significativamente. Sin embargo, el hito que marcaría un antes y un después en el proyecto *bombe* fue el refinamiento en el diseño de la máquina por parte de Turing. El matemático inglés cayó en la cuenta de que si *Enigma* era tan compleja de descifrar era debido a que probablemente esta fuese sometida a ajustes de manera regular. Y efectivamente así era. La solución pasaba, entonces, por crear un conjunto de máquinas similares a *Enigma* pudiendo, así, saber qué códigos creaban las máquinas según los diferentes tipos de ajustes que pudieran hacérsele[44].

¿En qué forma el proyecto *bombe* tuvo una influencia perentoria en las ideas de Turing sobre la posibilidad de máquinas pensantes? El modo en el que el conjunto de máquinas obtuvo el resultado que se esperaba provocó que Turing viera algo más profundo en ello. Copeland defiende que el sistema de búsqueda guiada (*guided search*) fue el detonante decisivo:

> Gracias a *bombe*, Turing vislumbró la posibilidad de lograr una inteligencia artificial de carácter general mediante la búsqueda guiada. Esta idea le fascinó por el resto de su vida. Pronto estaba hablando con entusiasmo con sus compañeros descifradores de códigos sobre el uso de este nuevo concepto de búsqueda guiada para mecanizar los procesos de pensamiento involucrados en el juego de ajedrez […] También quería mecanizar el proceso de aprendizaje en sí. En Bletchley Park, hizo circular un mecanografiado sobre inteligencia artificial -ahora perdido-, siendo este sin duda el primer trabajo en el campo de la IA (Copeland et al., 2017: 266).

Turing vio en la forma de actuar de bombe un destello de una forma de pensamiento. Que dicha forma de pensamiento tuviera la dependencia que tenía de la actividad humana no tendría que tener importancia. Los mismos seres humanos dependemos de otros seres humanos para aprender a pensar. En la cita anterior se menciona algo importante relacionado con este asunto y que da lugar a otra de las características del pensamiento del matemático inglés. A partir de lo descubierto en el proyecto *bombe*, Turing dará valor al aprendizaje de las máquinas, intentado de dar con el método adecuado para que estas lleguen a pensar.

Tal método de aprendizaje no estaría basado solamente en hacerlo de una manera mecánica. Las máquinas descifradoras de códigos mostraron un tipo remoto de pensamiento y ello era debido a la acción humana sobre ellos. Aquello que permitía que las máquinas mostrasen un tipo de pensamiento era que estas de alguna forma aprendían de lo que les era introducido por los humanos. El método de aprendizaje, por tanto, sería más semejante al humano de lo que en un principio pudiera parecer:

[44] Un hecho interesante es que se cree que si el bando nazi hubiera sido más meticuloso a la hora de realizar los ajustes, la tarea de descifrar a *Enigma* habría sido aún más engorrosa: […] Otra de las ironías es que los alemanes podrían haber hecho pequeñas modificaciones a la máquina *Enigma* que hubieran provocado que la ruptura de los mensajes fuera mucho más difícil. Por ejemplo, simplemente colocando las muescas de rotación en el mismo lugar en todas las ruedas habría socavado muchos de los métodos manuales de Hut 8 para resolver la configuración diaria (Greenberg, 2017: 95).

> Si estamos tratando de producir una máquina inteligente, y estamos siguiendo el modelo humano lo más cerca posible, deberíamos comenzar con una máquina con muy poca capacidad para llevar a cabo operaciones elaboradas o para reaccionar de manera disciplinada a las órdenes […] Luego, al aplicar la interferencia apropiada, imitando la educación, debemos esperar modificar la máquina hasta que se pueda confiar en ella para producir reacciones definitivas a ciertas órdenes. Este sería el comienzo del proceso (Turing, cit. por Proudfoot, 2017a: 315).

¿Y cómo aprendemos los seres humanos? ¿No es acaso cierto que los métodos de aprendizaje humano también son diversos? Sabemos que existen métodos que potencian la memoria, otros el cálculo, otros el pensamiento abstracto, otros el lenguaje… ¿Cuál de ellos es el correcto? Más que preguntarse por la búsqueda del método adecuado, Turing pensaba en el momento en el que los humanos tienen las capacidades más propicias para el aprendizaje. ¿Y cuál es ese? Pues la infancia. Los niños aprenden paso a paso y su aprendizaje se basa, en mayor medida, en la experiencia. Para Turing, por tanto, es imprescindible que una máquina aprenda a partir de la experiencia y progresivamente:

> […] (Deberíamos) comenzar desde una máquina relativamente simple y, al someterla a un rango adecuado de «experiencia», transformarla en una más elaborada y capaz de lidiar con un rango mucho mayor de contingencias […] Tal como lo veo, este proceso educativo sería en la práctica esencial para la producción de una máquina razonablemente inteligente en un espacio de tiempo razonablemente corto. La analogía humana por sí sola sugiere esto (Turing, cit. por Proudfoot, 2017a: 316).

Turing denominará a este proyecto como «niño máquina» (*child machine*) y constituirá uno de los pilares en sus convicciones acerca de las «máquinas pensantes». Por otro lado, ello no impediría que también fuera fruto de constantes frustraciones (Proudfoot, 2017a: 319). La meta no era sencilla, por mucho que la manera de aprender de los niños, en principio, pueda resultar más simple que la de los adultos. Finalmente concluiría que muy probablemente no se pudiera educar a las máquinas exactamente como a los niños «normales» (Proudfoot, 2017a: 319) y ello era debido a otro de los puntos importantes de su pensamiento: la creación de máquinas con aspecto humano. Más arriba vimos que el punto de vista A, en un momento dado, no encuentra necesario el aspecto físico en relación a la inteligencia y la consciencia. El mismo Turing fantaseaba con la idea de intentar saber qué podría hacer un cerebro sin cuerpo (Proudfoot, 2017: 319), pero ello no era óbice para que tuviese en cuenta la parte física. De hecho, el matemático inglés también sería pionero en la robótica:

> Una razón muy positiva para creer en la posibilidad de hacer maquinarias de pensamiento es el hecho de que es posible hacer maquinarias que imitan cualquier parte pequeña de un hombre […] La forma de establecer nuestra tarea de construir una «máquina de pensar» sería tomar a un hombre como un todo y tratar de reemplazar todas las partes de él por la maquinaria. Este reemplazo incluiría cámaras de televisión, micrófonos, altavoces, ruedas y «servomecanismos de manejo» (*handling servo-mechanism*), así como algún tipo de «cerebro electrónico» (*electronic brain*). Esto, por supuesto, sería una tarea tremenda. El objeto producido por las técnicas actuales sería de un tamaño inmenso, incluso si la parte del «cerebro» fuera estacionaria y controlara el cuerpo desde la distancia. Para que la máquina tuviera la oportunidad de descubrir las cosas por sí misma, se le debería

> permitir vagar libremente por el campo, pero el peligro para el ciudadano común sería grave (Turing cit. por Proudfoot, 2017a: 319-320).

El factor físico no es indiferente, pero llevar a cabo una equiparación total entre las máquinas y los seres humanos resulta una tarea de una dificultad incuestionable. Por otro lado, dicha dificultad no prueba que la posibilidad de llegar a la meta sea inalcanzable. El paso importante es intentar lograr que la máquina primero piense inteligentemente sin necesidad de cuerpo. Una vez conseguido llegar a dicho punto, sin duda, el camino se allanará considerablemente. Pero si no tiene cuerpo, no siendo equiparable al ser humano, ¿cómo podemos saber si piensa inteligentemente? Vimos arriba que Turing proponía el juego de la imitación. Si bien este juego fue propuesto por Turing para aclarar cuál era su postura, lo cierto es que ha sido objeto de no pocas malinterpretaciones. Una de las más extendidas y de la que el mismo Turing dio cuenta fue aquella que le relacionaba con el conductismo. A pesar de que el matemático inglés se pronunciara en contra ello no impidió que posteriormente se le siguiera considerando como tal[45].

No obstante, actualmente existen quienes intentan salvar a Turing del conductismo. Diane Proudfoot, por ejemplo, expone una triple razón que evidencia la incompatibilidad del pensamiento de Turing con dicha corriente. La primera de ellas tiene que ver, precisamente, con aquello que el mismo Turing postulaba en contra del conductismo. El conductista basa sus conclusiones a partir del comportamiento observado de un sujeto. ¿Y no es esto lo que se hace cuando se practica el juego de la imitación? Pues depende de la concepción que se tenga de la inteligencia. Turing defendía que la inteligencia es más emocional que matemática (Proudfoot, 2017b: 303). El conductismo, por su parte, entiende que la inteligencia es más mecánica (o matemática): ante cierto estímulo sólo existen ciertas respuestas limitadas. Por lo tanto, concebir como equivalentes estas dos perspectivas es claramente erróneo.

Turing quiso desmarcarse del conductismo de manera intencionada, ya que esta corriente estaba en alza en los años 1950. Hacer un paralelismo entre lo que decía el matemático inglés con su juego de la imitación y lo defendido por los partidarios del conductismo no era descabellado. Otro asunto es, como hemos visto, que fuese erróneo.

La segunda de las razones está relacionada con la concepción equivocada de creer que con el juego de la imitación se está probando el comportamiento de la máquina. Lo que se evaluaría realmente con el juego es la reacción del observador. Turing decía que si una máquina lograba engañar a un interrogador más que un hombre en un juego de imitar las respuestas de una mujer, dicha máquina podría ser considerada inteligente al modo humano (Proudfoot, 2017b: 303). Entender que la finalidad del juego de la imitación es determinar el comportamiento de la máquina es deducir de forma incorrecta el motivo por el que fue propuesto.

La tercera razón pone el foco en el conductismo directamente. Proudfoot señala que el conductismo no puede contemplar la estructura del juego de la imitación (Proudfoot, 2017b: 303), en el sentido de que en dicho juego se espera que la máquina engañe al interrogador. Es decir, que lo

[45] Véase Searle (1980: 423). Si bien Searle no se refiere expresamente a Turing sí que lo hace en relación a la IA *fuerte*.

verdaderamente relevante del juego es si el interrogador es engañado o no. Si los conductistas quisieran poner a prueba a la máquina no centrarían su interés en el engaño al interrogador, sino única y exclusivamente en las respuestas de la máquina. Así pues, es poco menos que un dislate creer que un conductista requiera de una prueba semejante al test de Turing para manifestar su postura.

Sin lugar a dudas, el test de Turing constituye un argumento fundamental dentro del punto de vista A. Pero es completamente necesario comprenderlo de la forma adecuada.

Penrose, por su parte, lo hace. Nuestro autor tiene claro hacia quien va dirigido el test y la poca relación que tiene con corrientes como el conductismo. Sin embargo, Penrose tiene claro que el test de Turing no constituye una prueba concluyente de la inteligencia de las máquinas. Por ello se pregunta: ¿hasta qué punto pasar la prueba lleva a inferir como conclusión necesaria que la máquina piensa? Penrose no lo tiene claro y expone sus argumentos.

4.4. Penrose y su juicio al test de Turing

Hemos visto que Turing defendía que las máquinas pueden imitar partes del ser humano y que era cuestión de tiempo que esta imitación se quedara en una simple anécdota porque las máquinas podrán llegar a pensar como lo hacemos los humanos. Una de las primeras facetas en las que el matemático inglés se interesó para que el avance del pensamiento de las máquinas fue en la destreza que estas podían adquirir en juegos de mesa. El ajedrez ocuparía un lugar privilegiado, ya que es un juego en el que el uso de la inteligencia es imprescindible.

Turing no pudo crear una máquina jugadora de ajedrez lo suficientemente potente como para batir a un ser humano. Pero vimos que la Inteligencia Artificial (ya reconocida como tal) sí que pudo conseguir tal logro. Las máquinas no sólo lograrán derrotar a cualquier humano, sino que algunas que conseguirán victorias antes grandes campeones (como vimos más arriba con el caso *Deep Blue* contra Kaspárov).

Pero, ¿este tipo de logros son lo suficientemente concluyentes como para determinar que las máquinas expertas en ajedrez sean inteligentes y piensen? ¿Puede el test de Turing permitirnos responder a dicha pregunta? Penrose responderá negativamente a ambas preguntas y para dar cuenta de ello expone un ejemplo en el que una máquina experta en ajedrez muestra no comprender[46] el juego en sí mismo más allá para lo que fue programada.

Tal ejemplo envuelve el caso de un potente ordenador experto en ajedrez, conocido como *Deep Thought* (*Pensamiento Profundo*). Pues bien, una vez esta máquina había demostrado con creces sus habilidades para jugar al ajedrez al más alto nivel había que probar su comprensión del juego. Los resultados, sin embargo, fueron más llamativos de lo esperado. Penrose destaca el resultado de una prueba en una partida en particular. Esta estaba dispuesta del siguiente modo:

[46] Entendamos este concepto tal y como lo vimos en la nota 12. Es decir, en relación con los términos conocimiento, inteligencia y consciencia.

- La máquina, que manejaba las fichas blancas, tenía dispuesta en el tablero una barrera de peones de tal modo que impedía el paso de las fichas negras (que también estaban colocadas en forma de barrera con sus peones impidiendo, de tal modo, el movimiento de otras figuras propias como su alfil y sus dos torres) hacia el rey blanco (que era la única figura con la que contaba aparte de los peones). Dicha barrera estaba formada de manera que uno de los peones blancos tenía la oportunidad de «matar» a una torre negra. Este movimiento provocaría la ruptura de la barrera y, en consecuencia, la única defensa que tendría su rey. El movimiento «más inteligente» en este caso es seguir moviendo el rey hasta que la partida tuviese que acabar en empate técnico.

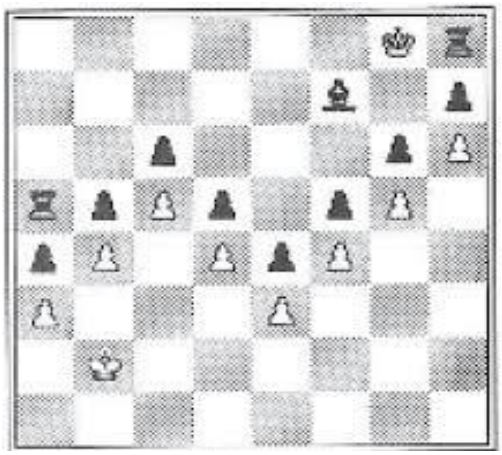

Figura 1. Ilustración de la partida planteada a *Deep Thought*

¿Cuál fue el movimiento que realizó la computadora? Sorprendentemente optó por «matar» a la torre sacrificando, así, la partida al completo. El resultado, empero, no debe sorprender. *Deep Thought* fue programada para atacar cuando tuviese la oportunidad y, de hecho, fue lo que hizo. Y este es justamente el argumento de Penrose. La máquina no puede ir más allá de su programa. La inteligencia que muestra no es más que el resultado de unas capacidades que se le introdujeron desde el principio. Pero, ¿es correcto condenar la posible inteligencia de *Deep Thought* por haber cometido este error en concreto? ¿Acaso los humanos estamos exentos de cometer errores? ¿O cuando los cometemos pasamos a ser automáticamente seres no-inteligentes? Es evidente que los humanos cometemos errores, y muchos, pero hay un rasgo en el error de *Deep Thought* que no debe ser pasado por alto. Muchos humanos podríamos (y me incluyo intencionadamente) haber cometido el mismo error que la máquina, pero no estaríamos hablando de un mismo caso. *Deep Thought* es experta en ajedrez, así que la situación equivalente con respecto a los humanos sería saber cuántos humanos expertos en ajedrez cometerían el mismo error. La respuesta es que muy probablemente ninguno, porque ¡estos comprenden el juego de ajedrez y la jugada era lo suficientemente clara como para dirigirla hacia el empate técnico!

En este caso particular *Deep Thought* no fue sometida al test de Turing hasta que se intentó comprobar su comprensión del juego en el que era experta. Con

su error, el observador de su respuesta pudo ver que existía una gran habilidad por parte de la máquina, pero no podía dar una respuesta concluyente sobre su inteligencia. A pesar de ello, ¿es definitiva la prueba de que *Deep Thought* no comprende el ajedrez? Penrose admite que no. Aunque esté convencido de que la máquina no comprendió lo que sucedió en dicha partida y que nunca llegará a hacerlo, nuestro autor es prudente a la hora de determinar un diagnóstico definitivo en este caso:

> Pero estos son aún días muy tempranos, si vamos a considerar lo que la inteligencia artificial podría alcanzar en última instancia. Los defensores de la IA (ya sea A o B) afirmarán que es solamente cuestión de tiempo, y quizá de algunos significativos desarrollos adicionales en su maquinaria, para que realmente empiecen a hacerse patentes elementos importantes de comprensión en el comportamiento de sus sistemas controlados por ordenador […] podría ser perfectamente posible que uno de tales sistemas construido con suficiente ingenio conserve una ilusión, durante algún tiempo considerable (como sucede con *Deep Thought*), del que posee alguna comprensión […] (pero) la falta real de comprensión general por parte del sistema ordenador debería quedar de manifiesto con el tiempo, al menos, en principio (Penrose, 2012: 63).

Es obvio que Penrose tiene sus convicciones, pero también tiene la sensatez suficiente como para percatarse de que intentar dar respuestas absolutas es una tarea vana, al menos por ahora.

Otras críticas al punto de vista A son menos precavidas y mantienen que es posible demostrar que este modelo es indudablemente incorrecto. Dicha crítica procede del punto de vista B, concretamente del filósofo John Searle, quien se erige como uno de los mayores representantes de la IA *débil*.

5. La Inteligencia Artificial *débil* y la respuesta de Penrose

5.1. IA *débil* vs IA *fuerte*: la Habitación China

Bien conocido es el argumento de John Searle contra la IA *fuerte* a través de su experimento mental de la Habitación China. Por ese mismo motivo obviaré volver a explicarlo y me centraré en los argumentos en contra de dicho argumento a través de la perspectiva de un defensor de la IA *fuerte* actual como lo es Jack Copeland.

A pesar de no explicar la Habitación China sí que es conveniente mencionar las conclusiones principales de dicho experimento. Searle las expone de la siguiente manera:

> Axioma l. Los cerebros causan a las mentes […].
> Axioma 2. La sintaxis no es suficiente para la semántica […].
> Axioma 3. Las mentes tienen contenidos; específicamente, tienen contenidos intencionales o semánticos.
> Axioma 4. Los programas son definidos formalmente o sintácticamente […].
> Conclusión 1. En sí misma, la instanciación de un programa nunca es suficiente para tener una mente (por los axiomas 2, 3 y 4) […]
> Conclusión 2. La manera en que el cerebro causa a las mentes no puede ser sólo por la instanciación de un programa (axioma 1 y conclusión 1).
> Conclusión 3. Cualquier artefacto que tenga una mente tendría que tener poderes causales equivalentes (al menos) a los del cerebro (por axioma 1 trivialmente).

> Conclusión 4. Para cualquier artefacto que tenga una mente, el programa por sí mismo no sería suficiente para proveerle tal mente. El artefacto tendría que tener poderes causales equivalentes al cerebro (por las conclusiones 1 y 3) (Searle, 1995: 442-443).

Searle señala que quien que manipula los símbolos (ya sea un humano o una máquina) puede ser experto en el manejo de la sintaxis, pero que ello es insuficiente para poder comprender tales símbolos, ya que para la comprensión es imprescindible dominar la semántica. El objeto de estudio puede prescindir perfectamente de la semántica, por lo que no podemos contar con *su* comprensión. Entonces, ¿cómo es posible defender que una máquina (o un ser humano que maneja símbolos) puede llegar a comprender si su conocimiento de la semántica es nulo?

Copeland opta por una postura que el mismo Searle se encarga de denominar *objeción de los sistemas* (Copeland, 1996: 195). ¿Y en qué consiste esta postura opuesta a la IA *débil* de Searle? La objeción de los sistemas consiste básicamente en defender que con el experimento de la Habitación China no se está analizando el sistema al completo, sino a una parte de este. La parte que corresponde al manejo de símbolos si bien es una parte indiscutiblemente importante ella sola no constituye la totalidad del sistema que se está estudiando.

Copeland defiende esta postura a partir de una conversación imaginaria con Searle, que divide en cuatro partes. Por nuestra parte, veremos las tres primeras, ya que la última, considero, no añade nada nuevo. Todas ellas tienen la misma estructura: Searle responde a una crítica recibida y expone su contraargumento. Una vez lo expone, Copeland responde a tal contraargumento.

Lo interesante de la conversación imaginaria es que aborda temas esenciales de la defensa de Searle. Sin embargo, pienso, no consigue obtener el propósito de Copeland, este es, demostrar la invalidez del argumento de la Habitación China. Pasemos al análisis de dicha conversación.

El primer punto que se trata es el más evidente. Si el humano dentro de la habitación del experimento es simplemente una parte y concedemos que dicha parte puede no entender el chino, ¿qué otras partes le hace falta para que podamos hablar del sistema al completo? El sistema al completo está formado por el libro de instrucciones que explica al humano las reglas con las que tiene que responder a las preguntas que se le realicen, de lápices, de cuadernos y del mecanismo por el que el sujeto ofrece las respuestas. En algunos casos se ha planteado, incluso, sin la necesidad de contar con lápices o con cuadernos. Pero admitámoslo en la primera forma expuesta para dotar al sistema de una mayor complejidad[47]. Ahora bien, que el humano por sí mismo no entienda el chino es totalmente legítimo (de hecho, así lo entienden los partidarios de la *objeción de los sistemas*), pero, ¿su «fusión» con los lápices, los cuadernos, el libro de instrucciones y el mecanismo le dan la comprensión que por sí mismo no tiene? Searle piensa que este supuesto es absurdo (Copeland, 1996: 195). ¿Cómo unos simples lápices, varios cuadernos, un libro de instrucciones y un mecanismo (que sólo tiene la función de reproducir las respuestas que el humano le da) logran que el

[47] Ya que el experimento realmente no critica la simplicidad del sistema de un programa, sino su insuficiencia para explicar la comprensión.

sistema al completo consiga comprender? El ejemplo de la habitación es el que es. Un humano con la ayuda de todo ese sistema no conseguiría la comprensión del chino. Plantear que sí lo hace es otorgar al resto de cosas algo así como un poder especial. Esto no tiene otro modo de verse que como un absurdo.

La respuesta de Copeland tiene dos aclaraciones. La primera de ellas tiene que ver con el hecho de que el experimento de Searle, al involucrar a un ser humano, no deja lugar, paradójicamente, a la imaginación. Es decir, que no es difícil ver a un ser humano como simplemente una parte de un sistema, aun cuando (tal y como está planteado en el experimento) ¡es una parte del sistema! Una vez que un ser humano entra en la escena parece perderse la referencia de que estamos hablando de un supuesto no real.

La segunda aclaración responde del mismo modo que Searle, es decir, declarando que el planteamiento de la Habitación China en sí mismo es absurdo, en términos de aplicación a la realidad. Cuando los partidarios de la *objeción de los sistemas* defienden la comprensión del sistema (al completo) se les acusa de no atender a la *realidad* de lo que sucedería en la habitación. Pero, replican ellos, la *realidad* en la Habitación China no tiene cabida. Es totalmente imposible que un humano con las condiciones de las que habla Searle en su experimento logre alcanzar el grado de eficiencia que tiene que tener para ejecutar el programa en la forma planteada.

O sea, o se habla o no se habla en términos de realidad. Plantear un experimento mental para luego pedir a tus adversarios que las implicaciones de su defensa se atengan a la realidad va en contra de las exigencias de una discusión seria. Y en este punto tienen razón los partidarios de la *objeción de los sistemas*[48].

Otro asunto es reconocer que el sistema al completo tenga comprensión. Searle cree, con acierto, que ese supuesto es una petición de principio (Copeland, 1996: 196), ya que se está suponiendo como verdadero aquello que Searle pretende demostrar como falso. Copeland en este caso da la razón a Searle, pero apuntala que esa no es su postura particular. Copeland insiste en que existen partidarios de la *objeción de los sistemas* que no piden la petición de principio, sino que centran su atención en la no validez lógica del argumento de Searle. ¿En qué sentido el argumento de Searle no tiene validez lógica? Copeland lo expone con el ejemplo del programa Sam[49]:

> Esto se puede expresar con más sencillez si volvemos a la versión de la escena de la habitación china en que interviene el programa Sam. Searle argumenta que José[50] no entiende la historia y por tanto, tampoco la entiende Sam. Con suerte, ahora estará usted de acuerdo conmigo en que no es un argumento válido. Para saber si Sam comprende, tenemos que fijarnos en Sam, no en José. En realidad, no me hizo falta fijarme mucho en Sam para darme cuenta de que definitivamente no entiende la narración. Sam literalmente no sabe de qué le están hablando. Cuando escribe «John pidió lasaña» no tiene una noción de a qué se refieren estás palabras. Sam […] puede manipular las palabras de forma que produce la ilusión de que

[48] Esta es una opinión personal, no tengo noticia de que Searle haya cedido en esta parte.

[49] Célebre programa dentro del ámbito de la Inteligencia Artificial, ya que puede mantener una conversación sobre acontecimientos típicos de restaurante a través de guiones que se le introducen. Copeland trata el caso de Sam en la misma obra anteriormente, concretamente en el capítulo 2 (1996: 55-59) y en el capítulo 6, precisamente con motivo de la utilización de Sam por parte de Searle como ejemplo de máquina que no comprende (1996: 190-193).

[50] Es el nombre del humano que usa Copeland en su versión de la Habitación China.

> comprende, pero no tiene ni idea de qué significan las palabras. Incluso los creadores de Sam están de acuerdo. Roger Schank escribe: «No se puede decir de ninguno de los programas que hemos escrito que verdaderamente entienda» (Copeland, 1996: 198).

Tenemos la premisa de Searle «José no entiende la historia», y la conclusión «por tanto, Sam tampoco». Estos partidarios de la *objeción de los sistemas* pueden llegar a compartir la no comprensión de Sam (de hecho, así lo reconoce Copeland), siendo la frase de la conclusión verdadera. Pero que la conclusión contenga una frase verdadera no implica que sea válida lógicamente. Alguien puede construir la frase que quiera y poner una frase verdadera como conclusión y ello no le servirá para que dicha conclusión adquiera validez lógica.

Entonces, al no suponer como falso lo que pretende demostrar Searle como verdadero, esta sección de la *objeción de los sistemas* se libra de pedir la petición de principio de la que el filósofo estadounidense les acusa estar pidiendo.

Los partidarios de la *objeción de sistemas*, al menos los de esta segunda parte, tienen que admitir que no pueden asegurar la comprensión del sistema. Pero, por otra parte, también creen que es preciso exigir que argumento de Searle se revise en su forma lógica. Ahora la pregunta pertinente es, ¿es tan necesaria esa revisión lógica? Searle no plantea su experimento de la Habitación China como un problema que precise de una solución de acuerdo a la lógica formal. De hecho, Searle no ha vacilado a la hora de dejar claro que su pensamiento no aspira a ser impecable lógicamente:

> […] Mi enfoque no intenta tratar a la mente como algo formal o abstracto, tal como hace la IA *fuerte*, ni tampoco intenta tratar a la mente simplemente como un conjunto neutral de poderes causales sin características mentales intrínsecas, tal como hacen ciertas formas de funcionalismo. Francamente pienso que el enfoque que voy a presentar es más bien un punto de vista obvio y de sentido común, y hasta que me vi envuelto en esas polémicas recientes, suponía que era ampliamente aceptado, tan ampliamente aceptado que hasta no merecía una enunciación expresa (Searle, 1995: 424).

Por tanto, pedir exquisitez lógica al planteamiento de Searle es querer entablar otro tipo de debate en el que el filósofo estadounidense parece no estar muy interesado.

Pero volvamos, para terminar, a la conversación imaginaria entre Copeland y Searle.

Ahora Copeland imagina que Searle cambia la Habitación China, porque ella implica estar constituida por partes, y, en su lugar, pone todo el peso del experimento en un ser humano. Es decir, el sistema pasa a ser *simplemente* un humano. ¿Cómo se desarrolla el resto del experimento? El planteamiento dice que en vez de tener que precisar de cuadernos, lápices, etc., este humano tiene la capacidad de realizarlo todo en su cabeza (sabiendo de la imposibilidad de que esto pudiera llevarse a cabo, ya que seguimos situados en el contexto de un experimento mental). Tenemos entonces que el humano realiza la misma tarea que en la habitación pero no en un lugar aparte, sino desde una parte de sí mismo (la cabeza). Y nos volvemos a preguntar, ¿comprende el sistema al completo, este es, el humano en su totalidad, el chino? La respuesta de Searle seguiría siendo que no.

Copeland, por su parte, tiene algo que decir a este nuevo planteamiento. Piensa que la conclusión «si una parte del humano no puede hacer esto, el resto del humano tampoco puede», no tiene más remedio que ser errónea. Para mostrar el error de Searle imagina otro escenario en el que un humano también es protagonista. A este humano se le implanta un chip en el cerebro que le permite hacer un tipo de cálculos que antes no era capaz de realizar. Es decir, una parte del humano es capaz de llevar a cabo una tarea que para el resto no es posible. ¿Ello desmonta el argumento de la Habitación China? Copeland piensa que este supuesto pone, al menos, en un dilema (Copeland, 1996: 200) a lo defendido en dicho experimento. Si Searle dice que el humano puede resolver los problemas, aunque este diga que no, entonces el filósofo estadounidense está faltando a una de las bases de su argumento en la Habitación China, esta es, a la declaración del humano. Y si reconoce que no comprende los cálculos, entonces está admitiendo que una parte del humano puede realizar lo que otras no pueden, ¡justo lo contrario que lo defendido en su humano como sistema completo!

En mi opinión, el supuesto de Copeland no pone en demasiados aprietos a la Habitación China, al menos como él pretende hacer ver.

En primer lugar, es de rigor admitir que no comparto la totalidad del pensamiento de Searle. De hecho, considero que tanto en el ejemplo del hombre como sistema completo como en la Habitación China, el filósofo estadounidense dictamina conclusiones como necesarias cuando en realidad no está nada claro que así lo sean. ¿De verdad los humanos de estos experimentos, con las capacidades y los medios que les otorga, son incapaces de aprender el chino? ¿El aprendizaje de la sintaxis, al nivel en el que los humanos de ambos experimentos, es insuficiente como para poder dominar en algún momento la semántica del chino? Que conste que no defiendo un sí rotundo, pero me parece igual de inapropiado el no categórico que ofrece Searle. Tal y como dice Juan Arana en *La conciencia inexplicada* en relación con esto precisamente (Arana, 2015: 55-57) el aprendizaje se asemeja más a lo que realiza el humano en la Habitación China (o la cabeza del humano en el otro ejemplo) de lo que quiere dar a entender Searle. Es cierto que la comprensión (o el aprendizaje) no consiste(n) exclusivamente en el manejo de la sintaxis, pero es legítimo preguntarse si un dominio de estas características podría permitir hablar de una comprensión de la semántica. Al tratarse de un experimento mental la duda debe seguir ahí, responder de manera definitiva no es la opción más conveniente.

Por otro lado, el argumentario de Searle es consistente, al menos mucho más de lo que pretende hacer ver Copeland. Cuando el filósofo británico utiliza el ejemplo del humano con el chip-resuelve-problemas en el cerebro no sigue el mismo criterio que Searle sigue en su exposición. Tanto en la Habitación China como en el humano como sistema completo la última palabra no la tiene el humano del experimento, sino el observador-planteador del problema. Es cierto que es importante lo que dice el humano, pero solamente porque aquello que dice responde a la conclusión de quien plantea el experimento. Si el humano dice que comprende el chino en realidad es irrelevante para Searle, porque no cabe esa posibilidad. Entonces, ¿qué ocurre con el ejemplo del humano con el chip? ¿Tiene que acabar dando la razón Searle a Copeland? No necesariamente. Cuando Searle plantea sus experimentos dota a esos humanos de unas capacidades que normalmente no

tienen los humanos. Aun así, no dejan de ser una exageración de capacidades que sí tenemos los humanos. El caso del humano con el chip en el cerebro que le permite solucionar problemas que antes no podía otorga al ser humano una capacidad que no podemos contemplar en nosotros ni mínimamente rebajada. Como supuesto en sí mismo es plenamente legítimo, pero como supuesto que equivale al de Searle no. Un experimento mental puede contraargumentarse con otro experimento mental siempre y cuando su equivalencia sea evidente. Y en este caso, insisto, entiendo que no lo es.

5.2. Penrose y la IA *débil*

Penrose no es partidario del punto de vista de Searle. Pero en el fondo los argumentos de ambos no se alejan tanto. Penrose incluso llega a admitir que está de acuerdo con buena parte de lo que plantea Searle en contra de la IA *fuerte* (Penrose, 1991: 42-43), aunque otro asunto es la totalidad del pensamiento del filósofo estadounidense. En este punto veremos tanto lo que les une como lo que les separa.

Comenzando con las desavenencias, Penrose critica fundamentalmente dos aspectos, a partir de críticas que Searle ha recibido anteriormente. El primero de estos aspectos está relacionado con la concepción de «comprensión» que maneja Searle. Cuando el filósofo estadounidense dice que el humano del experimento no comprende absolutamente nada ni al principio ni al final de la prueba, está siendo impreciso con el concepto de comprensión. En un sentido más amplio, la localización de cambios en los caracteres que observa el humano, por ejemplo, implica algún tipo de comprensión. Si el humano responde con una frase concreta un número determinado de veces y en un momento *percibe* que un carácter ha sido modificado, en ese percibir se hay necesariamente alguna comprensión. No estaríamos hablando de una comprensión –llamémosla así- *final*, pero sí cierto tipo de ella. Esto está relacionado de alguna forma con lo visto más arriba acerca del aprendizaje. Searle, al negar cualquier tipo de comprensión, parece no tener en cuenta que la repetición forma parte del aprendizaje[51]. Y si bien con la sola repetición no se obtiene una comprensión completa, sí que con ella logramos favorecer un tipo determinado de comprensión. Por otra parte, Penrose comparte la idea de que con el mero hecho de llevar a cabo la tarea de la que se encarga el humano del experimento (nuestro autor habla de «ejecutar los detalles del algoritmo») no le garantiza que pueda concretar los significados reales de las historias en chino (Penrose, 1991: 43). Esto último está relacionado con el segundo aspecto del planteamiento de Searle que Penrose critica.

El segundo rasgo tiene que ver con la complejidad que plantea la habitación. Dicha complejidad está dividida en dos partes: i) la imposibilidad de llevarse a cabo tal experimento, y ii) la poca relación, en términos de complejidad, que tiene un programa computacional y el programa que debería ejecutar el humano del experimento. La primera de las partes es notablemente superficial, ya que Searle habla de un experimento mental. De hecho, Penrose

[51] Para un ejemplo útil acerca de las diferentes facetas involucradas en el aprendizaje véase Minsky (1986 123-136).

la reconoce como poco seria (Penrose, 1991: 43). La segunda, sin embargo, la tiene en cuenta, ya que con ella se pone encima de la mesa cuestiones de *principio* y no cuestiones prácticas, como sucede con la primera (Penrose, 1991: 43). La cuestión de *principio* de esta parte se basa en que el programa debe tener una complejidad mínima como para que esta pueda pasar el test de Turing. Dicha complejidad se da pero hasta tal punto que resultaría imposible para cualquier ser humano llevar a cabo el experimento tal y como lo plantea Searle. La diferencia entre las dos partes reside en que para la segunda declara la realización del experimento le resulta imposible en relación a lo complejo que resultaría a un humano ejecutar un programa que intenta hacer las veces de una actividad mental humana. Mientras que en la primera se apelaba a la mera imposibilidad práctica de llevarse a cabo, sin tener en cuenta la ejecución del programa (Penrose, 1991: 44).

Siendo honestos es casi imperceptible la diferenciación en el segundo aspecto destacado por Penrose. Esto es consecuencia de la poca fuerza que tiene el argumento de la imposibilidad llevar a efecto el experimento. ¡Se trata de un experimento mental! Que ciertos aspectos de un planteamiento no sean realizables en nuestro mundo supone una parte esencial de los experimentos mentales. A partir de ahí, cualquier crítica en esa dirección no tiene más remedio que ser considerada como trivial.

En cambio, la primera de las críticas sí que aborda una cuestión, filosóficamente hablando, más rica. Ella obliga a replantear las características y métodos de aprendizaje. Es lícito plantear que el humano del experimento después de manejar innumerables veces los caracteres chinos pueda estar familiarizado de alguna forma con estos, hasta el punto de poder detectar alteraciones. La legitimidad de este planteamiento estriba en la posibilidad *real* de esta familiarización. De hecho, me atrevería a decir que aquellos que se autoproclaman negados para aprender idiomas podrían llegar a reconocer dicha posibilidad. Sin duda, la repetición es una herramienta eficaz a la hora de aprender y más si a esa repetición le añadimos reglas. Si en el experimento cambiásemos el libro de reglas para responder a las preguntas por otro de reglas sobre cómo hablar el idioma chino, por lo cual se respondería en base de lo que se entendiera, ¿seguiría siendo imposible aprender chino? El tiempo que tardaría cualquier humano en pasar el *test del idioma chino* probablemente sería amplio en la gran mayoría de los casos, pero al final se acabaría superando porque ¡se ha aprendido el idioma! Negar una capacidad que los humanos tenemos, en este caso distinguir caracteres, sí que plantea verdaderamente una cuestión de principio.

No obstante, el tema no puede adquirir una importancia mayor porque obligaría entrar en terrenos que nada tienen que ver con el propósito de este trabajo. Por este motivo es conveniente seguir viendo aquello que Penrose encuentra inadecuado en la defensa de Searle.

El filósofo estadounidense piensa que la clave de la diferencia entre las máquinas y los seres humanos se encuentra en el aspecto biológico. Aquello de lo que estamos hechos es lo que permite que el humano pueda tener la capacidad de adquirir el manejo de la semántica (si seguimos con el ejemplo de la habitación). Pero, ojo, Searle no niega una conexión entre el poder causal del cerebro y el de los ordenadores:

> Algunos suponen que sostengo que en principio es imposible para los chips de siliconas duplicar el poder causal del cerebro. Ése no es mi argumento; es más, no

> tiene ninguna conexión con mi argumento. Que los poderes causales de las neuronas puedan ser duplicados en algún otro material, como chips de siliconas, válvulas electrónicas, transistores, latas de cerveza, o alguna desconocida substancia química, es una cuestión fáctica, que no ha de ser resuelta apelando a bases puramente filosóficas o a priori (Searle, 1995: 220).

De hecho, llega a afirmar que el humano se asemeja más a las máquinas de lo que en un principio pueda parecer:

> Asumir si es posible producir artificialmente una máquina con un sistema nervioso, neuronas con axones y dendritas, y todo el resto, suficientemente como los nuestros, la respuesta […] parece ser obviamente, sí. Si se pueden duplicar exactamente las causas, se podrían duplicar los efectos. Y, de hecho, sería posible producir conciencia, intencionalidad y todo lo demás utilizando otros tipos de principios químicos distintos de los que utilizan los seres humanos. Es […] una cuestión fáctica.
> «De acuerdo, pero, ¿podría una computadora digital pensar?»
> Si por «computadora digital» queremos decir cualquier cosa que tenga un nivel de descripción en el que pueda describirse correctamente como la instanciación de un programa de computadora, entonces, de nuevo, la respuesta es, por supuesto, sí, ya que nosotros somos la instanciación de cualquier número de programas de computadora, y nosotros podemos pensar.
> «Pero, ¿podría algo pensar, entender, y demás solamente en virtud de ser una computadora con el programa correcto? ¿Podría instanciar un programa, el programa correcto por supuesto, por sí mismo ser una condición suficiente del comprender?»
> Esta, pienso, es la cuestión correcta por la que preguntar, puesto que es confundida normalmente con una o más cuestiones anteriores, y la respuesta a ella es no.
> « ¿Por qué no?»
> Porque las manipulaciones de símbolos formales no tienen por sí mismas ninguna intencionalidad; ellas son insignificantes; ni siquiera son manipulaciones de símbolos, ya que los símbolos no simbolizan nada. En la jerga lingüística, sólo tienen una sintaxis pero no semántica. La intencionalidad que las computadoras parecen tener está únicamente en las mentes de quienes las programan y quienes las usan, quienes envían los *inputs* y quienes interpretan los *outputs* (Searle, 1980: 422).

Es decir, que los humanos pensamos a partir de la instanciación de programas de computadora. Por eso las máquinas, que también piensan a través de la instanciación de programas de computadora, muestran un pensamiento similar al de los seres humanos. Pero la cosa cambia cuando se intenta dar una explicación de la intencionalidad de las máquinas. Según Searle, la mera manipulación de símbolos no puede producir ninguna intencionalidad, ya que estamos hablando de un aspecto superficial de la mente. Pero si los humanos y las máquinas pensamos del mismo modo, ¿qué es lo que permite al humano tener intencionalidad e impide que la máquina la posea? La parte biológica, tal y como vimos más arriba. La actividad de nuestro sistema biológico puede ser reproducida por otros componentes no-biológicos, pero su reproducción se quedaría en la pura imitación. ¿Tan importante es la composición biológica como para que sea el único factor que impida que máquinas y humanos sean iguales? Penrose discrepa en este asunto. Nuestro autor va más lejos y cree que en realidad las máquinas ni tan siquiera reproducen fielmente la actividad de la mente humana. Es decir, que si bien para Penrose el factor biológico tiene una importancia innegable en la diferenciación entre máquinas y humanos, también encuentra inoportuno hacer caer todo el peso en él:

> ¿Qué hay tan especial en los sistemas biológicos —aparte quizá de la forma «histórica» en que han evolucionado (y el hecho de que nosotros seamos uno de esos sistemas), que los discrimina como los únicos objetos a los que se permite alcanzar intencionalidad o semántica? La tesis me parece sospechosamente dogmática, ¡quizá no menos dogmática, incluso, que las afirmaciones de la IA *fuerte* que sostienen que la simple ejecución de un algoritmo puede producir un estado de consciencia! (Penrose, 1991: 48).

A pesar de que Penrose siente cierta simpatía por lo que expone el experimento de la Habitación China, su conclusión acerca del pensamiento de Searle no se aleja demasiado con respecto a la que tiene reservada para los partidarios de A[52]. El punto de vista B de Searle tiene la audacia de intentar adivinar el modo en el que el pensamiento de las máquinas y los humanos se asemejan para, luego, recular y no atreverse a comprobarlo de ninguna forma. Penrose es contrario a esta actitud. Su postura se sostiene en la idea de que la última palabra la deben tener las pruebas científicas y no un principio dogmático. Las ciencias de las que se sirve nuestro autor para sus argumentos son la física y la matemática. Pero antes de pasar a analizar algunos de esos argumentos volvemos a preguntarnos…

6. ¿Tienen inteligencia las máquinas?

6.1. Un último test al test de Turing

Damos por sentado que el desarrollo de las máquinas seguirá su curso y que el nivel de sofisticación será cada vez mayor. Sin embargo, ¿podremos contar también con que esas máquinas sofisticadas sean poseedoras de inteligencia? ¿Cómo podríamos saberlo?

Muchos encuentran en ese momento de perfeccionamiento de las máquinas el escenario idóneo para someterlas a un test de Turing realmente válido. Pero tampoco hemos podido decir que hasta ahora dicha prueba haya sido determinante a la hora de detectar inteligencia en las máquinas[53]. Esto nos lleva al planteamiento de otra cuestión: ¿no podemos detectar inteligencia en las máquinas a través del test de Turing porque estas aún no la tienen o porque el test es insuficiente para dicha tarea?

La respuesta de un fiel partidario de las ideas de Turing es que obviamente aún no disponemos de máquinas lo suficientemente inteligentes como para que superen el test de Turing. Una vez las máquinas alcancen el nivel de inteligencia requerido no supondrá un problema para ellas demostrar su inteligencia. El test habrá probado su eficacia y habremos de comenzar a aprender a convivir con aparatos [al menos] igualmente inteligente que nosotros.

Por otra parte tenemos los partidarios de la idea de una posible incapacidad intrínseca de la prueba. Estos partidarios no son necesariamente contrarios al

[52] […] Sin embargo, vale la pena dejar constancia de que yo considero que el argumento de la Habitación China ofrece una línea argumental convincente en contra de A, aunque no creo que esta línea argumental sea totalmente concluyente (Penrose, 2012: 57).
[53] Véase nota 31.

punto de vista A[54]. La crítica de estos está relacionada, precisamente, con el papel que ocupa el ser humano en el desarrollo de la prueba.

Se supone que si la máquina consigue engañar al ser humano la inteligencia de la primera quedaría demostrada. La primera parte de la suposición ya ha tenido lugar, pero la implicación no parece quedar manifiesta. Aquello que parece evidente con tal resultado es que quizá al ser humano no le pertenece el papel de determinar si la máquina es inteligente o no. Juan Arana, por ejemplo, dice al respecto lo siguiente:

> Por otro lado, tal como se planteó en sus términos iniciales, el test ha sido ya satisfactoriamente superado por las máquinas, gracias al desarrollo de computadores electrónicos que superaban ampliamente a los primeros modelos de máquinas de Turing en lo que se refiere a velocidad de procesamiento y tiempo de acceso a la información, por no hablar del tamaño físico de la memoria. Parece que si conversamos con ellas a través de teclado y pantalla nos resulta imposible en términos estadísticos descubrir su genuina identidad, a no ser que sea una conversación realmente larga y derrochemos sabiduría. Aquí topa nuestra especie con una nueva humillación, puesto que estos resultados, más que avalar la inteligencia de las máquinas, sugieren nuestra debilidad como jueces del contencioso (Arana, 2015: 93).

¿La solución pasa por resignarnos a no saber jamás si las máquinas son inteligentes o no? No es esto, ni mucho menos, lo que plantea Arana en la cita. Tampoco los críticos del test de Turing. Aquello que se quiere abordar es que este modo de hallar la inteligencia ha demostrado no tener la validez que Turing pensó que podría tener.

Penrose, por su parte, tampoco pone todo el peso de la prueba de inteligencia de las máquinas en el test de Turing. Pero de todos modos, como vimos más arriba con el ejemplo de *Deep Thought*, piensa que si a una máquina se le acaba haciendo un test de Turing en el que se la ponga a prueba su capacidad de comprensión esta está abocada al fracaso. Esto no cambiaría ni aunque la máquina llegase a su grado máximo de sofisticación, porque ¡no tiene capacidad de adquirir consciencia!

6.2. ¿Qué constituye, entonces, la consciencia?

La consciencia y la inteligencia son dos facultades que parecen escapársenos de las manos cuando intentamos atraparlas en el marco conceptual. Paradójicamente, justo cuando creemos que están más cerca de nuestro alcance. Podemos suponer que la consciencia es el resultado de una combinación sutil de algoritmos, la casual expresión del material que compone al ser humano o algo de lo que simplemente no podemos dar cuenta. Todo ello, sin embargo, no contribuye a un mayor entendimiento de lo que *es*. Penrose opta por no dar una definición de la consciencia. En su lugar, considera capacidades que *implican* inteligencia (por tanto, consciencia).

[54] De hecho, actualmente muchos de los partidarios de la IA sostienen que la clave para encontrar los procesos computacionales que acercarían a máquinas y humanos debe buscarse en el cerebro más que en una prueba tipo Turing (Churchland, 2007: 123-125)

Nuestro autor está convencido de que tales capacidades son de naturaleza no computacional y que ello es demostrable desde ellas mismas.

Entonces, ¿qué es la consciencia? Penrose no da una definición ni la identifica con una característica en particular como los partidarios de A o B. Tampoco renuncia a poder definirla, tal y como lo hace D. Nuestro autor sabe de las dificultades que acarrea concretizar un discurso sobre la consciencia, aunque también de lo inadecuado que resulta intuirlo:

> Esto no pretende sugerir que yo crea que de verdad conocemos intuitivamente lo que es realmente la consciencia, sino simplemente que existe un concepto semejante que estamos tratando de captar de algún modo –un fenómeno genuino y científicamente descriptible, que juega un papel activo como pasivo en el mundo físico (Penrose, 2012: 54).

¿Cuáles son las capacidades de las que habla Penrose? En concreto son dos: la capacidad del pensamiento matemático y la capacidad física. Nuestro autor cree que en estas dos capacidades se encuentra la clave del asunto:

> […] la propia ciencia y las matemáticas han revelado un mundo lleno de misterio. Cuanto más profundo se hace nuestro conocimiento científico, más profundo es el misterio que revela. Es quizá digno de señalar que los físicos, que son los que más directamente familiarizados con las formas enigmáticas y misteriosas en que se comporta *realmente* la materia, tienden a adoptar una visión del mundo menos clásicamente mecanicista que los biólogos. […] Podría suceder perfectamente que, para acomodar el misterio de la mente, necesitáramos una ampliación de lo que actualmente entendemos por «ciencia», pero no veo razón para hacer ninguna ruptura clara con los métodos que nos han servido tan extraordinariamente bien (Penrose, 2012: 66).

Es necesario que primeramente se conozca de manera amplia cómo funciona el pensamiento matemático. Una vez se conquiste este paso será posible hacer un estudio de la física adecuado. Penrose adopta, de este modo, lo que él mismo se encarga de definir como un punto de vista *platónico* de las cosas:

> […] Según Platón, los conceptos matemáticos y las verdades matemáticas habitan en un mundo real propio que es intemporal y sin localización física. El mundo de Platón es un mundo ideal de formas perfectas, distinto del mundo físico, pero en cuyos términos debe entenderse este mundo físico. También está más allá de nuestras imperfectas construcciones mentales; pese a todo, nuestras mentes tienen algún acceso directo a este reino platónico a través de un "conocimiento" de las formas matemáticas y nuestra capacidad para razonar sobre ellas (Penrose, 2012: 66).

Sobre la importancia que tendrá el platonismo dentro el pensamiento de Penrose volveremos en el siguiente capítulo.

¿Por qué cree Penrose que entender el pensamiento matemático es decisivo para poder dar una respuesta satisfactoria sobre la no computabilidad de la consciencia? La computación es intrínsecamente matemática, por lo que si queremos intentar llegar a conclusiones acerca de ella es totalmente necesario que la abordemos desde su naturaleza matemática y no desde fuera de ella.

Ante la pregunta « ¿qué constituye la consciencia?», Penrose no da una respuesta afirmando saber la respuesta. Aquello que hace, como hemos visto, es dotar de una característica que se le niega desde la Inteligencia Artificial:

la no-computabilidad. Veamos, pues, la exposición de los argumentos matemáticos que intentan dar fuerza a su postura.

Capítulo 2

La importancia del pensamiento matemático

1. La peculiaridad del pensamiento matemático

1.1. El quehacer del pensamiento matemático y el estatus de las matemáticas

> *[…] La filosofía se halla escrita en el gran libro que está siempre abierto ante nuestros ojos –quiero decir, el universo-; pero no podemos entenderlo si antes no aprendemos la lengua y los signos en que está escrito. Este libro está escrito en lenguaje matemático, y los símbolos son triángulos, círculos u otras figuras geométricas, sin cuya ayuda es imposible comprender una sola palabra de él y se anda perdido por un oscuro laberinto (Galileo, 1964: 631-632).*

Penrose defiende que en las matemáticas y, sobre todo, en el modo en el que los seres humanos las pensamos se encuentra una de las claves fundamentales que esclarecerían la no-computabilidad de la mente y la consciencia humana. En un principio esto puede resultar irónico si tenemos en cuenta el punto de vista de A. Nuestro autor lo expresa del siguiente modo:

> […] Si pensar es precisamente llevar a cabo una computación de cierto tipo, parece que deberíamos ser capaces de ver esto más claramente en nuestro pensamiento matemático. Pero, de forma notable, resulta todo lo contrario (Penrose, 2012: 78).

Pero la ironía es tan sólo superficial. Las matemáticas contienen elementos que son fácilmente reproducibles en términos computacionales. El cálculo es un ejemplo evidente de ello. Una máquina calculadora, sin necesidad de ser excesivamente avanzada, supera con creces a cualquier humano medio en la capacidad del cálculo matemático. No obstante, Penrose está mencionando concretamente el pensamiento matemático, que es diferente a las matemáticas en general[55]. En un sentido general, las matemáticas son el saber del que se ocupa el pensamiento matemático. Más adelante veremos de un modo más extenso qué son las matemáticas para Penrose y algunos aspectos más sobre estas. Sigamos viendo, pues, el pensamiento matemático y aquello que lo caracteriza.

Sin duda el pensamiento matemático tiene características especiales. Un aspecto que lo define es la dificultad de llevarlo a cabo. Tan peculiar es este modo de pensamiento que no pocas personas consideran a las matemáticas como un muro que no pueden atravesar por mucho que lo intenten. Aun así, no son esta clase de afirmaciones las que permiten explicar la *dificultad* del pensamiento matemático. Aquello que sí lo logra hacer es la idea de que este tipo de pensamiento requiere una «profundidad»[56] especial y nada deductiva («deductiva» en términos de simplicidad y no de deducción, que tan importante es en matemáticas).

El pensamiento matemático es un pensamiento que se conquista. Esta es una idea que en el campo de la filosofía ha tenido un gran número de partidarios. Filósofos como Platón o Descartes eran muy explícitos en defender esta idea en particular acerca del pensamiento matemático.

El filósofo griego, por ejemplo, entendía que si el pensamiento matemático es difícil de hacer frente es porque este participa de la Idea del Bien, la idea más difícil y menos asequible, aquella que nos da la Verdad. Alguien que pretenda divisar la Idea del Bien no puede prescindir del pensamiento matemático:

> -Sería conveniente, Glaucón, establecer por ley este estudio y persuadir a los que van a participar de los más altos cargos del Estado a que se apliquen al arte del cálculo, pero no como aficionados, sino hasta llegar a la contemplación de la naturaleza de los números por medio de la inteligencia; y tampoco para hacerlo servir en compras y ventas, como lo hacen los comerciantes y mercaderes, sino con miras a la guerra y a facilitar la conversión del alma desde la génesis hacia la verdad y la esencia (Platón, 1992: 354).

[55] Esta distinción ya había sido objeto de estudio anteriormente. Véase Russell (1983: 160-182).

[56] Utilizo este término sabiendo que puede ser controvertido. El matemático Godfrey Harold Hardy entendía que era necesario otorgar a las matemáticas (él hablaba concretamente de ideas matemáticas) de «profundidad». Pero también consideraba que era necesario distinguir dicha profundidad de la «dificultad»: […] Tiene algo que ver con dificultad pues las ideas más «profundas» son habitualmente las más difíciles de comprender, aunque estos dos términos no expresan en absoluto lo mismo. Las ideas subyacentes en el teorema de Pitágoras y en sus generalizaciones son bastante profundas, pero ahora ningún matemático las encontraría difíciles. Por otra parte, un teorema puede ser esencialmente superficial y, sin embargo, ser difícil de probar (como son muchos teoremas «diofánticos», que son los relativos a la solución de ecuaciones en los números enteros) (Hardy, 2014: 81). No obstante, mi utilización sí que entiende la profundidad con la dificultad, algo que cobra sentido con la explicación que sigue a continuación.

Platón, por ello, difícilmente reconocería el mérito de una máquina calculadora en términos de pensamiento matemático. El filósofo griego defiende que en el pensamiento matemático debe estar implicada el alma, algo que no es necesario si lo que se pretende es hacer meros cálculos mecánicos.

Por su parte, Descartes dice algo prácticamente igual[57] en su obra *Reglas para la dirección del espíritu*. En ella el filósofo francés defiende que el simple aprendizaje automático de las matemáticas no contribuye a la búsqueda de la verdad. Ni tan siquiera aquellas partes de las matemáticas que garantizan conocimientos claros y distintos. La matemática que permite emprender el camino hacia la verdad es la que obliga al pensamiento tomar parte actuante:

> Cuando por primera vez me dediqué al estudio de las matemáticas, leí desde luego la mayor parte de las cosas que suelen enseñarse por sus autores, pero cultivé principalmente la aritmética y la geometría, porque eran consideradas como las más sencillas y como camino para las otras. Pero no caían por entonces en mis manos autores que me satisficieran plenamente en ninguna de las dos; porque es verdad que leía en ellos muchas cosas respecto de los números que yo comprobaba que eran verdaderas, por cálculos hechos después; y por lo que toca a las figuras, presentaban, por decirlo así, ante los mismos ojos muchas verdades, que sacaban necesariamente de ciertos principios; pero me parecía que no hacían ver suficientemente al espíritu por qué tales cosas eran así y cómo se hacía su descubrimiento (Descartes, 2011: 12-13).

No basta con manejar las matemáticas, hay que pensarlas. Ello significa emprender un camino arduo pero (y, como hemos visto, de esto no dudaban ni Platón ni Descartes) es uno de los caminos que conducen a la verdad. Pero, ¿en qué consiste el pensamiento matemático? Pues precisamente en esto que estamos viendo: en la racionalización de las matemáticas.

El origen del pensamiento matemático como tal se sitúa en la Antigua Grecia. Se conoce que en Babilonia y en Egipto también existieron métodos prácticos basados en las matemáticas. Esto métodos, sin embargo, no requerían del pensamiento matemático que venimos viendo. La diferencia fundamental entre la cultura griega y las anteriores es que en estas últimas las matemáticas eran principalmente cuantitativas mientras que en la Antigua Grecia estas adquirieron también un marcado carácter cualitativo[58] (Kline, 1972: 49-50).

[57] Aunque criticara a los primeros filósofos por promover el conocimiento de las matemáticas precisamente por ser las más fáciles (Descartes, 2011: 13). Esto no deja de ser un error por parte de Descartes.

[58] Considerar a las matemáticas como predominantemente cuantitativas ha sido una corriente intermitente a lo largo de la historia. Si bien con los griegos, como hemos visto, las matemáticas tomaron una naturaleza más cualitativa, ello no impidió que siguieran entendiéndose por mucho tiempo como cuantitativas. No obstante, la entrada en escena de elementos matemáticos peculiares (como los números complejos, la teoría de conjuntos, etc.) las matemáticas han dejado de lado el ser mayoritariamente cuantitativas para pasar a ser cualitativas. Gauss, por ejemplo, entendía que «el matemático abstrae completamente de la calidad de los objetos el contenido a sus relaciones; él solo se ocupa de contar y comparar sus relaciones entre sí» (Gauss, cit. por Ferreirós, 2016: 109) para, más tarde, reconocer que esta comparación no tenía por qué ser necesariamente cuantitativa, pudiendo ser cualitativa con propiedades topológicas (Ferreirós, 2016: 109).

El pensamiento matemático, por tanto, es aquel que permite tener un conocimiento[59] de las matemáticas, diferenciándose, así, de los métodos meramente prácticos de las matemáticas.

Penrose defiende que ser un experto en cálculo matemático al nivel en el que lo son las máquinas no les garantiza la comprensión de las matemáticas. Es por esta razón por la que nuestro autor entiende que las matemáticas tengan una importancia capital a la hora de refutar la posibilidad de la computabilidad de la mente y la consciencia humana[60].

1.2. El alcance de las matemáticas

¿Hasta qué punto es conveniente otorgar a las matemáticas la importancia que Penrose les da? ¿No se ha podido comprobar no pocas veces que la infalibilidad de las matemáticas no siempre juega a su favor? La defensa de Penrose no está basada en la obtención de resultados, sino en un asunto de complejidad mayor: los fundamentos de las matemáticas. Nuestro autor considera que en el fondo del asunto de la computabilidad de la mente y la consciencia humana, sobre todo desde Turing, aquello que se está poniendo encima de la mesa es el debate de los fundamentos de las matemáticas (Penrose, 1991: 151-153)[61]. En este debate, entre otros muchos asuntos, se intenta dar una respuesta concluyente acerca del alcance de las matemáticas, para lo cual es plenamente necesario determinar su naturaleza. Este apartado concreto de los fundamentos de las matemáticas será el que veremos desarrollado, ya que es aquel que interesa a Penrose. Nuestro autor expone las posturas formalista, intuicionista y platonista[62], siendo esta última con la que él mismo se identifica.

2. Fundamentos de las matemáticas y su relación con la computabilidad de la consciencia

2.1. El formalismo

[59] En el sentido expresado en la nota 12.

[60] En el prefacio de EcalR lo expresa del siguiente modo: […] El conocimiento que tenemos de los principios que realmente subyacen el comportamiento de nuestro mundo físico depende, de hecho, de una apreciación de sus matemáticas (Penrose: 2006: 21-22). Es decir, que para Penrose las matemáticas no sólo son importantes por lo que puedan decir de sí mismas, sino por cómo podemos comprender nuestro mundo físico a través de la relación entre las unas y el otro. Sobre cómo entiende Penrose la relación de las matemáticas con el mundo físico lo podremos ver desarrollado más tarde en este mismo capítulo.

[61] De hecho no es el único. Véase asimismo Jack Copeland et al. (2017), concretamente en los capítulos de Copeland (2017: 49-56), y de Grattan-Guinness (2017: 437-442). Los capítulos citados tienen su importancia en el desarrollo de §2.4.

[62] Frente a la clasificación convencional de logicistas, intuicionistas y formalistas (Carnap, 1983:41). La diferencia reside en que Penrose incluye dentro de la postura formalista a los logicistas (Penrose, 1991: 136-157).

Penrose sitúa dentro del formalismo las ideas de Gottlob Frege, los sistemas desarrollados por Russell y Whitehead en los *Principia Mathematica* y el programa de David Hilbert[63]. De todos modos, quien representará en mayor medida esta corriente según nuestro autor (y en esto coincide con la mayoría de estudios acerca de este tema) será el programa de Hilbert.

A grandes rasgos, el programa de Hilbert consiste en *dar* con un sistema formal en el que se incluyan todos los razonamientos matemáticos correctos para cualquier área matemática particular (Penrose, 1991: 139-140). Dicho sistema formal debe estar compuesto de axiomas que deben ser completos. Es decir, que en estos axiomas deben estar entendidos los conceptos que ellos contienen (Ferreirós & Gray, 2006: 63). De este modo se asegura de manera irrevocable la verdad o la falsedad de cualquier elemento que participe en el sistema. No sería posible, entonces, la presencia de contradicciones, dando lugar, así, a una especie de nominalismo (Herce, 2014: 60-61). En esta construcción subyace una propiedad en la que recae el peso a la hora de determinar el sistema formal que permitiría dicha tarea: la *relación formal*:

> En mi opinión, un concepto puede ser fijado lógicamente sólo por sus relaciones con otros conceptos. Estas relaciones, formuladas en ciertas declaraciones (*statements*), que yo denomino axiomas, siendo evidente que tales axiomas (quizá junto con proposiciones asignando nombres a conceptos) están en la definición de los conceptos (Hilbert, cit. por Ferreirós & Gray, 2006: 63).

Es decir, puro formalismo. Cualquier enunciado matemático o lógico podría prescindir de su significado, mientras que nunca de su relación formal. Entonces, ¿todo comienza y acaba con las estructuras formales? Ciertamente, Hilbert pretende construir un sistema formal que no tenga que rendir cuentas más allá de sí mismo. Es en este sentido en el que podemos hablar del programa de Hilbert como un juego con símbolos:

> La idea detrás de estas imágenes de «juego», presumiblemente, es que cuando se borra todo rastro de significado, como en la visión de Hilbert de las matemáticas ideales, las matemáticas se convierten en última instancia en una actividad de manipulación de símbolos realizada de acuerdo con ciertas reglas; reglas que, además, no responden a nada tan serio como una preocupación por la verdad objetiva, sino solo a preocupaciones menos importantes como un impulso subjetivo o psicológico de unidad lógica en nuestro pensamiento (Detlefsen, 1996: 81).

Muchos ven en esta concepción un rasgo de radicalidad. Empero, no es estrictamente correcto definir el formalismo de Hilbert como radical. Georg Kreisel, gran estudioso del programa de Hilbert, comenta a tenor de este asunto que mientras que los problemas planteados por Hilbert están referidos a propiedades sintácticas de los sistemas formales, ello no impide que las

[63] Quienes pretendan ser rigurosos en la utilización de los términos puede encontrar en la clasificación de Penrose un problema. La perspectiva de Frege con respecto a la de Russell y Whitehead, por ejemplo, suele considerarse aparte, a pesar de que ambas estén situadas dentro del logicismo. La diferencia fundamental reside, precisamente, en la pertenencia al platonismo por parte de Frege (Maddy, 2003: 26). Si bien Russell y Whitehead no se desvinculan del platonismo de manera explícita, si es cierto que no aparecen rasgos platónicos dentro de su esquema (al menos en el modo en el que sí lo hace en el de Frege). Por su parte, Hilbert, como formalista, es [tal y como veremos más adelante] *anti-platónico*. He aquí el problema de incluir a estos autores en el mismo grupo. De todos modos, Penrose habla de un modo muy general con respecto a este tema, por lo que pedirle rigurosidad, en mi opinión, es algo que quedaría fuera de lugar.

soluciones se den a través de razonamientos intuitivos. Y no sólo eso, sino que dichos razonamientos podrían prescindir de cualquier tipo de formalización, algo que el mismo Hilbert se encarga de aclarar explícitamente (Kreisel, 1983: 208). Por tanto, hablar del formalismo de Hilbert como un formalismo radical carece de base, al menos de una firme[64].

A pesar de todo, este formalismo es marcadamente abstracto. Que se pueda prescindir del significado no es del agrado de todos. En este bando contrario se sitúa Penrose. Nuestro autor piensa que el significado es importante. Su defensa gira entorno a la idea de equiparar, no sin razón, al sistema formal con los algoritmos, mientras que el significado equivaldría al pensamiento matemático:

> […] A ciertas personas les gusta esta idea, según la cual las matemáticas se convierten en una especie de «juego sin significado». Sin embargo, no es una idea que me seduzca. Es, en realidad, el «significado» —y no la ciega computación algorítmica— lo que constituye la substancia de las matemáticas (Penrose, 1991: 144).

Las matemáticas no se basan, por tanto, exclusivamente en sus relaciones formales. Hay un «algo más» en ellas que está fuera del alcance de toda explicación formal. Este es, precisamente, el argumento que Gödel dará frente al formalismo a través de sus teoremas de incompletitud y que Penrose utilizará como apoyo a su postura. Pero esto lo veremos más adelante en este capítulo.

Los críticos del formalismo coinciden en que esta corriente limita de manera notable las capacidades del ser humano. Esto desde la perspectiva de los formalistas resulta paradójico, ya que su modelo surgió, precisamente, para acabar con los límites a los que podrían enfrentarse quienes se dedican a las matemáticas:

> Recordemos que *somos matemáticos* y, como tales, a menudo nos encontramos en una situación similar, y recordemos cómo el método de los elementos ideales, esa creación de genio, nos permitió encontrar un escape. Presenté algunos ejemplos brillantes del uso de este método al comienzo de mi conferencia. Del mismo modo que se introdujo $i = \sqrt{-1}$ para que las leyes del álgebra, por ejemplo, las relativas a la existencia y el número de las raíces de una ecuación, se pudieran preservar en su forma más simple, así como se introdujeron los factores ideales para que las leyes simples de la divisibilidad podría mantenerse incluso para enteros algebraicos (por

[64] Esto, sin embargo, no es óbice para que el programa de Hilbert sea aquel que mejor cumpla los cinco rasgos del formalismo. Estos cinco rasgos (Detlefsen, 2005: 236- 237) son:
1) la aritmética pasa a ser el centro de la investigación matemática (en lugar de la geometría),
2) rechazo de la concepción clásica de la prueba (*proof*) y el conocimiento (*knowledge*) matemático,
3) vuelco a la abstracción en la búsqueda de rigor (en detrimento de la intuición y el significado),
4) el lenguaje como no-representacional con respecto al pensamiento matemático, y
5) la componente creativista (*creativist*), en el sentido de quien hace matemática tiene la libertad de crear instrumentos de pensamiento que permiten promover sus objetivos epistémicos.
A pesar de Hilbert es considerado generalmente ya no sólo como perteneciente al formalismo, sino su mayor representante, ello no impide que esta afirmación haya sido puesta en duda. José Ferreirós (2009) defiende que la postura de Hilbert pasó por varias fases. En este trabajo en particular destaca su etapa como logicista, algo que podría ser muy conflictivo con la idea general del pensamiento de Hilbert, pero de lo cual se puede dar buena cuenta en sus escritos, sobre todo aquellos publicados antes de 1904.

> ejemplo, introducimos un divisor común ideal para los números *2* y *1 + √-5*, mientras que uno real no existe), por lo que debemos *unir las proposiciones ideales a las finitarias* para mantener el reglas formalmente simples de la lógica aristotélica ordinaria (Hilbert, cit. por Detlefsen, 2005: 289).

Sin embargo, para los críticos, este reclamo de Hilbert no tiene sentido (si tenemos en cuenta la crítica de Penrose, por ejemplo). Al depender la verdad o la falsedad del esquema formal del sistema, dejando a un lado el significado, ello inevitablemente le hace tener cierta independencia con respecto al quehacer del ser humano. Por tanto, esa creación de genio de la que habla Hilbert queda anulada de algún modo. Este aspecto entra en conflicto directo con otra de las corrientes que toman parte en el debate de los fundamentos de las matemáticas: el intuicionismo.

2.2. El intuicionismo

Penrose expone el intuicionismo para considerarlo también como una postura contraria a su pensamiento. No obstante, existen rasgos fundamentales en el intuicionismo que son compartidos de alguna forma por nuestro autor. Para verlos de un modo adecuado pasamos a analizar, a grandes rasgos, cuáles son las principales características de esta corriente.

El intuicionismo surge como respuesta contraria al formalismo. Las ideas del intuicionismo profesado por el matemático L.E.J. Brouwer son las más representativas de esta corriente. Esto no significa, sin embargo, que el nacimiento del intuicionismo se le deba al matemático holandés. El intuicionismo es una corriente que se fue gestando en la escuela francesa en el siglo XIX. Hay quienes, incluso, sitúan su origen más tempranamente. El mismo Brouwer concede sin ambages a la filosofía kantiana[65] el comienzo del intuicionismo. Esta no es una consideración, ni mucho menos, descabellada[66]. Bien es sabido que el filósofo prusiano entendía que los axiomas de la aritmética y la geometría eran juicios que no dependían de la experiencia:

> En Kant encontramos una vieja forma de intuicionismo, ahora casi completamente abandonada, en la que el tiempo y el espacio son formas de concepción inherentes a la razón humana. Para Kant, los axiomas de la aritmética y la geometría eran juicios sintéticos a priori, es decir, juicios independientes de la experiencia y no capaces de demostración analítica; y esto explicaba su exactitud apodíctica en el mundo de la experiencia, así como en abstracto. Para Kant, por lo tanto, la posibilidad de refutar las leyes aritméticas y geométricas experimentalmente no solo estaba excluida por una creencia firme, sino que era completamente impensable (Brouwer, 1983: 78).

De tal forma, queda de manifiesto que existen verdades inalcanzables para cierto tipo de fuentes del conocimiento. Pasando al intuicionismo de Brouwer, pero sin abandonar la idea kantiana expresada en la cita, el

[65] Por otra parte, el formalismo también es considerado como una corriente surgida del kantianismo (Detlefsen, 1996: 76).

[66] Sin embargo, sí hay quienes piensan que es llamativa, ya que consideran que el quehacer matemático de Kant en realidad podría ser calificado como mediocre (Arana, 2018: 13).

matemático holandés defiende que los principios de la lógica formal deben ser abstraídos de las intuiciones mentales[67]. Es decir, que estos dependen de la construcción[68] del pensamiento humano y no de una posible independencia de los sistemas formales. Brouwer parece, a todas luces, un seguidor del idealismo alemán:

> […] saben de esa frase muy significativa «conviértete en ti mismo». Parece que hay un tipo de atención que se centra en ti mismo y que, en cierta medida, está dentro de tu poder. Lo que este Ser es, no podemos decirlo más; ni siquiera podemos razonar al respecto, ya que, como sabemos, todo hablar y razonar es una atención a una gran distancia del Ser; ni tan siquiera podemos acercarnos a él mediante el razonamiento o las palabras, sino solo 'convirtiéndonos en el Ser' como se nos da [...]
> Ahora reconocerán su libre albedrío, en la medida en que [...] sean libres de retirarse del mundo de causalidad y luego permanecer libres solo entonces obteniendo una Dirección definida que seguirá libremente, reversiblemente (Detlefsen, 1996: 75-76).

Pasando por alto que en esta cita Brouwer se acerca más a Fichte que a Kant (Detlefsen, 1996: 76), de lo que no hay lugar a dudas es que este modo de entender las matemáticas conduce al intuicionismo brouweriano hacia una perspectiva peculiar. Ello se debe a la concepción filosófica de la conciencia que maneja este autor. Para Brouwer, el modo, o más bien el cuándo, en el que los seres humanos somos conscientes marca el porvenir del desarrollo de las ciencias:

> La conciencia primordial, según Brouwer, es un guiso «onírico» que oscila entre la sensación y el descanso; y, dice, la vida consciente se mantendría así si no fuera por «actos de atención» por los cuales el sujeto se enfoca en cambios individuales entre diferentes contenidos sensoriales. Él llama a tal cambio el «desmoronamiento de un momento de la vida». La conciencia objetiva comienza cuando uno se enfoca en tal evento. A través de estos actos de atención, el sujeto distingue elementos conscientes individuales junto con sus contenidos y discierne un orden entre ellos. El sujeto itera este proceso y por lo tanto forma secuencias mentales atenuadas. Estos a su vez son los bloques de construcción de la conciencia del sujeto de los objetos empíricos ordinarios.
> Brouwer habla luego de lo que él llama «atención causal». El sujeto discierne cierta similitud entre distintos momentos conscientes y luego crea secuencias con partes iniciales similares y partes posteriores similares («cadenas causales»). De esta manera, dice Brouwer, el sujeto produce conciencia del mundo causalmente ordenado. Reflexionar sobre esto es la raíz de la ciencia; manipularlo es tecnología (Posy, 2005: 329).

Que el pensamiento matemático se deba a la construcción del pensamiento humano lleva de manera inevitable a entrar en conflicto con ciertos conceptos matemáticos fundamentales. Aquel que recibe una crítica más feroz por parte del matemático holandés es el concepto «infinito», entendido este de manera

---

[67] Esta es una idea reminiscente, casi de forma inevitable, a dos posturas muy importantes dentro del ámbito científico-filosófico. Estas son las correspondientes a Husserl y a Poincaré. Sobre cómo puede estar relacionado el pensamiento de Brouwer y Husserl véase van Atten (2007). Para la relación con Poincaré véase asimismo Detlefsen (1996: 50-123).

[68] Puede entenderse con esta expresión que incluyo el intuicionismo de Brouwer con las corrientes constructivistas. Si bien esto no es plenamente rechazable es conveniente aclarar que no hago tal inclusión de esta corriente con las otras, ya que el intuicionismo brouweriano es innegablemente particular. Entendamos, por tanto, «construcción» en su sentido más general.

real. Brouwer es contrario a la teoría de conjuntos[69]. Para el matemático holandés no es posible que exista un conjunto infinito, al menos en acto. El conjunto de número naturales $\mathbb{N} = \{1, 2, 3, 4\ldots\}$, por ejemplo, contiene una serie de números que tiende al infinito, pero esa infinitud no podrá darse jamás en acto, ¡porque es imposible de forma constructiva[70]! Este tipo de afirmaciones demuestra lo extremo que resulta el intuicionismo brouweriano[71]. Precisamente el extremismo de los planteamientos intuicionistas es lo que lleva a Penrose a posicionarse en contra de esta corriente (Penrose, 1991: 156-157).

Penrose comparte con el intuicionismo el papel activo que debe tomar el pensamiento del ser humano para el desarrollo de las matemáticas. Pero la gran diferencia entre ambas posturas se sitúa en que para Penrose ese papel activo no puede ser el fundamento mismo de las matemáticas. Para nuestro autor, las matemáticas tienen existencia por sí mismas, es decir, que no operan según se las estudie (tal y como dice el intuicionismo), sino que son independientes. A pesar de que Penrose se digna a diferenciarse del intuicionismo, no menos cierto es que su aclaración no deja de ser poco contundente (Herce, 2014: 63-64).

Pero sí que existen aspectos en los que Penrose deja claro por qué las ideas intuicionistas no le convencen. Un ejemplo de ello es que mientras los intuicionistas rechazan la ley del tercio excluso o los argumentos matemáticos por reducción al absurdo, Penrose encuentra en ellos una herramienta indiscutiblemente útil para las matemáticas. El *quid* de la cuestión se encuentra en la concepción de «existencia»:

> […] para un intuicionista «existencia» significa «existencia constructiva». En un argumento matemático que procede por *reductio ad absurdum* uno desarrolla alguna hipótesis con la intención de mostrar que sus consecuencias conducen a una contradicción, contradicción que proporciona la deseada demostración de que la hipótesis en cuestión es falsa. La hipótesis podría tomar la forma de un enunciado acerca de que una entidad matemática con ciertas propiedades requeridas no existe. Cuando esto conduce a una contradicción, uno infiere, en *matemáticas ordinarias*, que la entidad requerida existe realmente. Pero tal argumento, por sí mismo, no proporciona medios para *construir* efectivamente tal entidad. Para un intuicionista, este tipo de existencia no es existencia en absoluto; y es en este sentido en el que rechazan aceptar la ley del tercio excluido y el método de *reductio ad absurdum* (Penrose, 1991: 155).

Según Penrose, si la existencia también depende de la construcción, ello deriva inevitablemente hacia un escenario en el que las matemáticas dejan de ser una fuente fiable de conocimiento, ya que podría ser sometida a cambios constantemente. Esto es algo inadmisible para Penrose (Penrose, 1991: 157), ya que, como vimos más arriba, él defiende una concepción platónica, la cual otorgan de una verdad y una existencia inamovible a las matemáticas:

> Esta intuición difiere del concepto de intuición directa usado por Penrose. Mientras que la intuición directa de Penrose es platónica, los intuicionistas utilizan el

[69] Véase, por ejemplo, Brouwer (1983: 77-89).

[70] Esta idea es incluida en los que se conoce como «finitismo».

[71] Cuando más adelante me refiera a los intuicionistas hay que entenderlos de esta forma precisamente, a los intuicionistas seguidores de las ideas de Brouwer. Como el mismo Penrose se encarga de puntualizar (Penrose, 1991: 156), existen partidarios del intuicionismo que no comparten la totalidad de las ideas del matemático holandés.

> 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 (Herce, 2014: 62).

Hasta ahora hemos visto algunos de los rasgos que caracterizan al platonismo matemático. En el siguiente punto veremos qué se entiende por este tipo de platonismo de forma general, para luego concretar a grandes rasgos el platonismo que adopta Penrose.

2.3. El platonismo

¿En qué consiste el platonismo matemático? En una definición muy actual[72] de Linnebo (2017) se define a esta corriente del siguiente modo:

> [El platonismo matemático] sostiene que los abstractos objetos matemáticos son tan reales como los ordinarios objetos físicos […] La objetividad matemática se basa en la existencia de objetos matemáticos, que aseguran y explican los valores de verdad objetivos de las declaraciones relacionadas con dichos objetos (Linnebo, 2017: 32).

En esta definición se encuentra implícito un rasgo fundamental del platonismo matemático: su íntima relación con el realismo. De hecho, María del Ponte, siguiendo a Penelope Maddy, entiende el platonismo como realismo aplicado a las entidades matemáticas (del Ponte, 2006: 2). El realismo, por su parte, acepta las siguientes condiciones:

> 1. Las entidades de las que estamos hablando (las matemáticas en el caso del platonismo, las entidades éticas en el caso del realismo moral, etc.) existen.
> 2. Es posible conocer dichas entidades y, de hecho, nuestra mejor teoría acerca de ellas es verdadera (al menos, aproximadamente verdadera).
> 3. Tanto las entidades como la verdad de los enunciados con los que hablamos acerca de ellas son independientes del sujeto. Es decir, las entidades no son construcciones del sujeto, existen con independencia de nosotros, de manera que existirían aún si nosotros no lo hiciéramos y seguirán haciéndolo cuando nosotros no estemos. Aplicado al caso matemático, por ejemplo, esto equivale a afirmar que los sujetos descubrimos las propiedades de las entidades matemáticas y sus relaciones, no las inventamos. Las matemáticas, según el platonismo, son un descubrimiento humano, no una construcción. Por otro lado, la independencia de la verdad trae consigo 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) (del Ponte, 2006: 2-3).

Estas tres condiciones pertenecerían tanto al realismo como al platonismo. No obstante, en el platonismo cabría la posibilidad de añadir una condición extra:

[72] A pesar de que el concepto platonismo se encuentra en la filosofía casi desde su origen, lo cierto es que este no fue adoptado dentro del ámbito matemático hasta el siglo XX, de la mano de Paul Bernays en su artículo "Sobre platonismo en matemáticas" (1935).

> 4. Las entidades matemáticas son abstractas, queriendo decir con esto que están situadas fuera del espacio y del tiempo y que son incapaces de interactuar causalmente (del Ponte, 2006: 3).

Una vez vistas estas condiciones es fácil entender por qué a las anteriores corrientes -el formalismo y el intuicionismo- se les conoce comúnmente como *anti-platónicas*. El formalismo es una corriente anti-platónica en tanto que rechaza la existencia de las entidades matemáticas (este es un aspecto de lo que se conoce como *platonismo ontológico*). El intuicionismo, por su parte, rechaza tanto la independencia de las entidades matemáticas (que también es un rasgo del *platonismo ontológico*) como que la verdad o la falsedad de los enunciados matemáticos no dependan de la actividad del sujeto (que es lo que defiende el *platonismo semántico*) (del Ponte, 2006: 5).

Hablar de los platonismos *ontológico* y *semántico* deja de manifiesto que el platonismo es una corriente que se acepta (y rechaza) por grados. Penrose, por ejemplo, acepta estos dos tipos de platonismo, considerándose a sí mismo un platonista de manera plena. Sin embargo, ello no impide que su pensamiento entre en conflicto con algunas de las condiciones que hemos visto más arriba. Con respecto a la número 3, nuestro autor no la aceptaría del todo, ya que esta determina que la actividad del sujeto es insignificante para la existencia de las matemáticas. Aunque coincide al entender que las matemáticas han existido, existen y existirán con independencia del ser humano, también otorga al ser humano un papel «privilegiado» con respecto a las matemáticas. Dicho «privilegio» consiste en la parte activa que tiene la mente humana para hacer frente a las matemáticas. Es decir, en cierto modo reconoce un tipo de constructivismo, pero no al modo de los intuicionistas. Este aspecto del pensamiento penroseano ha sido considerado en ocasiones como difuso. Pero este es un asunto al que no sólo Penrose se enfrenta, sino, como bien dice Francesco Berto, todo(a) platonista:

> […] Quien profesa el platonismo sostiene que tiene una intuición intelectual de la realidad matemática infinita. Alegremente reconoce que esta intuición no es una comprensión completa y perfecta; después de todo, tenemos un conocimiento extremadamente aproximado del mundo físico externo. Sin embargo, solo los sofisticados argumentos escépticos dan lugar a dudas sobre la existencia y actualidad de esta realidad física que nos rodea. Al admitir que tenemos un conocimiento imperfecto del modelo estándar, la (el) platónica(o) cree que también puede dar cuenta de la necesidad de pruebas en matemáticas: necesitamos pruebas porque en muchos o la mayoría de los casos nuestra imagen de los números no es clara en absoluto, y la prueba tiene para reemplazar la intuición (Berto, 2009: 161).

Más adelante veremos, primero (§5.1), cómo trata Penrose este tema desde su platonismo y, en segundo lugar (Cap. 4), cómo lo hace también a través de su teoría de los tres mundos.

A modo de aditivo y, así, también alcanzar una perspectiva más amplia de qué se entiende generalmente por platonismo, es conveniente tener en cuenta alguno de los tipos de esta corriente.

En primer lugar, tenemos el platonismo denominado como *interno*. Este platonismo se entiende como el componente ideal de las matemáticas. Es por ello por lo que comprende el modo de proceder de todas las matemáticas, incluso aquellas corrientes que niegan cualquier tipo de platonismo, como el constructivismo radical. De hecho, el constructivismo radical constituye uno de los polos de los grados de platonismo *interno*, siendo el platonismo

extremo el opuesto (Ferreirós, 1999: 454). Por su parte, el platonismo *extremo* es aquel que concibe un mundo ideal en el que toda la matemática está presente. Si la clasificación ha quedado lo suficientemente clara puede verse que la diferencia entre el platonismo *extremo* y el *interno* es que el primero pertenece al segundo.

En segundo lugar, tenemos el platonismo *filosófico* o *externo*, con el cual los matemáticos defienden estar desvelando y trayendo al mundo físico los principios de este saber. A pesar de que esta actitud implica concebir la independencia a las matemáticas, no todas las variantes del platonismo, al menos aquellas que son reconocidas como tal[73], sostienen esa misma idea. El platonismo *externo* mantiene que las matemáticas son abstracciones de nuestra experiencia (las figuras geométricas son idealizadas a partir de las formas que vemos en el mundo físico, etc.). De este modo no se está otorgando un mundo aparte, aunque sí una naturaleza diferente (en este caso abstracta).

José Ferreirós describe estos tipos de un modo más esquemático:

> 1. Platonismo interno o propiamente matemático: es característico de las teorías de la matemática abstracta o moderna, donde se hace referencia a elementos cuya existencia se postula y se considera dada (se podría hablar de existencia ideal).
> 2. Platonismo externo, ontológico, o propiamente filosófico (una de las posibles interpretaciones filosóficas de la matemática, en particular de la característica antes señalada de la matemática abstracta): consiste en la afirmación de que los objetos matemáticos gozan de una existencia real, análoga en algún sentido (aunque diferente) a la existencia de los objetos físicos. (Ferreirós, 1999: 448).

Volviendo a la perspectiva de Penrose, es evidente que nuestro autor no vacila a la hora de reconocer su creencia en la independencia de las matemáticas. Y este es un pensamiento, según él, compartido por la mayoría de los matemáticos:

> Debería señalar en primer lugar que cuando los matemáticos elaboran sus minuciosas cadenas de razonamiento consciente para establecer verdades matemáticas, no piensan que estén siguiendo ciegamente reglas inconscientes que son incapaces de conocer y creer. Ellos piensan que están basando sus argumentos en lo que son verdades incuestionables –en definitiva, esencialmente «obvias»- y que están construyendo sus cadenas de razonamiento a partir únicamente de tales verdades. Y aunque estas cadenas puedan a veces ser extraordinariamente largas, difíciles o conceptualmente sutiles, el razonamiento es, en principio y de raíz, incuestionable, firmemente creído y lógicamente impecable. No tienden a pensar que estén actuando en realidad de acuerdo con ciertos procedimientos completamente diferentes, desconocidos o no creídos que, quizá «entre bastidores», guían sus creencias de maneras incognoscibles (Penrose, 2012: 142).

Ya en su artículo de 1935, Paul Bernays reconocía que el platonismo matemático (al menos en su forma menos extrema) reina en las matemáticas (Bernays, 1983: 261). El motivo de esta autoridad del platonismo se debe, aunque en muchas otras ocasiones no esté reconocida, a la apertura a nuevas fuentes de investigación (Brown, 2012: 46). Efectivamente, el platonismo matemático no se queda anclado en su campo. Las consecuencias de su

[73] Ya que, aunque esta rama suele considerarse dentro del platonismo, en el fondo está más cerca del aristotelismo (Ferreirós, 1999: 458).

aceptación y rechazo son significativas en múltiples ámbitos, tal y como hemos podido observarlo en el plano de la filosofía.

¿Qué tienen que ver los fundamentos de las matemáticas con la posibilidad (o no) de una computabilidad de la mente y la consciencia humana? A simple vista parece que ambos temas están separados por un abismo irreconciliable. Sin embargo no es así para Penrose, ya no sólo tiene la absoluta convicción de que tal abismo no es tan descomunal, sino que este, incluso, no existe. A continuación veremos algunos de los argumentos de nuestro autor en favor de esta declaración.

2.4. La computabilidad de la consciencia y su relación con los fundamentos

El motivo principal por el que Penrose defiende que el problema de la computabilidad de la consciencia tiene su origen en los fundamentos de las matemáticas responde a una base histórica.

Cuando formalistas[74] e intuicionistas entablaron el debate acerca de los fundamentos de las matemáticas surgieron respuestas con respecto a diferentes problemas. Concretamente, en 1900 el ya citado David Hilbert propuso en el que sería el segundo Congreso Internacional de Matemáticos en París su célebre serie de 23 problemas que, según él, ocuparían la futura investigación matemática. De dicha serie, el décimo problema resultaría clave para el planteamiento de la computabilidad de la consciencia tal y como la conocemos hoy en día. El problema décimo propone «encontrar un algoritmo que determine si una ecuación diofántica[75] polinómica dada con coeficientes enteros tiene solución entera». Sin embargo, este décimo problema sólo daría lugar al debate de la computabilidad de la consciencia de manera parcial. No sería hasta 1928 cuando el planteamiento de otro problema (que seguía la

[74] Recordemos que este es el criterio de Penrose, el cual incluye a logicistas y formalistas en el mismo grupo.

[75] Las ecuaciones diofánticas deben su nombre al matemático griego del siglo III d.C., Diofanto de Alejandría, que fue quien las estudió por primera vez. Estas son ecuaciones polinómicas, con un número cualquiera de variables en las que todos los coeficientes y todas las soluciones deben ser números enteros*. Un ejemplo de un sistema de ecuaciones diofánticas es el siguiente:

$6w + 2x^2 – y^3 = 0, 5xy – z^2 + 6 = 0, w^2 – w + 2x – y + z – 4 = 0$

Y otro ejemplo es:

$6w + 2x^2 – y^3 = 0, 5xy – z^2 + 6 = 0, w^2 – w + 2x – y + z – 3 = 0$

El primer sistema queda resuelto, en particular, por

$$w = 1, x = 1, y = 2, z = 4,$$

mientras que el segundo sistema no tiene ninguna solución (porque, por su primera ecuación, *y* deber ser un número par, y por su segunda, *z* debe ser también par, pero esto contradice su tercera ecuación, cualquiera que sea *w*, porque $w^2 – w$ es siempre par y 3 es un número impar). Estos ejemplos están tomados de Penrose (2012: 43).

* Un número entero es un número de la lista…, -3, -2, -1, 0, 1, 2, 3, 4,…

misma idea que el décimo de la serie de Hilbert) cuando se allanó el terreno para que el problema de la computabilidad de la consciencia entrara en escena. En tal fecha, el mismo Hilbert junto a su colega Wilhelm Ackermann propusieron en Bolonia el que se conoce como el *Entscheidungsproblem*, o lo que es lo mismo, el problema de la parada. Aquello que se plantea en el problema de la parada es la existencia de un procedimiento mecánico-algorítmico que dé cuenta de ciertos problemas matemáticos[76] relacionado con los sistemas formales. La solución a este problema obtuvo respuesta relativamente pronto (concretamente en 1936) de la mano de Alonzo Church y Alan Turing[77], siendo la respuesta al problema negativa. Es decir, con las aportaciones de Church y Turing se demostró que tal algoritmo no existe.

Lejos de hace un análisis detallado de las soluciones propuestas por Church y Turing, veremos algunas de las características de ambas, aunque haciendo algo más de hincapié en la particular de Turing. El motivo de ello es que la de Turing contribuye al debate de la computabilidad de la consciencia de forma más directa que la de Church, a pesar de que ambas digan lo mismo. A continuación veremos el porqué.

La solución de Church consistía en la realización de un esquema abstracto que daba finalmente con la respuesta negativa al problema de la parada. Este esquema es conocido como cálculo lambda[78]. Esta solución pone de manifiesto la naturaleza matemática de la noción de computabilidad. Pero la noción que maneja Church, sin embargo, tiene poco que ver con las máquinas computadoras, al menos en primera instancia[79]. El cálculo lambda es tan abstracto que su aplicación a algo *más acá* de las matemáticas es de difícil contemplación.

No ocurre así con la solución de Turing. El matemático inglés planteó el problema en términos de una máquina hipotética, que más tarde, como bien se sabe, llevaría su nombre. La máquina de la que habla Turing en su solución al problema es una máquina abstracta. Es decir, es imposible crearla en el mundo material. No obstante, esta condición no impidió que dicha máquina se convirtiera en la propulsora de la creación de las computadoras tal y como las conocemos hoy en día. El motivo por el que la máquina de Turing ha sido tan importante para el desarrollo de las máquinas computadoras actuales reside, precisamente, en la propuesta de cómo debe funcionar dicha máquina. Las características de su funcionamiento y su composición son las siguientes[80]:

> - Cinta: Aun cuando las computadoras modernas utilizan un dispositivo de acceso aleatorio con capacidad finita, la memoria de la máquina de Turing es infinita. La cinta, en cualquier momento mantiene una secuencia de caracteres del conjunto de

[76] El problema matemático concreto es conocer si es posible que tal algoritmo pudiera decidir si las reglas de un sistema formal pueden ser demostradas.

[77] Las respuestas de ambos fueron dadas de forma independiente. Pero es de rigor reconocer la posible influencia que el trabajo de Church tuvo en la respuesta de Turing (hasta el punto, por ejemplo, de que el primero fue director de tesis del segundo), si bien no en su enfoque sí que en su línea de investigación.

[78] Para una explicación en detalle del cálculo lambda véase (Penrose, 1991: 100-106).

[79] No deja de ser cierto que el cálculo lambda acabó siendo importante para con la computación práctica, como por ejemplo en la creación lenguaje computacional LISP, que sigue la estructura básica del cálculo de Church (Penrose, 1991: 105-106).

[80] Existen varias versiones de la máquina de Turing, pero la presentada en la cita contiene los rasgos comunes a todas estas versiones.

> caracteres aceptado por la máquina (teniendo la capacidad la máquina de constituir un número muy extenso o muy escueto)[81].
> - Cabeza de lectura/escritura: La cabeza de lectura/escritura en cualquier momento señala a un símbolo en la cinta. Llamamos a este símbolo el símbolo actual. La cabeza de lectura/escritura lee y escribe un símbolo a la vez desde la cinta. Después de leer y escribir se mueve a la izquierda, a la derecha o permanece en su lugar. La lectura, la escritura y el desplazamiento, todos se realizan bajo instrucciones del controlador.
> - Controlador: El controlador es la contraparte teórica de la unidad central de procesamiento (CPU) en las computadoras modernas. Es un autómata de estado finito, una máquina que tiene un número finito predeterminado de estados y se mueve de un estado a otro con base en la entrada. En cualquier momento puede estar en uno de estos estados […].
> Para cada lectura de un símbolo, el controlador escribe un carácter, define la siguiente posición de la cabeza de lectura/escritura y cambia el estado […]. Para cada problema, debemos definir la tabla correspondiente (Forouzan, 2003: 321-322).

A pesar de la sencillez de la máquina, esta tiene la capacidad de realizar un gran número de procedimientos. Sin embargo, aquello que demuestra Turing con la hipótesis de la existencia de esta máquina es precisamente que el modo de proceder (mecánico-algorítmico) no puede dar cuenta del problema de la parada. Es decir, que incluso obteniendo una máquina con capacidad ilimitada en el manejo de algoritmos esta no puede resolver el *Entscheidungsproblem* planteado por Hilbert y Ackermann. Por tanto, la máquina de Turing pone de manifiesto todo el potencial de los algoritmos, pero, a su vez, también alerta de las limitaciones de estos.

Visto lo visto, resulta paradójico que una hipótesis que tenga por conclusión la limitación de los sistemas algorítmicos contribuyera al planteamiento [filosófico] de la posibilidad de una inteligencia artificial. Pero en realidad la paradoja no tiene lugar, ya que los intereses de la investigación de Turing apuntaban en otra dirección[82].

Uno de los temas que marcaron el camino de la investigación del matemático inglés fue (tal y como cree Penrose) el perteneciente a los fundamentos de las matemáticas. ¿Cuál fue el nexo que unió dos materias, en principio, tan distantes? El historiador y filósofo de las matemáticas Ivor Grattan-Guinness concede a la influencia de uno de los mentores de Turing, Max Newman[83], un peso fundamental en el papel de dicho nexo.

El argumento de Grattan-Guinness se basa en que Newman era experto en el debate de los fundamentos de las matemáticas. Y no sólo eso, sino que también contribuyó de manera notable a dicho debate. Newman centró su investigación en topología, siendo pionero en Gran Bretaña en investigar en esta rama de las matemáticas; y también en lógica, teniendo una gran importancia el logicismo de Russell (Grattan-Guinness, 2017: 441). Su interés en estos campos no era fruto de la casualidad. Newman perteneció a

[81] Paréntesis añadido por mí.

[82] Tal dirección era la construcción de una máquina que respondiera a los principios de la máquina de su hipótesis. Aunque en la actualidad se le reconozca de forma unánime el papel en el origen de las máquinas computadoras, esto no fue siempre así. Véase (Randell, 2017: 67-75).

[83] El motivo de destacar el papel de Newman es doble: i) la innegable influencia en Turing (como veremos a continuación); pero también ii) por su relación con Penrose, a través del padre de este, Lionel, quien fue su compañero en su estancia en Viena y, sin duda, fue importante en la vida intelectual de Newman.

la corriente de matemáticos que pensaban en la necesidad de resolver la hendidura entre la lógica y las matemáticas[84] y su investigación estaría ocupada en dicho menester.

Un acontecimiento que fue definitivo en lo que respecta a la influencia de Newman en el enfoque de la investigación de Turing fue un curso que Newman impartió en Cambridge en 1935. Tales cursos tenían como tema principal los fundamentos de las matemáticas, teniendo especial importancia la perspectiva de los intuicionistas brouwerianos. Dentro del tema de los fundamentos de las matemáticas también se exploraban distintos problemas derivados de este debate, como lo era el problema de la decisión y las aportaciones hechas por Gödel unos años antes. A pesar de que Turing no mencionara la importancia de dichos cursos, se presume muy probable que el contacto entre este y Newman se diera para el desarrollo del artículo de 1936 acerca de la computabilidad (Grattan-Guinness, 2017: 439).

Hasta ahora, una parte importante en el debate de los fundamentos de las matemáticas ha quedado relegada a un segundo plano, y esta es la aportación de Gödel al mismo[85]. Tal aportación no es otra que el teorema de incompletitud, que tendrá una gran importancia en el pensamiento de Penrose, tal y como podremos ver en §3.

Tenemos entonces que Kurt Gödel es, junto a Turing y Church, uno de los tres pilares en lo que concierne a la búsqueda del significado y de los límites de la computabilidad (Copeland, 2017a: 57). Si el papel de estas tres personalidades en ocasiones no ha llegado a tener la apreciación merecida es debido a que sus trabajos estuvieron centrados en el plano abstracto y no práctico del término computabilidad. Esto también ha pasado factura a la hora de hacer reconocible la conexión sustancial existente entre la computabilidad y los fundamentos de las matemáticas. La simple (entre comillas) diferenciación del plano práctico y abstracto de «computabilidad» ha sido el obstáculo que ha impedido que tal conexión no fuera manifiesta.

Sea como fuere, lo cierto es que el papel de los tres es innegable. Sin embargo, aquello que defendían dirigía hacia caminos diferentes. La propuesta de Church y Turing sirvió como parapeto para los defensores de la posibilidad de una computabilidad de la consciencia humana, mientras que la de Gödel para los detractores de esta postura.

Como vimos más arriba, Penrose es de esos pensadores que utilizan las ideas de Gödel para debatir los argumentos de los defensores del punto de vista A. Veamos a continuación de qué manera lo hace nuestro autor.

3. Cómo influye el teorema de Gödel en el punto de vista de Penrose

3.1. El teorema de Gödel

[84] Para una referencia recomendable de algunos de estos matemáticos véase asimismo (Grattan-Guinness, 2017:437-438).

[85] El motivo es seguir una línea argumental que conecte con §3.

¿Qué determina el teorema de Gödel que lo hace tan importante dentro del pensamiento de Penrose? Con la intención de no hacer una exposición técnica de los teoremas, lo que veremos serán sus rasgos más importantes con respecto al debate de la computabilidad de la mente y la consciencia humana.

Es conveniente conocer, *grosso modo*, en qué contexto surgió este teorema. El lógico de origen austríaco Kurt Gödel, a los 25 años de edad, presentó los teoremas que llevarían su nombre en respuesta a la tendencia de las matemáticas del momento:

> Como es bien sabido, el progreso de la matemática hacia una exactitud cada vez mayor ha llevado a la formalización de amplias partes de ella, de tal modo que las deducciones pueden llevarse a cabo según unas pocas reglas mecánicas. Los sistemas formales más amplios construidos hasta ahora son el sistema de *Principia Mathematica* (PM) y la teoría axiomática de conjuntos de Zermelo-Fraenkel (desarrollada aún más por J. von Neumann) (Gödel, 2006: 53).

Las matemáticas, de la mano sobre todo de Hilbert, Russell y Whitehead (recordemos que son los que, para Penrose[86], constituyen el bloque formalista), se traducían a sistemas formales (teniendo una importancia capital los sistemas lógicos), pudiéndose resolver todo problema matemático a través de ellos. Esto tenía como consecuencia la concepción de las matemáticas como un saber mecánico:

> Sin embargo, al restaurar así el razonamiento matemático a su estado lógico clásico, Hilbert observó que los operadores lógicos ya no se concebían ni se empleaban de manera semántica o contenciosa como expresiones para operaciones sobre proposiciones significativas. Más bien, se estaban utilizando de una manera puramente sintáctica como parte de un dispositivo computacional-algebraico más grande para manipular fórmulas (Detlefsen, 1996: 80).

Esto, empero, no convencía al joven Gödel:

> Estos dos sistemas son tan amplios que todos los métodos usados hoy en la matemática pueden ser formalizados en ellos, es decir, pueden ser reducidos a unos pocos axiomas y reglas de inferencia. Resulta por tanto natural la conjetura de que estos axiomas y reglas basten para decidir todas las cuestiones matemáticas que puedan ser formuladas en dichos sistemas. En lo que sigue se muestra que esto no es así, sino que, por el contrario, en ambos sistemas hay problemas relativamente simples de la teoría de los números naturales que no pueden ser decididos con sus axiomas (y reglas) (Gödel, 2006: 53-54).

Lo que se proponía con este teorema, por tanto, era tirar por tierra la esperanza de poder explicar las matemáticas a través de un número determinado de axiomas y reglas. Es decir, de volver a tratar la cuestión sobre los fundamentos de las matemáticas. Algo que finalmente consiguió.

Aunque la explicación del teorema gödeliano vaya a ser panorámica es de rigor aclarar algunos conceptos. La finalidad del teorema consiste en que una proposición perteneciente a un sistema formal concreto[87] pueda declararse a

[86] A pesar de que comúnmente aquellos que son conocidos como *formalistas* son los seguidores de Hilbert, mientras que los seguidores de Russell y Whitehead son denominados *logicistas*.

[87] Del modelo ofrecido en los *Principia Mathematica*. El motivo por el que elige este tipo de sistemas formales es para que su crítica esté situada dentro de los sistemas que critica y no desde fuera de ellos.

sí misma como indecidible. Esto tendría como consecuencia la prueba de su indeducibilidad. Y no sólo eso, sino que también podría ser traducible a cualquier sistema formal.

En primer lugar veamos qué es una proposición. Se define proposición a aquellos enunciados que generalmente responden a un valor de verdad. El valor de verdad de una proposición se define de un modo básico, siendo este verdadero cuando la proposición es verdadera y falso cuando esta es falsa (Russell, Whitehead, 1997: 7). Este concepto, por tanto, es de una importancia capital dentro de la lógica[88].

El segundo concepto importante del propósito es la deducibilidad. La deducibilidad se entiende como la capacidad de dar cuenta de la consistencia lógica[89], en términos sintácticos, es decir, de estructura.

Y el tercer concepto, que no aparece en el propósito pero que sí tiene peso, es el de recursividad. Dicho término se corresponde con la capacidad que tiene un procedimiento de definirse a sí mismo. Dentro de los procesos recursivos, aquellos que le interesaban a Gödel para su teorema eran las funciones recursivas primitivas, que son aquellas que se definen a sí mismas cuando sus operaciones principales están compuestas de recursión y composición de funciones. Resulta apropiado que retengamos esta idea de la capacidad de explicarse a sí mismo, porque ella tiene una importancia fundamental.

Volviendo de nuevo al propósito, tenemos que giraba en torno a la idea de que una proposición pudiera dar cuenta de su indecidibilidad. Es decir, que el teorema tiene como meta hacer manifiesto hasta qué punto un sistema formal puede dar cuenta de sí mismo, o más bien, hasta qué punto no puede dar cuenta de ello. Para llegar a este resultado se requerían dos pasos: i) la construcción de dicha proposición y ii) la demostración del carácter indecidible de la proposición[90]. El caso es que Gödel acaba logrando con éxito estos dos pasos[91] y consigue, de este modo, su propósito.

Penrose explica a través de unos sencillos pasos lógicos las implicaciones de dicho teorema de la siguiente forma:

> Hemos numerado todas las funciones proposicionales que dependen de una sola variable, de modo que la que acabamos de escribir debe tener asignado un número. Escribamos este número como k. Nuestra función proposicional es la k-ésima de la lista. Por consiguiente:
>
> $$\neg\exists x\,[\prod x \text{ demuestra } Pw\,(w)] = Pk\,(k)$$
>
> Examinaremos ahora esta función para el valor w particular: w = k. Tenemos:
>
> $$\neg\exists x\,[\prod x \text{ demuestra } Pk\,(k)] = Pk\,(k).$$
>
> La proposición específica Pk (k) es un enunciado aritmético perfectamente bien definido (sintácticamente correcto). ¿Tiene una demostración dentro de nuestro sistema formal? ¿Tiene demostración su negación ¬ Pk (k)? La respuesta a ambas preguntas debe ser «no». Podemos verlo examinando el significado subyacente en

[88] Sobre todo si la lógica que manejamos es bivalente, es decir, aquella que contempla solo la verdad o la falsedad de las proposiciones.

[89] Esta es una propiedad de un sistema formal, la cual consiste, a grandes rasgos, en entender como imposible la aceptación de un sistema concreto y su contradicción al mismo tiempo.

[90] Esta división en dos pasos concretos es propuesta por lógico belga Jean Ladrière para entender de un modo más esquemático el teorema. Para dicho análisis del teorema gödeliano véase su obra *Limitaciones Internas de los Formalismos: Estudio sobre la significación del Teorema de Gödel y teoremas conexos en la teoría de los fundamentos de las matemáticas* (1969).

[91] Para una explicación detallada del teorema véase Gödel (2006: 57-87).

> el procedimiento de Gödel. Aunque Pk (k) es sólo una proposición aritmética, la hemos construido de modo que afirma lo que se ha escrito en el lado izquierdo: no existe demostración, dentro del sistema, de la proposición Pk (k)». Si hemos sido cuidadosos al establecer nuestros axiomas y reglas de inferencia, y suponiendo que hayamos hecho bien nuestra numeración, entonces no puede haber ninguna demostración de esta Pk (k) dentro del sistema. En efecto, si hubiera tal demostración, el significado del enunciado que Pk (k) realmente afirma, a saber, que no existe demostración, sería falso, de modo que Pk (k) tendría que ser falsa como proposición aritmética. Nuestro sistema formal no debería estar tan mal construido como para permitir que se demuestren proposiciones falsas. Por consiguiente, debe ser el caso que, de hecho, no hay demostración de Pk (k). Pero esto es precisamente lo que Pk (k) está tratando de decirnos. Por lo tanto, lo que afirma Pk (k) debe ser un enunciado *verdadero*, de modo que Pk (k) debe ser verdadera como proposición aritmética. ¡Hemos encontrado una proposición *verdadera* que *no tiene demostración dentro del sistema*! (Penrose, 1991: 146-147).

El enunciado de la proposición Pk (k), que está dentro de un sistema formal previamente establecido (y con la forma de los sistemas formales propios de los *Principia Mathematica*) es verdadero. Pero, tal y como hemos visto en la explicación de Penrose, dicho valor de verdad no puede ser demostrado dentro del sistema, al menos sin caer en una contradicción. Tenemos entonces que la recursividad no es una capacidad que le pertenezca a los sistemas formales. Cuando los sistemas formales intentan dar una explicación desde ellos mismos sin caer en contradicción sólo pueden conseguir dar vueltas y vueltas sin llegar a ninguna respuesta concreta. Esta situación es lo que se conoce en lógica como *dialelo* o *círculo vicioso*, que viene a ser un punto de bloqueo del cual no se puede salir.

Un algoritmo tiene la misma estructura y funcionalidad que un sistema formal[92]. Si los sistemas formales no son capaces de dar cuenta de sí mismos, los algoritmos, por su parte, tampoco podrán hacerlo. ¡De hecho es lo que sucede! El ser humano tiene la capacidad de poder dar cuenta de sí mismo o, al menos, el ejercicio de la reflexión no lo conduce a una situación como la del círculo vicioso al que están condenados en ocasiones los sistemas formales. Este es, básicamente, el modo en el que Penrose entiende que el teorema de Gödel puede ser decisivo a la hora de poner en jaque al punto de vista A. Pero veámoslo más detenidamente.

[92] Esta no es una consideración personal. Normalmente está aceptado de este modo. Penrose lo expresa del siguiente modo: Una de las propiedades esenciales de un sistema formal es que debe ser realmente un procedimiento algorítmico (i.e. «computacional») para comprobar si las reglas de (cierto tipo de sistema formal)* han sido o no correctamente aplicadas (Penrose, 2012: 108). De todos modos es conveniente tener cuidado con este tipo de consideraciones, porque, tal y como dice Juliette Kennedy: «Ver la computabilidad en términos de cálculos lógicos implica primero restringir la clase de sistemas formales en los que se representarán las funciones computables» (Kennedy, 2017: 71). Es por cuestiones como esta por las que nuestro autor se apresura en aclarar que dicha equivalencia se debe a su propósito argumentativo (el acceso a las verdades matemáticas) y no como una propiedad evidente entre ambas (Penrose, 2012: 109).
*Paréntesis añadido por mí

3.2. Penrose y Gödel I: cómo adopta Penrose el teorema de Gödel

Penrose no es el primero vez que quiere destacar la potencialidad del teorema de Gödel como base argumental en contra de entender las facultades mentales como meras computaciones. El papel de pionero pertenece al filósofo británico John Lucas, que en el año 1961 usó el teorema de Gödel como apoyo al mentalismo, en su enfrentamiento con el fisicalismo. El mismo Penrose reconoce tal influencia, aunque también reivindica una contribución particular, la cual, considera, supera las dificultades por las que tuvo que pasar el argumento de Lucas (Penrose, 2012: 65).

De todos modos, en un primer momento[93], Penrose trató el teorema de Gödel de manera sucinta, casi de pasada. Sería más tarde[94] cuando consideró que dicho teorema tiene un alcance mayor de lo que en un principio supuso[95]. Nuestro autor utilizará tanto las implicaciones del teorema en sí mismo como lo que este supuso en su contexto.

Como vimos más arriba, el teorema de Gödel surgió para hacer manifiesto que las matemáticas están *más allá* de los sistemas formales. Penrose hace el mismo alegato pero refiriéndose a la consciencia humana y los procesos computacionales.

Uno de las bases sobre las que Penrose sostiene la adopción del teorema de Gödel para sus argumentos en contra de la computabilidad de la mente y la consciencia humana tiene que ver con los límites de los procesos computacionales. Pero ojo, nuestro autor no defiende que tales límites se hacen manifiestos cuando sometemos a las máquinas a procesos complejos. Penrose no tiene problemas en admitir la capacidad de las máquinas para resolver problemas de una complejidad que puedan resultar, incluso, imposible para los seres humanos. Su defensa se basa en que es imposible para una máquina adquirir una consciencia similar a la de los seres humanos. De hecho, nuestro autor piensa que es en los problemas que requieren solución a través del sentido común (humano) donde las máquinas se alejan de forma notable de los seres humanos:

> […] Por el momento, ningún robot controlado por ordenador podría siquiera empezar a competir con un niño pequeño en la ejecución de algunas de las actividades cotidianas más sencillas: por ejemplo, reconocer que un lápiz de colores que está en el suelo en el otro extremo de la habitación es el que se necesita para completar un dibujo, atravesar la habitación para recoger ese lápiz y luego utilizarlo. […] Pero, por el contrario, el desarrollo de potentes ordenadores que juegan al ajedrez proporciona un ejemplo sorprendente en el que los ordenadores pueden ser enormemente eficaces (Penrose, 2012: 60).

Por tanto, la parte «menos brillante» de la consciencia humana sería aquello que comprometería de manera más profunda a las máquinas en su intento por alcanzar a nuestra especie. Penrose es consciente (valga la expresión) de que

[93] En NME.

[94] En SM.

[95] […] Espero que mi exposición servirá para corregir no sólo algunos equívocos aparentemente muy extendidos sobre la importancia del argumento de Gödel, sino también la evidentemente inadecuada brevedad de mi análisis en NME (Penrose, 2012: 65).

su postura no pocas veces ha sido interpretada de forma contraria y por ello aclara:

> […] En efecto, parece que estoy afirmando que el comportamiento computacional tiene que encontrarse en áreas muy complejas del conocimiento matemático, más que en el comportamiento de sentido común. Pero no es esto lo que yo afirmo. Lo que estoy afirmando es que la «comprensión»[96] implica el mismo tipo de proceso no computacional, ya resida en una percepción matemática genuina, digamos la de la infinitud de los números naturales, o resida meramente en la percepción de que un objeto de forma oblonga puede ser utilizado para mantener abierta una ventana, o en la comprensión de cómo podría ser atrapado o liberado un animal con unos pocos movimientos seleccionados de un cabo de cuerda, o en comprender los significados de las palabras «felicidad», «lucha» o «mañana», o en darse cuenta de que cuando el pie izquierdo de Abraham Lincoln estaba en Washington, su pie derecho estaba también en Washington casi con toda seguridad –¡por utilizar un ejemplo que puso sorprendentemente en dificultades a un sistema IA real! (Penrose, 2012: 69).

Ahora bien, ¿qué tiene esto que ver con el teorema de Gödel? Penrose defiende que esta forma de *saltarnos* las reglas computacionales es lo que relaciona de manera directa el teorema gödeliano con la no-computabilidad de la consciencia humana. Recordemos que este teorema nos dice, precisamente, que los sistemas formales no pueden dar cuenta de sí mismos, ya que estos pueden tener reglas adicionales que no están contenidas en ellos, sino fuera. Es decir, que sólo desde un procedimiento no-computacional se podría dar cuenta de tales reglas adicionales. Sabemos que el ser humano cuenta con los procedimientos no-computacionales, pero plantear que una máquina también los posea supone una contradicción flagrante porque, ¡las máquinas deben su comportamiento a procesos computacionales!

A pesar de la conexión entre los argumentos de Penrose y el teorema de Gödel, ello no implica que el pensamiento de ambos autores esté en consonancia. Recordemos que Gödel no descartaba que el pensamiento matemático respondiera a un algoritmo sutil:

> […] Él admitía la posibilidad lógica de que las mentes de los matemáticos humanos pudiesen actuar siguiendo algún algoritmo del que no eran conscientes, o quizá podrían ser conscientes del mismo con tal de que no pudieran estar incuestionablemente convencidos de su validez. […] Gödel […] se vio […] llevado en la dirección mística que yo he designado por D- que la mente no puede explicarse de ninguna manera en términos de la ciencia del mundo físico (Penrose, 2012: 143-144).

¿Quiere decir esto que Penrose fuerce de un modo inadecuado las implicaciones del teorema de Gödel? En mi opinión creo que no. Y por norma general este no es un aspecto que se le suela reprochar a nuestro autor. Otro asunto es el grado en el que se admita como argumento firme. En el punto que sigue veremos en qué modo se critica la postura de Penrose y en qué medida dichas críticas tienen, en mi opinión, peso o no.

[96] Véase nota 12.

3.3. ¿Está obsoleto el argumento Penrose-Gödel en materia de Inteligencia Artificial?

Que el teorema de Gödel pueda ser una herramienta útil para argumentar en contra de los defensores de la Inteligencia Artificial es algo que no parece del agrado de todos. Esta, empero, no es la dirección que tomo. Pienso que la idea de enfrentar el concepto de reflexión y de recursividad es un gran acierto y, considero, que es un argumento filosófico muy potente. Penrose dice con respecto a esto: «Los principios de reflexión proporcionan la propia antítesis del razonamiento formalista. Si se es cuidadoso, nos permiten salir fuera de los rígidos confinamientos de cualquier sistema formal y obtener nuevas intuiciones matemáticas que no parecían disponibles antes» (Penrose, 1991: 151). La reflexión es una característica que parece fuera del alcance de las máquinas.

A pesar de que este argumento permita abarcar cuestiones que en un principio pueden parecer distantes, ello no impide que sea objeto de crítica.

Algunas de las críticas que veremos a continuación están dirigidas tanto a la utilización del teorema de Gödel por parte de John Lucas como a la llevada a cabo por Penrose. Por otro lado también se incluirán críticas indirectas a ambas, pero que guardan una relación íntima con las implicaciones de tales argumentos.

La primera de las críticas pertenece a la llevada a cabo por Russell y Norvig en su obra conjunta, anteriormente citada, *Inteligencia Artificial: un enfoque moderno*. Esta crítica en particular también tiene en cuenta la perspectiva de Penrose, aunque es reconocible que el interés de la misma está centrado en la de Lucas[97]. Russell y Norvig defienden que la adopción del teorema de Gödel es criticable en tres puntos diferentes.

El primero de estos puntos destaca cómo Lucas (Penrose también lo hace) entiende(n) que los procesos computacionales son aquellos que llevan a cabo las máquinas de Turing. Por tanto, hablar de computadoras y de máquinas de Turing es hacerlo de la misma cosa. Si bien esto es fácilmente aceptable, no menos cierto es que esto no debe entenderse de forma absoluta, ya que las máquinas de Turing son infinitas, mientras que las computadoras son finitas. Dicho carácter infinito de las máquinas de Turing implica que estas no estén sujetas a lo que el teorema de Gödel dicta. Y no sólo eso, sino que ello podría trasladarse a cualquier computador (Russell & Norvig, 2004: 1078).

Esta parte de la crítica no tiene el peso que se le pretende dar. Se plantea como si al emplear la adopción del teorema de Gödel no se hubiese tenido en cuenta todas las características de la máquina de Turing. Por su parte, Penrose es bastante claro en este aspecto. De hecho, defiende que su punto de vista (recordemos, C) entra en conflicto con la tesis de Turing y su máquina, mientras que con la tesis de Church (más abstracta) no tiene por qué necesariamente chocar de frente (Penrose, 2012: 35). Penrose sitúa en el mismo escenario la máquina de Turing y el teorema de Gödel, porque, defiende, que Turing mismo lo habría aceptado (Penrose, 2012: 35).

[97] Es llamativo que siendo esta obra más cercana a la de Penrose (siendo SM de 1994 y la obra de Russell y Norvig de 1995, habiendo ediciones posteriores en 2003 y 2009) la crítica la sigan centrando en su mayor medida en el trabajo de John Lucas, unos 30 años más antiguo.

El segundo punto dice que las limitaciones del teorema gödeliano no son ni tan dramáticas para las máquinas ni tan definitivas como argumento. Para hacer manifiesta dicha idea Russell y Norvig defienden que los seres humanos podríamos vernos en un brete similar al que se enfrentan los sistemas formales (y computacionales). Ellos lo expresan del siguiente modo:

> […] Consideremos la sentencia siguiente:
> *J. R. Lucas no puede consecuentemente afirmar que esta sentencia es verdadera.*
> Si Lucas afirmara esta sentencia, entonces se estaría contradiciendo a sí mismo, por tanto Lucas no puede afirmarla consistentemente, y de aquí que esta sentencia sea verdadera. (La sentencia no puede ser falsa, porque si lo fuera Lucas entonces no podría afirmarla consecuentemente, por tanto sería verdadera.) Así pues, hemos demostrado que existe una sentencia que Lucas no puede afirmar consecuentemente mientras que otras personas (y máquinas) sí pueden. Sin embargo, esto no hace que cambiemos de idea respecto a Lucas (Russell & Norvig, 2004: 1078-1079).

Por supuesto que la situación que se describe no cambia la idea con respecto a Lucas, pero también es necesario entender que no se está describiendo una situación real. En cuanto a términos lógicos es cierto, el ser humano correría la misma suerte que un sistema formal o que una computadora, es decir, que se vería acorralado por la sentencia. Pero la diferencia reside en que el ser humano puede dar cuenta de tal laberinto lógico, mientras que un sistema formal o cualquier máquina (por muy potente que sea), no. Si nuestra idea con respecto a Lucas no cambia es precisamente ¡porque no debe hacerlo!

Siguiendo con este mismo punto de la crítica, Russell y Norvig plantean que si la inferioridad que el ser humano tiene con respecto a las máquinas en términos de cálculo rápido no es tenida en cuenta para desprestigiar la inteligencia humana, por qué sí sucede lo contrario con las máquinas y sus limitaciones (Russell & Norvig, 2004: 1079).

Esto de nuevo se encuentra en un escenario distinto al que Penrose plantea su argumento. Hemos podido ver en varias ocasiones que nuestro autor no tiene problemas en admitir las tremendas capacidades *intelectuales* de las máquinas. Este asunto, curiosamente, suele ser un punto común entre los detractores de la perspectiva penroseana. Y todo parece responder a un malentendido. Aquello que defiende Penrose es que es imposible para una máquina que pueda llegar a obtener una consciencia semejante a la del ser humano. Esa es la única capacidad que niega nuestro autor a las máquinas[98] (recordemos del punto anterior cómo las máquinas pasan sus mayores dificultades con problemas que requieren sentido común). Por tanto, intentar insinuar que se le nieguen a las máquinas cualquier tipo de inteligencia es incorrecto. Las computadoras son entidades muy capaces en numerosos ámbitos. Otra cuestión es intentar equiparar tales capacidades a la de los seres humanos. En algunos aspectos los humanos somos superiores a las máquinas y en otros al contrario. Pero lo importante para Penrose no es eso, sino comprender el abismo que nos separa de forma irremediable.

Y el tercer punto de la crítica de Russell y Norvig es básicamente el mismo que la primera parte del segundo punto. Se vuelve a plantear la cuestión de si el ser humano es inmune a las implicaciones del teorema gödeliano y cómo ello no deslegitima ninguna de las dos inteligencias (Russell & Norvig, 2004:

[98] Sin intentar de quitar, claro está, ni la más mínima transcendencia que ello tiene (tal es la transcendencia que es el argumento principal de su trabajo tanto en NME como en SM).

1079). Acabamos de ver en qué modo discrepo con este planteamiento, así que sería un ejercicio fútil volver a desarrollar mis ideas al respecto.

La fuerza del argumento del teorema de Gödel puede verse afectada por otorgarle a este un papel que no le corresponde[99]. Y precisamente de esta acusación no se libra Penrose. Defender que el teorema gödeliano tiene una utilidad fáctica como argumento en contra de la computabilidad de la consciencia y la mente humana es dispensar una carga al teorema que realmente no le incumbe. Pero, ¿de veras esto es así? Es cierto que es indispensable ser cautos a la hora de establecer equivalencias tanto epistemológicas como ontológicas entre argumentos de ámbitos distintos. Esto, en mi opinión, no supone un problema para el planteamiento de Penrose. Ya vimos más arriba que nuestro autor se encarga de aclarar que los sistemas formales y los computacionales están profundamente relacionados, no habiendo, de tal modo, dicho conflicto (ni ontológico ni epistemológico).

Pasemos ahora a ver diferentes críticas que ha recibido la perspectiva de Penrose en concreto. Empezamos con las de Solomon Feferman y Drew McDermott, las cuales pertenecen a una serie de críticas de varios autores, contenidas en la revista *Psyche* entre los años 1995 y 1996[100]. El motivo de la elección de estas dos críticas en concreto es porque, considero, son aquellas que contienen un tono más antagónico con lo expresado por Penrose.

Empezando con Feferman, es de destacar que este autor deja claro desde el principio que el problema del planteamiento de Penrose no es el contenido del mismo, sino la manera en la que pretende defenderlo. Es más, Feferman incluso admite que está de acuerdo con aquello que defiende Penrose.

Uno de los puntos que Feferman[101] destaca es la claridad con la que Penrose ve la relación entre los sistemas formales y las máquinas de Turing. Aunque

---

[99] Douglas Hofstadter, en su obra *Gödel, Escher, Bach: Una eterna trenza dorada*, expresa esta idea en concreto del siguiente modo: Pienso que puede ser sumamente interesante la traducción del Teorema de Gödel a otros ámbitos, a condición de dejar especificado en forma previa que la traducción es metafórica, y que no se proyecta una equivalencia literal (Hofstadter, 1982: 826).

Otros autores más actuales y que se refieren de modo directo a Penrose, como Francesco Berto, dicen lo siguiente: […] o la mente en realidad tiene una naturaleza no algorítmica y no completamente «mecanizable», o bien existen problemas matemáticos absolutamente indecidibles. Pero G1 y G2* no nos permiten ir más allá y concluir que la verdadera disyunción es la primera. Según Gödel, entonces, lo que se desprende de G1, y especialmente de G2, es que si nuestra mente es una máquina informática, es tal que «no es capaz de comprender completamente su propio funcionamiento». Si en realidad solo somos máquinas Turing, entonces no podemos saber exactamente qué máquinas de Turing somos, y esta es una conclusión muy sugerente. Como dijo Paul Benacerraf en "Dios, el diablo y Gödel": «Si soy una máquina de Turing, mi propia naturaleza me impide obedecer el profundo mandato filosófico de Sócrates: CONÓCETE A TI MISMO» (Berto, 2009: 188).

* Este es el modo en el que Berto se refiere a los teoremas de Gödel.

Y, por último, hay quienes incluso llegan más lejos en su argumento con respecto a este tema: «Volviendo al problema del alcance (*scope problem*), aunque el segundo teorema de incompletitud apunta a demostrar la imposibilidad de dar una prueba de consistencia finitaria en los sistemas considerados, esto no resuelve el problema general al cual los sistemas formales se aplican los teoremas de incompletitud» (Kennedy, 2017: 74).

[100] Los números de tales críticas están disponibles on-line en la página de dicha revista en la siguiente dirección (que es aquella que me ha permitido hacerme con ellos): http://journalpsyche.org/archive/volume-2-1995-1996/

[101] He decidido resaltar aquellos aspectos de la crítica de Feferman que no fueron tratados por Penrose en su réplica a la serie de críticas de la revista *Psyche* (1996). En dicha réplica nuestro autor da cuenta, sobre todo (con respecto a esta crítica en concreto), de los

comparte la idea de que la reformulación del teorema de incompletitud como argumento en contra de entender el pensamiento matemático, por otro lado, no deja de cuestionar si tal equivalencia es o no forzada. El motivo de su duda es que piensa que Penrose otorga a los sistemas formales un *modus operandi* que no le corresponde (o al menos no es tan evidente que así sea) y que claramente le aleja de los seres humanos. Para Feferman no está tan claro que los sistemas formales lleven a cabo el pensamiento matemático tal y como lo describe Penrose. De hecho, defiende que en realidad no podemos garantizar el conocimiento de aquello *en concreto* que permite el pensamiento matemático:

> […] La forma en que realmente llegan a la prueba es a través de una maravillosa combinación de razonamiento heurístico, perspicacia e inspiración (basándose, por supuesto, en el conocimiento y la experiencia previos) para los cuales no hay reglas generales, aunque algunos patrones hayan sido discernidos por Polya y otros: *no* existe una fórmula para el éxito matemático. Es solo cuando finalmente se llega a una prueba que se puede verificar (mecánicamente, en principio, pero no en la práctica) la que efectivamente establece el teorema en cuestión (Feferman, 1995: 8-9).

Feferman comparte con Penrose que el *verdadero* pensamiento matemático no es mecánico y que la «comprensión» es definitiva en este apartado, ya que es en esta noción donde las máquinas *podrían* diferenciarse de los seres humanos. Sin embargo, la prueba a través del teorema de Gödel no tiene tanta fuerza como pretende Penrose, sino que esta tan solo traduce una convicción que plantea más preguntas que las que responde (Feferman, 1995: 9).

A pesar de que Feferman lance una crítica muy concreta al planteamiento de Penrose, lo cierto es que no ofrece una alternativa. Aquello a lo que apela Feferman es a la ambigüedad de Penrose a la hora de utilizar ciertos conceptos del campo de la lógica. Tal y como reconoce Penrose, Feferman tiene razón en su crítica, pero la solución pasa por entender tales conceptos en su modo más general para, así, no correr un riesgo desmedido. Conviene recordar que Penrose no pretende entrar en un debate puramente lógico. Con su argumento, nuestro autor solo está ofreciendo un caso que puede entenderse (y de un modo manifiesto) un símil entre los sistemas formales y las máquinas computacionales. Pasemos a ver a continuación la crítica de McDermott.

De la crítica de McDermott[102] cabe destacar que tiene un tono mucho menos delicado que el profesado por Feferman. La diferencia entre ambas actitudes reside, probablemente, en que, a diferencia de Feferman, McDermott no comparte el punto de vista de Penrose con respecto a la Inteligencia Artificial. No obstante, esto no es motivo para que McDermott critique aspectos estrechamente relacionados a los vistos en la crítica anterior.

Un ejemplo de ello es poner en entredicho la defensa de Penrose acerca del modo de proceder de los matemáticos, en lo que a *pensar* las matemáticas se refiere. Para McDermott no está tan claro que entre seres humanos y las máquinas exista ese abismo insalvable. En primer lugar, Penrose comete un error imprudente al generalizar sobre cómo piensan los matemáticos; y en

---

pormenores que destaca Feferman en el aspecto lógico y también acerca de la vaguedad del concepto solidez (*soundness*). Para críticas de la misma naturaleza que la de Feferman véase Lindström (2001) y Shapiro (2003).

[102] Al igual que con la crítica de Feferman, he decidido tratar aquello que Penrose no trata, al menos extensamente, en su réplica (en este caso, a McDermott).

segundo lugar, comete el error de extrapolar el teorema de Gödel a tal supuesto.

La tónica general de la crítica de McDermott es la de reprochar a Penrose su falta de precisión en términos fundamentales para afrontar este tipo de debates. Esto queda reflejado en el último párrafo de su paper:

> […] el computacionalismo (*Computationalism*) es apenas examinado, y mucho menos refutado, por este libro, que apuesta todas las canicas sobre el argumento de la brecha de Gödel y pierde. Una teoría computacional de la conciencia tiene muchos problemas, pero está mejor elaborada que cualquier otra alternativa, incluida especialmente la de Penrose. No es la arrogancia, sino un humilde deseo de verdad, lo que lleva a algunos investigadores a seguir la teoría computacional como una hipótesis de trabajo. El mayor obstáculo para el éxito de esta teoría no es la ausencia de una explicación de la conciencia consciente *per se*, sino el hecho de que la IA todavía ha progresado poco en el problema de la inteligencia general, y ha decidido centrarse en una estrategia más modesta de estudiar habilidades cognitivas individuales. La responsabilidad de la inteligencia artificial es demostrar que este programa de investigación conducirá a una teoría de la inteligencia general. La gente como Penrose debería declarar la victoria y retirarse (McDermott, 1995: 16).

Al igual que sucede con Feferman, en McDermott podemos ver muchos errores en el argumento de Penrose, aunque también muy pocas alternativas. Por un lado, a pesar de que Penrose da un papel muy importante a su argumento gödeliano, no lo deja absolutamente todo en manos de este. Y por el otro, McDermott al fin y al cabo no deja de apelar al «en el futuro veremos qué ocurre» (propio de quienes defienden la IA), con lo que exigir más contundencia a Penrose se antoja una exigencia, al menos, fuera de lugar.

Penrose ya en NME se percata de las posibles críticas a las que su postura tiene que hacer frente. No obstante, considera que los argumentos que puedan sobrevenirle no tienen la fuerza suficiente:

> Sin embargo, para convencernos *realmente* a nosotros mismos de la verdad de P_k (*k*) necesitaríamos *saber* cuál es realmente el algoritmo del matemático, y también estar convencidos de su validez como medio de llegar a la verdad matemática. Como los defensores de la IA fuerte se apresuran en señalar, si el matemático estuviera utilizando un algoritmo muy complicado en su cabeza no tendríamos *ninguna posibilidad* de conocer realmente cuál es dicho algoritmo y, por consiguiente, no seríamos capaces de construir su proposición de Gödel, y mucho menos convencernos de su validez. Este tipo de objeción se plantea a menudo contra las afirmaciones como las que hago aquí, de que el teorema de Gödel indica que los juicios matemáticos humanos son no algorítmicos. Pero yo personalmente no encuentro convincente esta objeción (Penrose, 1991: 517).

El motivo por el que no le convence este tipo de argumentos es porque estos no describen de manera fiel cómo los seres humanos pensamos las matemáticas. Para el ser humano las matemáticas no se le presentan de modo tal que resulte imposible poder dar respuestas por muy abstracto que sea el problema. Pero sobre cómo las matemáticas pueden llegar a ser tremendamente abstractas y aun así poder ser conocidas por los seres humanos volveremos más adelante (§4).

A pesar de los esfuerzos *filosóficos* de los autores vistos, en la actualidad este tipo de críticas no constituyen el argumento principal de quienes defienden la Inteligencia Artificial. Estos, más bien, se centran en expresar las enormes capacidades que las máquinas actuales tienen para, de este modo, tirar por tierra las posibles limitaciones que quieran otorgarles sus detractores.

Un claro ejemplo de esta corriente lo encontramos en el filósofo sueco Nick Bostrom, que en su obra *Superintelligence: Paths, Dangers, Strategies*, destaca el poder de las máquinas del siguiente modo:

> […] Si los métodos que utiliza el software para buscar una solución son lo suficientemente sofisticados, pueden incluir disposiciones para gestionar el proceso de búsqueda en sí de manera inteligente. En este caso, la máquina que ejecuta el software puede comenzar a parecer menos una simple herramienta y más un agente. Por lo tanto, el software puede comenzar desarrollando un plan sobre cómo buscar su solución. El plan puede especificar qué áreas explorar primero y con qué métodos, qué datos recopilar y cómo hacer un mejor uso de los recursos computacionales disponibles. Al buscar un plan que satisfaga el criterio interno del software (como generar una probabilidad suficientemente alta de encontrar una solución que satisfaga el criterio especificado por el usuario dentro del tiempo asignado), el software puede tropezar con una idea poco ortodoxa. Por ejemplo, podría generar un plan que comience con la adquisición de recursos computacionales adicionales y la eliminación de posibles interruptores (como los seres humanos). Dichos planes "creativos" aparecen cuando las habilidades cognitivas del software alcanzan un nivel suficientemente alto. Cuando el software pone en práctica un plan de este tipo, puede producirse una catástrofe existencial (Bostrom, 2014: 153)

La sofisticación del *software* ha llegado hasta tal punto que la inteligencia de las máquinas (superinteligencia, como lo denomina Bostrom) se asemeja en muchos aspectos a la humana (¡incluso hoy en día!). Pero, ¿en qué medida esta corriente puede reflejar tal grado de optimismo sin entrar en conflicto con la perspectiva de Penrose? Pues, según creo, todo lo que quiera. Hemos visto en repetidas ocasiones que nuestro autor reconoce de buen grado los méritos de los avances tecnológicos. Por ello, una descripción como la de Bostrom no supone un argumento en contra de lo que defiende Penrose (esto es, que las máquinas adquieran una consciencia similar a la del ser humano ya que esta contiene procesos no-computacionales).

Ahora bien, ¿cuál es el motivo por el que Penrose adopta el teorema de Gödel sin que esa convicción se vea amenazada por las críticas que esta pueda recibir? En pocas palabras, a las matemáticas. Pero siendo más precisos, a la relación de las matemáticas con la verdad. Nuestro autor entiende que este campo de conocimiento nos permite tener un contacto directo con la verdad. Por tanto, ¡este camino no debe ser abandonado!

4. Matemáticas: la naturaleza de los números y su relación con el mundo físico

4.1. Problemas y números

¿Qué hace de las matemáticas una herramienta tan útil y eficaz? Entre las muchas facultades que posee, esta disciplina admite una condición que la caracteriza entre las demás, esta es, la demostración.

Más arriba vimos que se sabe con exactitud que en Egipto y Babilonia ya se hacía uso de las matemáticas, pero que este tipo de matemáticas no son las que interesan a Penrose, ya que estas eran usadas para su aplicación práctica. Para nuestro autor, las utilizadas por Tales de Mileto y Pitágoras de Samos,

por ejemplo, son sensiblemente diferentes, porque con ellas (a través de la introducción de la *demostración matemática*) comienza realmente la tradición del pensamiento científico tal y como la conocemos (Penrose, 2006: 50).

A Tales se le atribuyen ciertas medidas y teoremas relacionados con la geometría. Por ejemplo, Diógenes Laercio decía:

> Jerónimo afirma que (Tales) midió también las pirámides por su sombra, tras haber observado el momento en que nuestra sombra es igual a nuestra altura (Diógenes, cit. por Kirk et al., 2008: 122).

Proclo, por su parte, destacaba:

> Eudemo atribuye este teorema a Tales en la historia de la geometría; pues dice que es necesario que lo utilizara para la explicitación del método mediante el que dicen que demostró la distancia de las naves en el mar (Proclo, cit. por Kirk et al., 2008: 122).

Este tipo de aportaciones, sin duda, hace manifiesto que las matemáticas que Tales llegó a manejar tenían cierta profundidad (en el sentido visto más arriba). Pero no menos cierto es que dichas matemáticas eran aún rudimentarias, ya que ellas respondían *ligeramente más allá* del ámbito práctico de las mismas (Kirk et al., 2008: 114-115).

Por otro lado tenemos a Pitágoras[103]. De este autor se sabe que en sus enseñanzas los números tienen una importancia fundamental, allende el plano práctico. De hecho, Penrose siente un interés particular por las aportaciones de los pitagóricos, ya que entiende que estas posibilitan una idea muy clara de la relación entre la verdad matemática y el mundo físico.

Pero antes de emplazar a las matemáticas en el mundo físico Penrose trae a colación un principio de demostración matemática del cual se obtienen numerosos resultados fructíferos: el principio de demostración por contradicción. El tratamiento de este principio lo lleva a cabo en EcalR, que se aleja un poco del debate de la computabilidad de la consciencia humana, pero no de lo que estamos viendo en este capítulo.

Este principio consiguió dar respuesta, por ejemplo, al problema matemático de encontrar un número racional multiplicado por sí mismo diera como resultado 2. Veamos cómo.

En primer lugar, es conveniente aclarar, al menos de forma breve, los distintos conceptos básicos que se manejarán de aquí en adelante en este apartado. Tales conceptos son los números «naturales», números «enteros», números «racionales», números «irracionales» y números «reales».

Por números «naturales» entendemos aquellos que son mayores que 0 (en algunas ocasiones el 0 se incluye en este grupo), pero no en forma de fracción, ni decimal. Este tipo de números son los que nos permiten contar elementos,

[103] Aunque haya encapsulado la corriente en torno a la figura de Pitágoras, es más correcto aclarar que cuando me refiera a este lo estoy haciendo a los pitagóricos en general (y con ello aludo a lo más general posible, no entrando en el debate que existió entre los llamados pitagóricos acusmáticos –de quienes se decían que sus prácticas intelectuales procedían directamente de Pitágoras- y los reconocidos como matemáticos –que las recibían de Hipaso (Kirk, Raven, Schofield, 2008: 313-314). Es decir, que al nombrar a los pitagóricos o Pitágoras estaré apuntando hacia todos ellos), ya que muy poco se sabe con certeza acerca del filósofo de Samos y de la corriente en general.

siendo los que más relación directa guardan con la vida cotidiana del ser humano. Los números «enteros» comprenden los números naturales, el cero y sus opuestos (negativos). Los números «racionales» son aquellos que se representan como fracción tipo *a*/*b*, donde *a* y *b* son números «enteros» y *b* es distinto de 0. Este tipo de número no es consecutivo, tal y como sí lo son, por ejemplo, los números «naturales». Si bien los números «racionales» son aquellos que se expresan mediante fracciones, los «irracionales» tienen la característica de no poder ser expresados de tal forma, ya que sus decimales, los cuales no siguen las pautas necesarias requeridas, no hacen posible dicha expresión. La expresión decimal de los números «irracionales» no es ni periódica ni exacta. El ejemplo más característico de este tipo de números es el valor de pi (π). Y por último, los números «reales», por su parte, son aquellos que comprenden a todos los anteriores, es decir, a «naturales», «enteros», «racionales» e «irracionales».

A modo de axioma, también es conveniente esclarecer dos aspectos acerca de los números. El primero es que el cuadrado de un número impar es un número impar y si un entero no es impar es par. Y el segundo es que cada secuencia estrictamente decreciente de números enteros positivos debe llegar a un final[104].

Una vez visto los conceptos principales y sus aspectos pasemos a ver un problema que fue objeto de estudio de los pitagóricos y que Penrose retoma. Dicho problema es encontrar la solución de $\sqrt{2}$. Bien se sabe que la respuesta al problema es que no existe tal número («racional») cuyo cuadrado tenga como resultado 2.

Para los pitagóricos esto supuso un desastre, ya que estos tenían «la esperanza de que toda su geometría podría expresarse en términos de longitudes que podrían medirse en términos de números racionales[105]» (Penrose, 2006: 105).

El problema se hace patente precisamente en el teorema de Pitágoras, y Penrose lo expone (2006: 106) presentando un cuadrado de lado longitud unidad (es decir, en el que sus lados miden 1), donde su diagonal –aplicando el teorema- es la suma de esos lados al cuadrado, teniendo como resultado $1^2 + 1^2 = 2$. ¡El resultado es dramáticamente erróneo, ya que ello deja la respuesta a la diagonal de este cuadrado en suspenso!

[104] Estos aspectos vienen muy al caso a tener en cuenta en los distintos problemas que vendrán en el desarrollo de este apartado.

[105] Todo lo que no supusiera harmonía entablaba un problema para los pitagóricos. Sexto Empírico describe esta actitud del siguiente modo: Y, al indicar esto, los pitagóricos, a veces, tenían la costumbre de decir «todas las cosas son semejantes al número» y de jurar, a veces, su juramento más poderoso así: «No, por aquel que nos dio la *tetractys*, que contiene la fuente y la raíz de la siempre fluyente naturaleza». Con la expresión «aquel que dio» querrían decir Pitágoras (pues le deificaban) y con el término «la *tetractys*», un número, que, por estar compuesto de los cuatro primeros números, causa el número más perfecto, como p.e., el diez (porque uno, dos, tres y cuatro son diez). Este número es la primera *tetractys* y es llamada «fuente de la siempre fluyente naturaleza», puesto que el universo entero es gobernado harmónicamente y la harmonía es un sistema de tres concordancias, la cuarta, la quinta y la octava y las proporciones de estas tres concordancias se encuentran en los cuatro números mencionados –en el uno, dos, tres y cuatro- (Sexto Empírico, cit. por Kirk et al., 2008: 312).

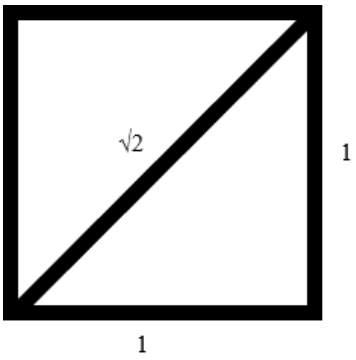

Figura 2. En esta ilustración podemos contemplar el error al aplicar el teorema de Pitágoras.

Los pitagóricos intentaron hacer frente a esto con la noción de «número en acto» que pudiera describirse en términos de razones de números «enteros» (Penrose, 2006: 106), pero esta solución no funcionaría.

Penrose lo explica de forma clara, a través de la demostración por contradicción. El asunto es ver por qué[106] $(a/b)^2 = 2$, no tiene solución para «enteros» que se consideran positivos (Penrose, 2006: 106). Como la herramienta que utiliza el matemático inglés es la demostración por contradicción, lo que sigue es suponer que la ecuación es verdadera, es decir, que existen tales a y b como números «enteros». Una vez tomado este supuesto multiplicamos la ecuación por b^2, dando como resultado $a^2 = 2b^2$. Esto nos lleva necesariamente a que $a^2 > b^2 > 0$. De $2b^2$ sabemos que es par, siguiéndose, por tanto, que a también debe ser par. Entonces, tenemos un «entero» positivo c tal que $a = 2c$. El siguiente paso sería sustituir en la ecuación anterior a por este equivalente y elevarlo al cuadrado, obteniendo $4c^2 = 2b^2$. Si dividimos los miembros de la última ecuación por 2 tenemos como resultado $2c^2 = b^2$, llevándonos necesariamente a que $b^2 > c^2 > 0$. Si nos fijamos bien, este resultado es el mismo de antes, sólo que podemos cambiar la b por la a y la c por la b, lo cual significa que este proceso puede repetirse indefinidamente. Habiendo supuesto previamente que hablamos de que estos miembros son números «enteros» positivos encontramos un problema, y es que si tienen esta condición por definición una serie decreciente de este tipo de números debe tener un final, pero ¡ya hemos visto que puede repetirse indefinidamente! Al obtener esta contradicción nos vemos forzados a rechazar el supuesto inicial, que era admitir la existencia de un número «racional», el cual su cuadrado tiene como respuesta 2.

Si bien todo el proceso parece estar totalmente dirigido, Penrose sólo se limita a realizar los pasos que le están permitido acorde a las condiciones que los pitagóricos pusieron encima de la mesa, a saber, que los números manejados tenían que ser «enteros» positivos. Como se ha demostrado, esta condición no permite encontrar una respuesta a la pregunta de la raíz cuadrada de 2. Pero, ¿realmente no tiene respuesta afirmativa? ¿El misterio de la raíz cuadrada de 2 es tal que permanecerá oculto para siempre? Obviamente la respuesta es no. Basta con echar la mirada atrás en nuestras vidas y recordar que en la escuela hemos manejado el número que responde a la pregunta, y

[106] Como aclaro más arriba, utilizo la notación y los pasos que sigue Penrose (2006: 106-107).

efectivamente este no se correspondía a un número racional positivo, sino a los denominados como números «reales». Los números «reales» tienen la diferencia fundamental con respecto a los «enteros» es que una sucesión de «reales» positivos de manera decreciente no tiene que llegar forzosamente a un final (pongamos por ejemplo la secuencia 1/2, 1/4, 1/8, 1/16..., la cual puede seguir dándose hasta el infinito). Además, los números «reales» no plantean un problema a la hora de saber si es «par» o «impar», ya que *todos* los números «reales» tienen que contar como «pares» (Penrose, 2006: 109).

Después de descubrir que se puede dar una respuesta pero con números diferentes, esto condujo a los matemáticos a querer saber más acerca del alcance de los números «reales». Al contrario de la explicación de la imposibilidad de encontrar un número «entero» tal que su cuadrado sea 2, no entraremos en los detalles que Penrose ofrece acerca de las características de los números «naturales», ya que ello sólo lograría extender más de lo necesario una explicación de la que podemos prescindir. De todos modos, Penrose reconoce que los mismos griegos[107] ya tenían alguna idea del alcance de este tipo de números, aunque obviamente no del modo en el que se tiene hoy en día (la definición número «natural» como tal es bastante reciente). La diferencia fundamental entre la concepción de número «natural» griego y la moderna reside en la relación que guardan dichos números con el mundo físico.

Para los griegos un cuadrado dibujado o un cubo esculpido podrían haber sido considerados como una aproximación razonable o incluso excelente al ideal geométrico (aunque hay que tener en cuenta que siempre se concebiría como una aproximación y no como una demostración, en el sentido de que las figuras en sí mismas no se *manifiestan*). Mientras que para los modernos (entendamos que estos son principalmente los matemáticos del siglo XIX) las matemáticas podrían tener independencia del mundo físico, ya que se había descubierto que existían geometrías matemáticamente consistentes diferentes de la de Euclides (Penrose, 2006: 114).

Las matemáticas pasaron a ser tal y como la concebimos[108] una vez abandonó por completo su papel como herramienta y tomó una naturaleza propia, independiente del mundo físico.

4.2. El mundo físico y más números

Las matemáticas griegas, como hemos visto, aunque cada vez precisaran de una *profundidad* mayor no dejaban de estar enfocadas (e incluso subordinadas) hacia el ámbito práctico, con lo que su relación con el mundo físico era casi trivial. Caso aparte era Platón. Bien se sabe que el filósofo

[107] Penrose ya no se refiere a los pitagóricos. Y esto podría ser problemático si se pretende ser riguroso en el plano de las influencias filosóficas. Estrictamente hablando, Penrose en ese preciso apartado no habla de los seguidores directos de Pitágoras. A quien se menciona es a Eudoxo, quien era discípulo de Platón. Como bien se sabe, Platón sentía un gran respeto por las ideas pitagóricas. Así que apear a Eudoxo de ser pitagórico de manera categórica no es del todo correcto.

[108] Parafraseando a Juan Arana: […] no solo como ciencia de la cantidad, sino como ciencia de las relaciones formales abstractas en general (Arana, 2012: 120).

griego reconocía la independencia en la existencia de las matemáticas con respecto mundo físico[109]. Pero incluso en la perspectiva de Platón no existía una independencia plena, ya que las matemáticas ejercían influencia en el mundo físico. En cierto sentido, las matemáticas tenían su *razón de ser* en influir en el mundo físico.

Algo diferente sucedería con las matemáticas del siglo XIX. Estas conservaban la idea platónica de que las matemáticas eran independientes del mundo físico, pero añadirían que la idea influir de alguna manera para la practicidad del mundo físico es prescindible. Parte importante de esta concepción la tienen determinados números, que si bien pueden encontrarse algunas de sus características de algún modo muy sutil en el mundo físico (Penrose, 2006: 120) aquello que realmente los define es su posibilidad de independencia con respecto a este.

Penrose, por su parte, destaca el papel de los números «reales». Para el caso de este tipo de números resalta la figura de Richard Dedekind (y también la de Georg Cantor, aunque en un grado menor[110]). El trabajo de Dedekind, sin duda, marca un hito en el ámbito de las matemáticas, sobre todo porque este supone una pieza fundamental en la interacción entre estas y el ámbito filosófico[111] (Avigad, 2006: 160). En concreto, su modo de entender los números «reales», al igual que el de Cantor, es peculiar, ya que expone el conjunto teórico de este tipo de números a través del campo de los números «racionales» (Ferreirós, 2016: 229). Esta concepción allanó el terreno para posteriores investigaciones acerca de tales números, no limitándose tal influencia al ámbito matemático. Veamos cómo lo explica Penrose en EcalR:

> […] Lo que se hace, básicamente, es considerar que los números racionales, tanto positivos como negativos (y el cero), están dispuestos en orden de tamaño. Podemos imaginar que este ordenamiento tiene lugar de izquierda a derecha, considerando que los racionales negativos se extienden indefinidamente hacia la izquierda, y los racionales positivos se extienden hacia la derecha, estando el 0 en el centro. (Esto es solo para propósitos de visualización; de hecho el procedimiento de Dedekind es completamente abstracto.) Dedekind imagina un «corte» que divide esta disposición […] en dos […]. Cuando el «filo del cuchillo» que hace el corte no «incide» sobre un número racional sino que cae entre ellos, entonces decimos que define un número real *irracional*. […] Cuando se añade el sistema de los «irracionales», definidos en términos de tales «cortes», al sistema de los números racionales que ya teníamos, se obtiene la familia completa de los *números reales* (Penrose, 2006: 115).

[109] Otra de las grandes diferencias entre los griegos y los matemáticos del siglo XIX es que mientras que los primeros centraban su atención en la geometría, los segundos lo haría en la aritmética: […] la idea de que la matemática pura es totalmente autónoma del mundo físico es antigua, típica de los neoplatonistas. Los matemáticos alemanes del siglo XIX combinaron este tipo de orientación racionalista-neoplatonista con el cambio de la geometría a la aritmética (Ferreirós, 2016: 227).

[110] Esta no es una consideración personal. Más bien es una deducción que hago de la consideración de Penrose (por tal y como trata a una y otra figura con respecto al asunto de los números «reales»).

[111] Y no sólo eso, sino que también constituye una de las corrientes con prerrequisito formal mínimo para un fundamento neo-fregeano (algo que pretendían las corrientes emergentes de la época). Estas corrientes, según Hale, «deben encontrar principios de abstracción presuntamente consistentes que, nuevamente en conjunción con una lógica adecuada -presumiblemente segundo orden-, sea suficiente para la existencia de una serie de objetos que se comporten colectivamente como los números reales clásicos; esto es, que compongan un campo completo y ordenado». La propuesta de Dedekind se aproxima de forma atractiva a este fin (Hale & Wright, 2005: 186).

A pesar de que estos «cortes» permitan la entrada en escena de los números «reales», es evidente que su concepción es altamente abstracta. De hecho, antes de que aparecieran los esquemas de Cantor y Dedekind la idea de los números reales y sus operaciones carecían de sentido (Ferreirós, 2016: 230). Pero con estos se estaba dando paso a unas matemáticas *puras*, unas matemáticas con una independencia –valga la expresión- real. De repente, el mundo físico carecía de toda importancia[112].

La idea de que cierto tipo de matemáticas abstractas (las puras en concreto) prueben su independencia total del mundo físico no seduce demasiado a Penrose. Esto no debe entenderse erróneamente. A estas alturas volver a repetir la simpatía de nuestro autor por el platonismo (y con ello de la independencia de las matemáticas) carece de sentido. No obstante, la perspectiva de Dedekind le resulta extrema.

De hecho, para Penrose los números «reales» no están alejados del mundo físico. Una vez estos números se asentaron en las matemáticas también lo hicieron en física, y con ello las descripciones de la naturaleza se volvieron más precisas:

> […] En la época de Euclides había escasa evidencia para apoyar siquiera la pretensión de que tales «distancias» euclídeas se extendían hacia fuera hasta más allá de, digamos, unos 10^{12} metros, o hacia dentro, hasta algo tan pequeño como 10^{-5} metros […] En efecto, nuestras modernas y satisfactorias teorías cosmológicas nos permiten ahora ampliar el rango de nuestras distancias en números reales hasta aproximadamente 10^{26} metros o más, mientras que la precisión de nuestras teorías de la física de partículas extiende este rango hacia dentro, hasta 10^{-17} metros o menos […] Puede considerarse una notable justificación de nuestro uso de idealizaciones matemáticas el hecho de que el rango de validez del sistema de los números reales se ha ampliado desde un total de aproximadamente 10^{17}, desde lo más pequeño a lo más grande, que parecía adecuado en la época de Euclides, hasta al menos 10^{43} que nuestras teorías actuales utilizan directamente, lo que supone un extraordinario incremento en un factor de 10^{26} (Penrose, 2006: 117-118).

Penrose no niega el alto grado de abstracción que requieren los números «reales», pero de ahí a que estos estén completamente alejados del mundo físico no responde a los resultados que las mismas matemáticas ofrecen.

Con motivo de seguir exponiendo una contraposición a la idea de que las matemáticas puras sólo pueden ser entendidas como *absolutamente* independientes del mundo físico, Penrose toma como ejemplo otro tipo de números sumamente abstractos pero que guardan relación con el mundo físico: los números «complejos».

Nuestro autor mismo reconoce que la relación no es muy evidente (al modo práctico griego), pero defiende que si se hace el esfuerzo de comprender tales números se puede detectar. Por ahora veamos cuáles son los números «complejos».

La utilización de los números «complejos» surge de otra pregunta matemática, la cual, en primera instancia, parece no tener una respuesta satisfactoria. Tal pregunta es si es posible encontrar la raíz cuadrada de -1. La

[112] Es de rigor apuntar que Dedekind era más radical que Cantor en su exposición acerca de la idea de la independencia de las matemáticas con respecto al mundo físico (Ferreirós, 2016: 229). La diferencia entre ambas perspectivas se encuentra en la postura filosófica de ambos. Mientras que Cantor es considerado como perteneciente al platonismo (Ferreirós, 1999: 451, 448-449), Dedekind es incluido dentro del logicismo (Ferreirós, 2009: 40).

aparente insolubilidad está igualmente relacionada con aquello que ocurría en el planteamiento de dar *a través de un número entero* con la raíz cuadrada de 2. Es decir, en tanto que planteamos la pregunta en términos de números «enteros», por ejemplo, volvemos irremediablemente a no poder dar con una respuesta. Lo conveniente, por tanto, es emplear un sistema mayor de números que el de los números enteros para, de este modo, *buscar más allá*[113].

El procedimiento que da lugar a la solución del problema fue propuesta de forma relativamente temprana (en el siglo XVI), de la mano de Gerolamo Cardano primero y más tarde por Raffaele Bombelli (Penrose, 2006: 131). Ello, empero, no significó que su uso fuera inmediato, ya que los resultados fueron tratados con desconfianza[114] (Penrose, 2006: 126- 127).

El proceso que da paso a los números «complejos» es el siguiente: introducir una cantidad, llamada «i» cuyo cuadrado sea el resultado que estamos buscando (en este caso, -1). Esta cantidad se añadiría al sistema de los números «reales», haciendo posible de esta manera el combinar ambos elementos, obteniendo expresiones tales como $z = x + iy$, donde x e y son números «reales» arbitrarios. Dichas expresiones son conocidas como de *forma binómica*. Esta está compuesta de:

a) la *parte real*, el número «real» x, denotado como $Re\ (z)$,

b) la *parte imaginaria*, el número «real» y, denotado como $Im\ (z)$.

Por lo que tenemos que: $z = Re\ (z) + iIm\ (z)$.

En lo referente al conjunto de los números «complejos» tendríamos el siguiente:

$C = \{z = x + i.\ y;\ x, y \in \mathbb{R}\};\ Re\ (z) = x;\ Im\ (z) = y$

Por tanto, los números «reales» se entienden como un subconjunto de los números «complejos». Por otro lado, un número «complejo» de parte imaginaria nula sería «real» (es decir, números «complejos» de la forma $z = x + i.\ 0$ son números reales). Por su parte, se denominan números «imaginarios» aquellos con la forma $0 + i.\ y$ (es decir, con su *parte real* nula).

La última característica destacable de este tipo de números es que dos números «complejos» (tales como $z_1 = x + iy$ y $z_2 = u + iv$) son iguales si y sólo si tienen iguales sus partes reales y sus partes imaginarias ($x = u$, $y = v$)[115].

Penrose además expone algunas de las posibles combinaciones y cálculos[116] que se pueden llevar a cabo con los números «complejos» para hacer a estos más *asequibles*. No obstante, aquello que queda claro que este este tipo de números es igualmente abstracto que los reales o incluso más. No es así, como lo entiende nuestro autor. Allí adonde el ojo inexperto sólo ve oscuridad, Penrose ve un camino necesario para poder llegar a la luz:

> Presumiblemente, el motivo de este recelo era que la gente no podía «ver» que los números complejos se les presentasen de un modo obvio en el mundo físico. En el

[113] Por otra parte, la pregunta en su origen no está planteada concretamente en términos de números «enteros». Este es un ejemplo que he tomado para que su comprensión sea más asequible, ya que el caso de los límites de los números «enteros» lo pudimos ver en el problema anterior.

[114] No sería hasta Gauss cuando este tipo de números adquirieron un estatus diferente. Sobre el papel de Gauss con respecto a los números complejos véase Goldstein et. al [eds.] (2007).

[115] Esta explicación de los números «complejos» es tomada de Jorge Muñoz y Paco Moya, en una serie de apuntes on-line llamados *Libros Marea Verde*.

[116] Véase Penrose, (2006: 131-138).

> caso de los números reales existía la sensación de que las distancias, los tiempos y otras magnitudes físicas proporcionaban la realidad que tales números requerían [...]. Pero habría que recordar que la conexión que tienen los números reales matemáticos con aquellos conceptos físicos de longitud o tiempo no es tan clara como habíamos imaginado […]. Se podría decir que los números reales son tan producto de la imaginación de los matemáticos como lo son los números complejos. Pese a todo encontraremos que los números complejos, tanto como los reales, y quizá más incluso, componen una notable unidad con la naturaleza. Es como si la propia naturaleza estuviera tan impresionada por el alcance y consistencia del sistema de los números complejos como lo estamos nosotros, y hubiera confiado a estos números las operaciones detalladas de su mundo en sus escalas más minúsculas (Penrose, 2006: 133-134).

La estima que nuestro autor siente por los números «complejos» se debe a que con la *aparición* de estos muchos de los problemas que no tenían solución comenzaron a tenerla. ¿Se trata de una cuestión de magia inexplicable o la magia estaba contenida en estos números?

4.3. La magia de los números complejos

Este punto estará centrado en la explicación de una de las singularidades de las que son propietarias las series de potencias, como lo es la cualidad de ser divergente o convergente y qué relación guarda esto con la magia (expresión de Penrose) de los números «complejos».

Para empezar, las series de potencias son aquellas de tipo $a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots$, siendo esta en particular divergente si esta suma de términos tiene que hacerse infinita y convergente si la suma tuviera un valor límite. Para seguir con los mismos términos que utiliza nuestro autor, el siguiente paso que realiza es considerar la misma serie pero planteada de modo tal que sea $1 + x^2 + x^4 + x^6 + x^8 + \ldots$, siendo los valores $a_0 = 1$, $a_1 = 0$, $a_2 = 1$, $a_3 = 0$, $a_4 = 1$, $a_5 = 0$, $a_6 = 1,\ldots$, con respecto a la serie de potencias anteriormente vista. Penrose propone a continuación dar valores a x tales como 1, 2 y 1/2. Para[117] $x = 1$ tenemos que la serie es divergente y para $x = 2$ también, mientras que para $x = 1/2$ tenemos que la serie es convergente.

Hasta aquí no parece haber ningún problema, pero nuestro autor plantea escribir la «respuesta» a la suma de toda la serie, obteniendo $1 + x^2 + x^4 + x^6 + x^8 + \ldots = (1 - x^2)^{-1}$, y con ello surge que si bien para $x = 1$, la respuesta dada sigue ofreciéndonos que este valor hace que la serie sea divergente[118], y para $x = 1/2$, convergente[119]. Sin embargo para $x = 2$ parece que el resultado, al menos en principio, plantea otro tipo de problemas:

> […] Pero, ¿qué pasa con $x = 2$? Ahora hay una respuesta dada por la fórmula explícita, a saber, $(1 - 4)^{-1} = -1/3$, aunque no obtengamos dicho valor sumando simplemente los términos de la serie. Difícilmente podríamos obtener esta respuesta porque estamos sumando cantidades positivas, mientras que -1/3 es negativo. La razón de que la serie diverja es que, cuando $x = 2$, cada término es mayor que el

[117] Para los detalles de las distintas sustituciones véase EcalR (2006: 139).

[118] El resultado es $(1 - 1^2)^{-1} = 0^{-1}$, siendo «infinito», haciendo manifiesta la divergencia de la serie (Penrose, 2006: 139).

[119] El resultado es $(1 - 1/4)^{-1} = 4/3$, cuyo valor concreto muestra su convergencia (Penrose, 2006: 139).

término correspondiente para x = 1, de modo que la divergencia para $x = 2$ se sigue lógicamente de la divergencia para $x = 1$. En el caso $x = 2$ no se trata de que la respuesta sea realmente infinita, sino de que no podemos llegar a esta respuesta intentando sumar la serie directamente (Penrose, 2006: 139-140).

Es decir, que la divergencia resultante de $x = 2$ se entiende a través de $x = 1$ y no a través de $x = 2$ directamente. Esto, obviamente, resulta problemático, ya que la primera consideración que podemos hacer acerca del resultado que se obtiene a partir de $x = 2$ es que su solución (recordemos, $(1 - 4)^{-1} = -1/3$) es «sin sentido». Pero Penrose nos anima a que no ignoremos a «lo que sucede fuera de escena», ya que incluso es posible dar un sentido matemático a la respuesta «-1/3» de la serie infinita obtenida a partir de $x = 2$. Ello se debe a que el sentido que podemos ser capaces de dar a este tipo de expresiones que en principio no tienen [precisamente] «sentido» es a que en muchas ocasiones estas obedecen a las propiedades de los números «complejos» (Penrose, 2006: 142).

Pero volviendo a la convergencia, Penrose considera ahora una «respuesta» a la suma de toda la serie con una sutil diferencia. Si antes obteníamos $(1 - x^2)^{-1}$, ahora pasamos a considerar que obtenemos $(1 + x^2)^{-1}$, para de esta forma comprobar si se tiene un desarrollo en serie de potencias razonable (Penrose, 2006: 142). Este supuesto tiene como consecuencia que ya la serie no sería una suma, sino que se daría una alternancia entre suma y resta (quedando, por tanto, de modo tal que $1 - x^2 + x^4 - x^6 + x^8 - \ldots = (1 + x^2)^{-1}$).

Realizando el mismo paso que en el anterior caso, es decir, dar valores a x, tales como 1, 2 y 1/2, y tenemos[120] que en el caso de $x = 1$, su divergencia consiste en que las sumas parciales de la serie no se asientan, y no a que la serie muestre una secuencia infinita (Penrose, 2006: 143). El motivo de este problema vuelve a ser que estamos manejando números «reales», y por ello es preciso echar mano, de nuevo, de los números «complejos».

Detectar el alcance de los números «complejos» no es ni mucho menos una tarea fácil, y para hacer esta tarea de un modo más ilustrativo, Penrose trae a colación el plano complejo de Caspar Wessel. Con este plano se puede hacer interpretaciones geométricas de distintas operaciones con números «complejos». Ello parece no tener una relación directa con los distintos problemas que habían suscitado las series de potencias, pero nuestro autor demuestra que en el fondo sí que lo hace y que, incluso, puede hacerse manifiesto que los resultados que se obtienen a través de ello tienen un alcance más *profundo*.

Para ello comienza tomando las funciones de la variable «real» x que habíamos visto anteriormente, es decir, $(1 - x^2)^{-1}$ y $(1 + x^2)^{-1}$, para extenderlas de modo tal que se apliquen a una variable «compleja» z, quedando de forma tal que $(1 - z^2)^{-1}$ y $(1 + z^2)^{-1}$ (Penrose, 2006: 144). Al realizar esto ocurre que el problema de poder ver la divergencia con $(1 + x^2)^{-1}$, al pasar a considerarlo en términos de variable «compleja» z parece que dicho problema no existe, ya que las funciones parecen ser más semejantes (Penrose, 2006: 145). El mismo Penrose sabe que aún no queda claro del todo su argumento, y por ello continúa:

[120] Los resultados nos los ofrece Penrose (2006: 142), y son los siguientes:
$x = 1$: 1, 0, 1, 0, 1, 0, 1, etc.,
$x = 2$: 1, -3, 13, -51, 205, -819, etc.,
$x = 1/2$: 1, 3/4, 13/16, 51/64, 205/256, etc.

> […] Ahora debemos considerar nuestras series de potencias como funciones de la variable compleja *z*, en lugar de la variable real *x*, y podemos preguntar por aquellas localizaciones de *z* en el plano complejo para las que la serie converge y aquellas para las que diverge. La notable respuesta general para cualquier serie de potencias
>
> $$a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \ldots,$$
>
> es que existe un círculo en el plano complejo, centrado en 0, llamado círculo de convergencia, con la propiedad de que si el número complejo *z* está estrictamente dentro del círculo, entonces la serie converge para dicho valor *z*, mientras que si *z* está estrictamente fuera del círculo, entonces la serie diverge para dicho valor *z* (Penrose, 2006: 145).

Tenemos entonces que en el plano «complejo» la explicación sobre la convergencia y divergencia de las funciones tratadas, entendidas en términos de la variable «compleja» *z*, queda perfectamente aclarada y no sufre esa –llamémosla así- deficiencia que sí tenían dichas funciones cuando las manejábamos en términos de variables «reales» *x*. Como consecuencia, el alcance de los números «complejos» es incluso más patente que el de los números «reales», a pesar del grado de abstracción que se necesita para poder dar cuenta de ello.

Lo que Penrose quiere exponer con este tipo de explicaciones es que a pesar de que las matemáticas parezcan estar completamente alejadas del mundo físico, de algún modo siguen pudiendo explicar los fenómenos de la naturaleza. Y no sólo eso, sino que las matemáticas más abstractas tienen un poder de descripción mayor que las más básicas. Por tanto, las matemáticas (todas ellas) son imprescindibles para la obtención de respuestas que se adecúen mejor a la realidad que nos rodea.

Todo este tipo de planteamientos y convicciones que Penrose se sostienen sobre una idea filosófica que nuestro autor defiende sin ambages, este es, el platonismo, cuya importancia en su esquema veremos a continuación.

5. Penrose y el platonismo

5.1. ¿Qué dice Penrose del platonismo?

Resulta llamativo que el platonismo sea fundamental en el conglomerado del pensamiento de Penrose y que, sin embargo, no constituya una parte muy amplia en su trabajo. Y no es que nuestro autor vacile a la hora de declararse seguidor de esta corriente (al menos en el plano matemático). De hecho, se define a sí mismo como platonista en numerosas ocasiones. Veamos un par de ejemplos en los que habla un poco más extensamente del platonismo al que se adscribe:

> Pero quizá la cuestión no sea tan sencilla como esto. Como ya he dicho, hay cosas en las matemáticas, tales como los ejemplos que acabo de citar, para las que el término «descubrimiento» es mucho más apropiado que «invención». Existen los casos en que sale de la estructura mucho más de lo que se introdujo al principio. Podríamos adoptar el punto de vista de que en tales casos los matemáticos han tropezado con «obras de Dios». Sin embargo, existen otros casos en los que la

> estructura matemática no tiene esa compulsiva unicidad; por ejemplo, cuando en medio de la demostración de algún resultado el matemático encuentra necesario introducir alguna construcción artificial, y de ningún modo única, para conseguir algún fin muy específico. En tales casos, no es probable que se obtenga nada más de la construcción que lo que se puso al principio, y la palabra «invención» parece más apropiada que «descubrimiento». Estas son «obras del hombre». Desde este punto de vista, los auténticos descubrimientos matemáticos serían considerados, de forma general, como consecuciones o aspiraciones más altas que lo que serían las «meras» invenciones (Penrose, 1991: 134).

> […] Uno puede muy bien adoptar el punto de vista de que el *mundo platónico* contiene otras ideas, tales como la *Bondad* y la *Belleza*, pero aquí solo me interesaré por los conceptos platónicos de las matemáticas. Para algunas personas resulta difícil concebir que este mundo tenga una existencia independiente. Preferirán considerar los conceptos matemáticos como meras idealizaciones de nuestro mundo físico y, desde esta perspectiva, el mundo matemático se concebiría como algo que emerge del mundo de los objetos físicos.
> Pero no es así como yo concibo las matemáticas, ni creo que la mayoría de los matemáticos y los físicos matemáticos tengan esa idea del mundo. Lo conciben de un modo bastante diferente, como una estructura gobernada de manera precisa y de acuerdo con leyes matemáticas intemporales. Por eso encuentran más apropiado considerar el mundo físico como algo que emerge del *intemporal* mundo de las matemáticas (Penrose, et. al, 2008: 14-15).

Realmente podemos encontrar pocos textos semejantes a los citados en la obra de Penrose en los que nuestro autor dé cuenta de su platonismo particular. Como hemos podido ver en las citas, el platonismo al que se refiere Penrose no dista en exceso del más elemental, este es, el promulgado por Platón. ¿Es correcto, por tanto, definir [o incluir] a Penrose como platonista en el sentido más amplio del término, entendido esto como fiel seguidor de las ideas del filósofo griego? Aunque en un principio pueda parecer que sí, lo cierto es que el pensamiento de nuestro autor contiene trazas que lo hacen diferenciarse de cualquier platonista al uso.

Como dice en la cita (Penrose et al., 1999), Penrose sólo hace referencia al platonismo de los conceptos matemáticos. Para nuestro autor, los conceptos matemáticos son eternos e inamovibles. Esto se ha traducido siempre (es decir, desde cualquier tipo de platonismo) en la independencia de las ideas con respecto al mundo físico. Penrose también lo entiende de ese modo. No obstante, no es tan tajante a la hora de prescindir del quehacer humano. Precisamente este aspecto comporta un problema en la exposición de Penrose, ya que nuestro autor parece posicionado claramente, pero el matiz que añade lo sitúa en una postura difícilmente definible.

Recapitulemos. Los conceptos matemáticos tienen existencia y verdad propias, desde siempre y para siempre. Pero también reconoce el papel importante que tiene la construcción. Esta construcción no puede ser entendida del mismo modo en el que lo entendían los intuicionistas brouwerianos (ya que, recordemos, estos son partidarios de la dependencia total de las matemáticas con respecto al pensamiento humano). Para muchos, la postura de Penrose –un híbrido entre el platonismo y cierto tipo de constructivismo- no queda del todo explicada (Herce, 2014: 63).

Por mi parte no lo considero de tal manera, al menos en grado sumo. Aunque sí puedo percibir el brete en el que nuestro autor se inmiscuye sin pretenderlo (a mi modo de entender), también es reconocible que Penrose ofrece suficientes ejemplos con los que su postura queda suficientemente

clara. Uno de los más esclarecedores es el que concierne al conjunto de Mandelbrot[121].

La idea principal que Penrose pretender extraer de la exposición del conjunto de Mandelbrot consiste en hacer manifiesto que una descripción matemática simple puede dar lugar a un problema matemático muy complejo. Bajo este planteamiento subyace la idea de que, efectivamente, las matemáticas guardan en sí mismas mucho más de lo que en un principio pueda parecer:

> […] Benoît Mandelbrot, el matemático polaco-estadounidense (protagonista de la teoría fractal) que primero estudió el conjunto no tenía ninguna concepción previa acerca de la fantástica elaboración inherente al mismo, aunque sabía que estaba en la pista de algo muy interesante. ¡De hecho, cuando empezaron a surgir sus primeras imágenes de computadoras, él tuvo la impresión de que las estructuras difusas que estaba viendo eran el resultado de un mal funcionamiento de la computadora! Sólo más tarde llegó a convencerse de que estaban realmente en el propio conjunto. Además, los destalles completos de la compleja estructura del conjunto de Mandelbrot no pueden ser aprehendidos realmente por ninguno de nosotros, ni pueden ser completamente revelados por una computadora. Parecería que esta estructura no es sólo parte de nuestras mentes sino que tiene una realidad autónoma. Cualquiera que sea el entusiasta matemático o computadora que decida examinar el conjunto, encontrará aproximaciones a la *misma* estructura matemática fundamental. No hay ninguna verdadera diferencia que dependa de la computadora que se utilice para hacer los cálculos (siempre que la computadora tenga una precisión suficiente), aparte del hecho de que las diferencias en velocidad y memoria de la computadora, y capacidades de representación gráfica, puedan conducir a diferencias en los detalles finos que saldrán a la luz y en la velocidad con que se produce este detalle. El computador está siendo utilizado esencialmente de la misma forma en que el físico experimental utiliza un aparato experimental para explorar la estructura del mundo físico. El conjunto de Mandelbrot no es una invención de la mente humana; fue un descubrimiento. Al igual que el Monte Everest ¡el conjunto de Mandelbrot está *ahí*! (Penrose, 1991: 132)

En la descripción de Penrose existen dos puntualizaciones que es conveniente tener en cuenta: a) la forma elaborada que tiene el conjunto a

---

[121] Dicho conjunto está planteado en los siguientes términos:
Sea c un número complejo cualquiera. A partir de c, se construye una sucesión por inducción:

z0 = 0 es el término inicial
zn+1 = zn2 + 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 (Herce, 2014: 57). La sucesión queda acotada cuando todos los términos de esta son mayores o iguales a *c*.

Esta idea, obviamente, puede ser sometida a debate. Un ejemplo muy actual que es contrario a la visión de Penrose es el llevado a cabo por José Ferreirós (2015). Mientras que Penrose habla de que cuando los matemáticos ponen encima de la mesa algo nuevo en su campo lo que se produce es un descubrimiento, ya que ello *estaba ahí* en todo momento. Ferreirós, por su parte, defiende que el desarrollo de nuevas teorías normalmente se dan en «las interconexiones sistemáticas con el conocimiento y las prácticas previas»* (2015: 253), no otorgándole a las matemáticas una independencia total, tal y como lo hace Penrose al modo platónico.

* Ferreirós expone esta idea hablando concretamente de la teoría de conjuntos y la aceptación previa que es necesaria en matemáticas para poder hablar de invención o descubrimiento en estas. Según Ferreirós, la aceptación previa hace que se dé invención con descubrimiento, es decir, se inventa una regla a través de la práctica (Ferreirós habla concretamente de la invención del conjunto N) y a partir de ello podemos descubrir sus implicaciones.

pesar de provenir de una regla matemática de una simplicidad concreta y b) no es producto de un diseño humano. Con la primera de las características tenemos algo no muy diferente de otros problemas matemáticos. Con respecto a la segunda, nuestro autor toma parte en el platonismo pero sin dejar clara la particularidad de su postura. Sin embargo, esta no es la descripción completa del conjunto por parte de nuestro autor.

Tenemos que el conjunto de Mandelbrot se desarrolla a partir de una regla sencilla y a partir de ella se extiende alcanzando una «variedad infinita y complicación ilimitada». Esto sucedería aunque el ser humano no hubiese dado con ella nunca. Pero el *quid* del asunto para Penrose no se centra tan sólo en tal independencia, sino también en el hecho de que el ser humano logró *dar* con ello. Nuestro autor destaca la especial relación que los seres humanos tenemos con las matemáticas[122]. Si esta relación no fuese especial, ¿cómo sería posible que los seres humanos tengamos la capacidad de *dar* con tales entes independientes y con sus relaciones, a pesar de que estas puedan llegar a ser tremendamente complejas?

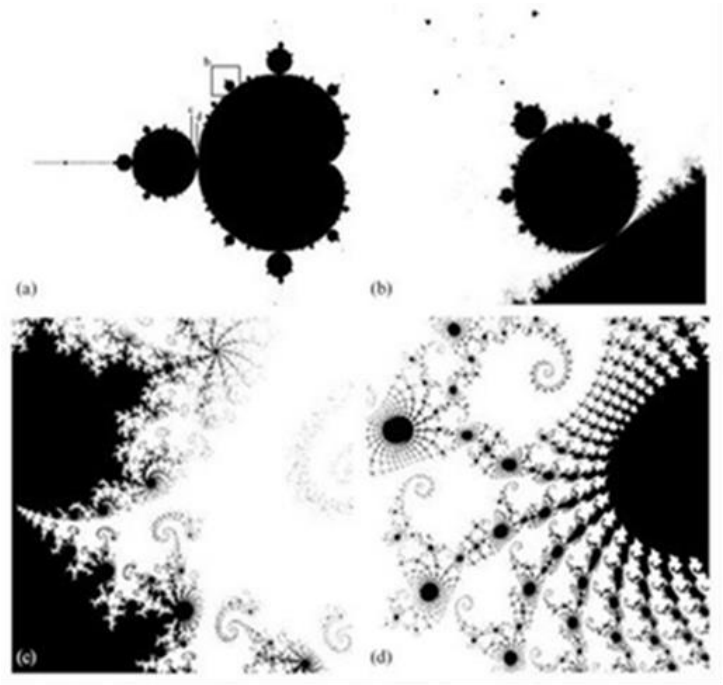

Figura 3. Representación del conjunto de Mandelbrot. Se puede ver cómo a cuanto más detalle se conoce del conjunto más complejo se vuelve.

De todos modos, cabe destacar que Penrose no pone todo el énfasis de la relación en la complejidad de las matemáticas. Para nuestro autor lo importante es que tengamos la capacidad de pensar las propiedades mismas que las matemáticas tienen, sin necesidad de recurrir a los planteamientos más complejos de esta. Penrose defiende que esta relación la que nos diferencia de las máquinas: ¡la totalidad de las matemáticas no puede ser

[122] Sobre cómo entiende Penrose esta relación volveremos más adelante (Cap. 4), cuando veamos la teoría de los tres mundos que nuestro autor desarrolla.

traducida en términos de computabilidad, porque es obvio que hay algo más allá a lo que los procesos computacionales no tienen acceso!

Otro ejemplo del modo especial que tienen los humanos de relacionarse con las matemáticas es el denominado el *problema de la teselación*. Al igual que sucede con el anterior ejemplo, la exposición penroseana de este problema es para argumentar en contra de la IA *fuerte*. Sin embargo, el ataque de este ejemplo es más directo que el anterior, en el sentido de que con el conjunto de Mandelbrot podemos prescindir de la perspectiva de las máquinas. Con el problema de la teselación, por el contrario, no sucede lo mismo. Veamos por qué.

El *problema de la teselación* es un problema geométrico el cual plantea que en un plano euclidiano compuesto de formas poligonales, estas puedan cubrir dicho plano en su totalidad de manera solapada (esta forma de cubrir el plano es lo que se conoce como *teselar*). La cuestión tiene respuesta, pero el resultado se obtiene a través del manejo de números «reales». Que este tipo de números sea definitivo para la resolución del problema resulta dramático para las máquinas, ya que estas no pueden operar a través de ellos. Es por ello por lo que precisamente Penrose trae a colación este problema en concreto. ¡Estamos ante una limitación de las máquinas que no afecta a los seres humanos!:

> Como hecho curioso, la insolubilidad computacional del problema de la teselación depende de la existencia de ciertos conjuntos de *polinomios* llamados conjuntos aperiódicos –que teselarán el plano *sólo no periódicamente* (i.e. de una forma tal que la estructura completa nunca se repite por mucho que se extienda) (Penrose, 2012: 45).

Como nota histórica cabe decir que la insolubilidad en términos computacionales del *problema de la teselación* ya fue expuesta previamente al trabajo de Penrose. La investigación de Robert Berger en 1966, como extensión de los argumentos de Hao Wang en 1961, dio cuenta de la solución insoluble. Pero esto sólo debe contar como nota histórica, ya que Penrose no pretende reclamar nada acerca de este problema. Si lo destaca es porque considera que ha tenido una importancia especial en su vida intelectual. Prueba de cómo ha influido en su pensamiento se traduce en dos de sus aportaciones en el plano de las matemáticas: el *triángulo sin fin* y la *escalera sin fin* (esta última realizada junto a su padre Lionel). Aparte de las aportaciones sobre este asunto de Berger y Wang, Penrose también destaca la influencia de la obra del pintor holandés M. C. Escher[123], con quien nuestro autor llegó a tener una relación personal[124].

---

[123] Sobre la influencia de Escher, Manjit Kumar en su «Cycles of Time: An Extraordinary New View of the Universe by Roger Penrose – review», relata:
Como estudiante en 1954, Penrose estaba asistiendo a una conferencia en Ámsterdam, cuando por casualidad se cruzó con una exposición de la obra de Escher. En aquel momento estaba tratando de conjeturar figuras imposibles propias y descubrió el «tri-bar» -un triángulo que parece un objeto tridimensional real y sólido, pero que no lo es. Junto a su padre, físico y matemático, Penrose siguió diseñando una escalera que simultáneamente gira hacia arriba y hacia abajo. A esto le siguió un artículo del cual envió una copia a Escher. Completando un flujo cíclico de creatividad, el maestro holandés de las ilusiones geométricas se inspiró para producir sus dos obras maestras (Kumar cit. por Herce, 2014: 28-29).

[124] Aunque esta no fue muy profunda en un principio sí que es destacable, ya que, como cuenta Penrose en una entrevista para el Canal 22 de México, el artista y el padre de nuestro

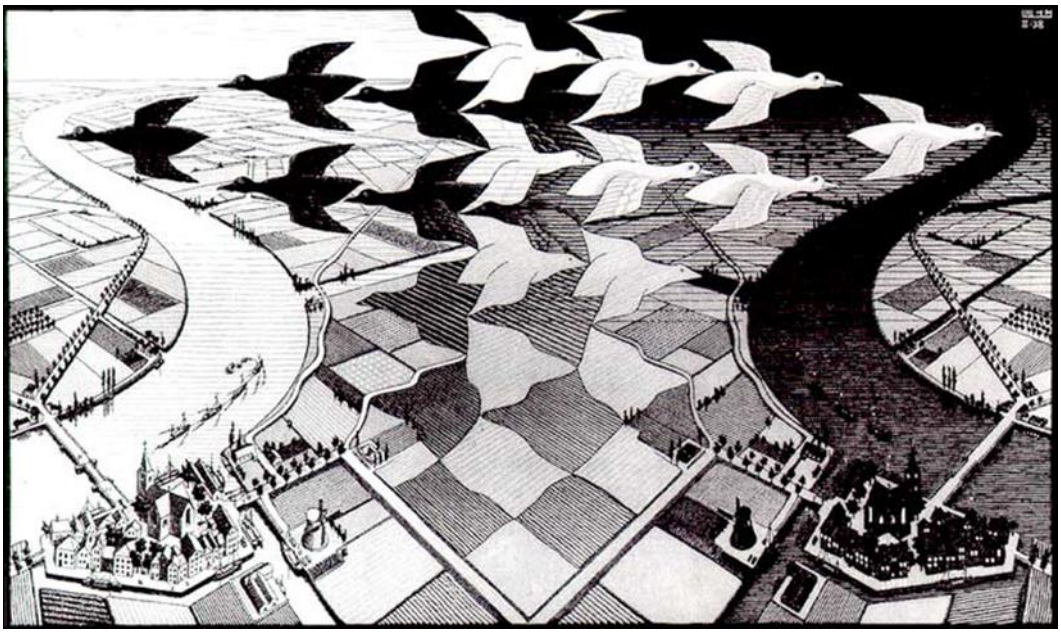

Figura 4. Ilustración de la obra de Escher «Día y noche» (1938). El problema de la teselación en concreto ocupó un lugar importante en la obra del artista holandés, la cual se caracteriza en general por su impecable utilización de la geometría.

autor iniciaron una correspondencia, la cual sirvió para que la retroalimentación fuese más fluida. Más tarde, Penrose también comenzaría a cartearse con Escher, enviándole diseños que respondían a principios geométricos avanzados. Nuestro autor de un diseño concreto: un nonaedro de tejas. El principio en el que se sostenía dicho diseño fue utilizado por el artista holandés en su última litografía (se tituló *Fantasmas*).

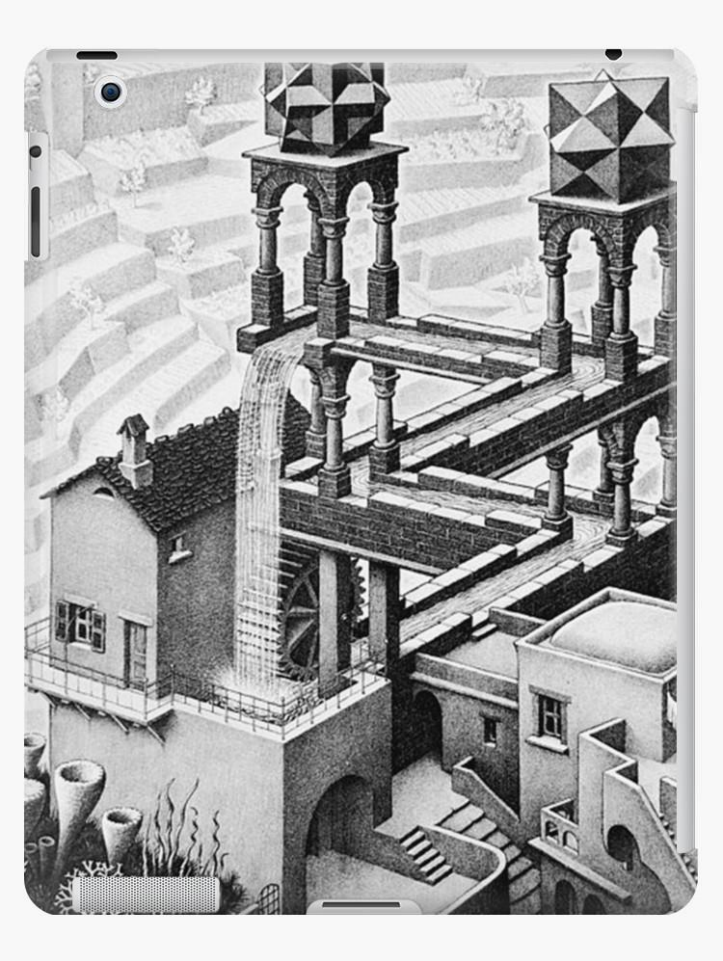

Figura 5. Ilustración de la obra de Escher «Cascada» (1961). En ella se ve cómo el artista utiliza el «tri-bar» de Penrose (en el último recorrido donde el agua, antes de caer, forma dicho triángulo).

Aunque Penrose se empeñe en utilizar ejemplos científicos para justificar su platonismo, podemos percibir que este tema desemboca en un discurso que responde en último término a convicciones internas. Y es aquí donde, considero, Penrose es claro. Si bien su defensa se vuelve un poco dispersa, no es porque su (por seguir con la misma expresión) convicción interna sea poco clara, sino más bien porque llega al punto en el que no puede dar más cuenta de ella (científicamente hablando[125]).

En mi opinión, pienso que Penrose no entra en definir en exceso el tipo de platonismo que defiende precisamente para no entrar en un debate que concierne al quehacer de los filósofos. Es decir, que prefiere seguir hablando de física, terreno en el que se siente seguro, en lugar de arriesgarse a entrar en el campo de la metafísica. Sin embargo, acabar adentrándose en este ámbito, como hemos visto, parece inevitable. Esto es algo que ha ocurrido siempre y, presumo, seguirá ocurriendo. El caso de [nuevamente] Gödel,

[125] Por otro lado esto no deja de ser frustrante para el mismo Penrose, ya que su punto de vista C defiende que la ciencia debe garantizar las respuestas a los misterios del universo. No obstante, también defiende que la ciencia actual no puede proporcionar tales resultados, de ahí a que considere necesario una reforma en la misma.

como personalidad importante dentro del pensamiento de Penrose, es paradigmático en el aspecto concreto del platonismo.

5.2. Penrose y Gödel II: el platonismo como unión

Ya vimos anteriormente la influencia de Kurt Gödel en los planteamientos de Penrose a través del teorema de incompletitud que el lógico austríaco propuso en contra de logicistas y formalistas[126]. El hecho de que surgiera el teorema gödeliano en el momento en el que lo hizo no suponía una tarea fácil, ya que tales corrientes imperaban en la escena de las matemáticas y con ello su rasgo más característico (en el plano filosófico): el anti-platonismo.

Gödel era totalmente contrario a este movimiento y su platonismo supuso un punto de vista filosófico llamativamente sutil y, para muchos (entre ellos Penrose), acertado.

La sutileza del platonismo gödeliano reside en varios aspectos. El primero de ellos es la relación de las matemáticas con el mundo físico. A Gödel le cuesta entender que se tenga más confianza en la intuición que proporcionan los sentidos que de la facilitada por las matemáticas. Aunque reconoce que existen cierto tipo de matemáticas que son difíciles de *encajar* en el mundo físico, ello no las hace menos fiables que la experiencia sensorial:

> Los objetos de la teoría de conjuntos transfinitos claramente no pertenecen al mundo físico e incluso su conexión indirecta con la experiencia física es muy flexible (debido principalmente al hecho de que los conceptos teóricos de conjuntos juegan un papel menor en las teorías físicas de hoy). Pero, a pesar de su lejanía de la experiencia sensorial, tenemos algo así como una percepción también de los objetos de la teoría de conjuntos, como se ve por el hecho de que los axiomas se imponen sobre nosotros como verdaderos. No veo ninguna razón por la que deberíamos tener menos confianza en este tipo de percepción, es decir, en la intuición matemática, que en la percepción sensorial, lo que nos induce a construir teorías físicas y esperar que las futuras percepciones sensoriales estén de acuerdo con ellas y, además, creer que una pregunta no decidible ahora tiene significado y puede decidirse en el futuro (Gödel, cit. por Berto, 2009: 149-150).

Hay, por tanto, una «evidencia» (o «realidad») propia de las matemáticas en sí mismas, pero tal «evidencia» no es tan evidente en tanto a lo que los humanos conocemos de ellas. Obviamente existen rasgos de las matemáticas que se escapan a nuestras capacidades. A pesar de dicha incapacidad que tenemos de *alcanzar* la totalidad de la «evidencia» matemática, ello no impide que podamos *acercarnos* a ella, con el paso del tiempo, incluso, cada vez más. Ya vimos más arriba cómo la utilización de ciertos números, como los «reales» o los «naturales», lograron ampliar el abanico de respuestas con «evidencia» matemática. De hecho, Gödel considera que, en concreto, los

[126] Además, el mismo Penrose siente necesario acudir a Gödel para dar cuenta de su platonismo, tal y como lo recuerda en NME: […] Imagino que siempre que la mente percibe una idea matemática toma contacto con el mundo platónico de los conceptos matemáticos […] Cuando «vemos» una verdad matemática, nuestra conciencia irrumpe en este mundo de ideas y toma contacto directo («accesible por vía del intelecto»). He descrito esta «visión» en relación con el teorema de Gödel pero es la esencia de la comprensión matemática (Penrose, 1991: 531).

números «naturales» y sus leyes nos son de gran utilidad para el *acercamiento* a la «evidencia» matemática (Ferreirós, 1999: 460). Esta idea es compartida por el platonismo penroseano. Sin embargo, existe un punto en el planteamiento de Gödel que si bien no es imposible que Penrose lo acepte, no menos cierto es que en el discurso de nuestro autor dicho punto no aparece claramente reflejado. Tal aspecto dicta que aquello que surge de los números «naturales» y sus leyes (hipótesis, axiomas, etc.) pertenece al plano de la convención[127]. Sin embargo, para el lógico austríaco esta convención no es plena, al menos en el sentido estricto del término (Ferreirós, 1999: 460). Así las cosas, lo cierto es que con esta idea Gödel confiere dos características a la relación del ser humano con las matemáticas:

i) Ese *algo especial* que permite entrar en contacto con el (aunque Gödel no hable en estos términos) mundo matemático y,

ii) conceder cierta libertad al quehacer humano (¡siempre dentro de los parámetros de la «evidencia» matemática!).

En principio, Penrose no aceptaría la segunda de las características. Nuestro autor entiende que aquello que descubrimos de las matemáticas *es* tal y como *son* ellas mismas, no dependiendo de ningún tipo de convención. No obstante, la diferencia, considero, es superficial, ya que, como hemos visto, Gödel no aboga por una convención plena. Esta idea está relacionada con otro de los aspectos a destacar del platonismo gödeliano: la relación que establece entre las matemáticas y las ciencias naturales.

En este aspecto en particular Gödel se sitúa muy en la dirección de Russell (Ferreirós, 1999: 460). Para hacer manifiesta la semejanza entre las matemáticas y las ciencias naturales el lógico austríaco defiende que los objetos físicos no son «datos». En lugar de ello entiende que tales objetos son

---

[127] Este concepto puede ser muy controvertido. Gödel lo entiende de un modo similar al concepto general que usa Poincaré (teniendo en cuenta la diferencia de contexto), es decir, que aunque las teorías y sus leyes se estimen a través del acuerdo, dicho acuerdo no puede ser arbitrario. María de Paz Amerigo expone de un modo claro este concepto en su obra *Mecánica y Epistemología en Henri Poincaré*: […] Poincaré reconoce los límites del conocimiento, en primer lugar, porque la ciencia es un producto de confección humana, que exige la toma de decisiones por parte del científico, lo que pone de manifiesto que no puede ser un mero reflejo especular de la naturaleza. Obviamente, no hay ciencia sin experiencia, y dadas las variaciones a las que esta se encuentra sometida, aquella estará también sujeta a cambio, en la medida en que es su base. Así, se opone a que la ciencia pueda ser un sistema en cuanto cuerpo de conocimientos establecido de manera definitiva en el que las proposiciones se enlacen unas con otras, pero sin que ello signifique que sea un mero conjunto desorganizado

[…] la convención como una categoría epistemológica diferente de lo a priori y de lo empírico tiene el papel tanto de rebajar las pretensiones de verdad de la ciencia empírica, como de poner de manifiesto que la mayor parte de sus elementos no pueden ser considerados verdaderos a partir de nuestras capacidades intelectuales.

[…] Poincaré equipara lo real con lo objetivo. Al redefinir la objetividad como intersubjetividad, nuestro autor demarca el único campo posible del conocimiento. Ahora bien, el garante de la objetividad del mismo tendrá que ser la posibilidad de comunicarlo (de Paz Amerigo, 2014: 127-128).

En mi opinión, tanto Gödel como Penrose, conceden al ser humano un grado importante de construcción, pero en el que esta construcción no constituye la fuente de conocimiento sino más bien el medio (¡la fuente siempre lo es la verdad matemática!). Por su parte, Poincaré concede un grado más a la convención, al acuerdo. Pero, y de nuevo esto es una opinión personal, esto se debe a que el intelectual francés se detuvo más en este apartado que Gödel o Penrose, y no a una diferencia sustancial entre ellos.

«entidades teóricas que postulamos para dar cuenta de los fenómenos que percibimos» (Ferreirós, 1999: 460):

> Debería observarse que la intuición matemática no tiene que ser concebida como una facultad que proporcione un conocimiento *inmediato* de los objetos que le conciernen. Parece más bien que, como en el caso de la experiencia física, *formamos* también nuestros conceptos de estos objetos a partir de algo más que *es* inmediatamente dado. Sólo que este algo más *no* es aquí, o no principalmente, la sensación. Que además de las sensaciones hay algo real e inmediatamente dado se sigue (independientemente de las matemáticas) del hecho de que incluso nuestros conceptos referentes a los objetos físicos contienen constituyentes cualitativamente diferentes de las sensaciones o meras combinaciones de sensaciones, por ejemplo, el concepto mismo del objeto, mientras que, por otro lado, mediante nuestro pensamiento no podemos crear ningún elemento cualitativamente nuevo, sino sólo reproducir y combinar los que están ya dados. Lo «dado» que subyace a las matemáticas está, evidentemente, muy relacionado con los elementos abstractos contenidos en nuestros conceptos empíricos. De esto no se sigue, sin embargo, que los datos de este segundo tipo sean algo puramente subjetivo, porque no pueden asociarse con acciones de ciertas cosas exteriores a nuestros órganos sensibles, como Kant afirmaba. Pueden representar más bien un aspecto de realidad objetiva, pero, en oposición a las sensaciones, su presencia en nosotros puede deberse a otro tipo de relación entre la realidad y nosotros mismos (Gödel, 2006: 427- 428).

Sin duda, la postura de Gödel es aguda y contiene un atrevimiento al que Penrose no aspira (por los problemas filosóficos que ello podría acarrearle). El lógico austríaco no se amedrenta a la hora de tomar partido en un asunto tan peliagudo como lo es nuestra relación con la realidad. Pero, al fin y al cabo, ¿por qué no expresar las convicciones internas de cada uno(a) en lo concerniente a cuestiones filosóficas fundamentales?

Volviendo al asunto en sí mismo, cabe destacar que Gödel defiende que la semejanza es necesariamente bidireccional, en el sentido de que no todo en las matemáticas es intuible al igual que no todo lo físico es perceptible:

> Pero esta analogía de la intuición con la percepción, del realismo matemático con el realismo del sentido común, no es el final de la elaboración de Gödel de la analogía del realista matemático entre las matemáticas y las ciencias naturales. Así como hay hechos sobre objetos físicos que no son perceptibles, hay hechos sobre objetos matemáticos que no son intuibles (*intuitables*). En ambos casos, nuestra creencia en tales hechos «no observables» está justificada por su papel en nuestra teoría, por su poder explicativo, su éxito predictivo, sus fructíferas interconexiones con otras teorías bien confirmadas, etc. (Maddy, 2003: 32).

Por su parte, Penrose no llega tan lejos como Gödel en este apartado. Aunque también sea defensor de la relación entre lo físico y lo matemático, Penrose acaba declarando que lo máximo que puede decirse de ello es que es un misterio[128] (Penrose, 2012: 434-443).

No obstante, el camino a la realidad de ambos tiende a encontrarse, sobre todo en lo concerniente a las epistemologías de sus respectivos platonismos. Penelope Maddy define la de Gödel del siguiente modo:

> […] la epistemología platónica de Gödel tiene dos niveles: los conceptos y axiomas más simples se justifican intrínsecamente por su intuición; las hipótesis más teóricas se justifican extrínsecamente, por sus consecuencias. […] Las hipótesis justificadas

[128] Este tema lo trata de un modo más extenso y de forma explícita en su teoría de los tres mundos. Por ello lo veremos más adelante en el capítulo 4 y no en el presente.

> extrínsecamente no son seguras y, dado que Gödel permite la justificación por fecundidad en física y en matemáticas, tampoco son a priori. […] Gödel le da todo el crédito a las formas puramente matemáticas de justificación (autoevaluación intuitiva, pruebas y justificaciones extrínsecas dentro de las matemáticas) y la facultad de intuición hace justicia a la evidencia de las matemáticas elementales (Maddy, 2003: 33).

Este [en mi opinión, acertado] análisis de Maddy sobre Gödel sitúa, por ende [y de nuevo en mi opinión], a Penrose en el mismo orden epistemológico en cuanto al platonismo. Nuestro autor comparte la idea de que el conocimiento no puede sustentarse en ideas teóricas, sino en certezas. Es por ello por lo que cree, al igual que Gödel, que las matemáticas constituyen una luz fundamental para enfrentar a los misterios del universo. Por otra parte, esto no impide que ambos pensadores reconozcan la imposibilidad para los seres humanos de poder dar cuenta de todos los misterios. No obstante, tanto Gödel como Penrose son defensores aguerridos de la idea de que las matemáticas tienen un poder explicativo tremendamente amplio y fiel a la realidad.

5.3. ¿Es sostenible el platonismo de Penrose?

Una cuestión importante desde el plano filosófico es determinar hasta qué grado el platonismo de Penrose es legítimo. Resulta cuanto menos llamativo que las ideas de Penrose hayan sido objeto de numerosas críticas en todas sus facetas pero, en cambio, su platonismo haya pasado casi desapercibido.

En mi opinión, la crítica directa más elaborada y trascendente es la que lleva a cabo el filósofo Mark Steiner. La crítica de Steiner se centra, sobre todo, en la idea de que Penrose no entiende adecuadamente el platonismo. ¿En qué aspecto Penrose malinterpreta el platonismo? Steiner defiende que Penrose, al poner el foco en el pensamiento matemático, hace una distinción entre los conceptos que Platón no realiza en el planteamiento de su filosofía:

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

La «idea del Bien» es la garante de la Verdad y no las matemáticas, por lo que concebir lo contrario sería un error. De hecho, Steiner va más lejos y piensa que la exposición penroseana se ajusta más a la filosofía cartesiana en lugar de la platónica (Steiner, 2000: 134), ya que Descartes sí que da prioridad a las matemáticas (concretamente a sus conceptos[129]).

[129] Esta no es una consideración sin fundamento, ya que Descartes llegó a decir literalmente: La matemática nos acostumbra a reconocer la verdad, porque en ella se hallan razonamientos correctos que no encontrarás en ninguna otra parte. Por lo tanto, quien haya acostumbrado

Si bien por un lado considero que Steiner tiene razón, por el otro pienso que una parte del análisis es erróneo. Es fácil conceder que Penrose pone el énfasis de su defensa en el pensamiento matemático. Pero a pesar del estatus especial que le otorga, es complicado observar que nuestro autor realice una distinción entre los conceptos matemáticos y los demás. El mismo Penrose sabe que de hacerlo sólo conseguiría entrar en un terreno farragoso. Uno de los motivos de peso por el que se aferra al ámbito matemático es porque este es su campo de estudio. Es cierto que defiende cierta superioridad de las matemáticas, pero, sin embargo, en ningún momento parece hablar en términos absolutos. Las matemáticas son aquellas en las que Penrose se siente cómodo y las que le permiten explicar de un modo más adecuado la conexión con el platonismo. Podemos ver expresada dicha idea casi al final de SM:

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

Si Penrose considera que cuestiones como la ética, la moralidad o la estética no dejan de ser igual de «reales» que las matemáticas, ¿cómo puede defender que existe una superioridad en términos absolutos como lo defiende, en cierto modo, Descartes? Realmente, defiendo, no lo hace, sino que comprende cuál es el terreno que domina (y en el que, por supuesto, cree) y se limita a dar explicaciones desde este:

> […] Sólo respecto a esta cualidad mental he sido capaz de hacer la necesaria afirmación fuerte: que es esencialmente *imposible* que una cualidad semejante pueda haber surgido como una característica de la mera actividad computacional, ni puede ser siquiera simulada adecuadamente por la computación –y recalcaré que no hay aquí ninguna sugerencia de que exista algo especial en la comprensión *matemática* frente a cualquier otro tipo de comprensión. La conclusión es que cualquiera que sea la actividad cerebral responsable de la consciencia (al menos en esta manifestación particular) debe depender de una física que está más allá de la simulación computacional (Penrose, 2012: 433).

Sin embargo, Steiner está convencido de que Penrose es más partidario del cartesianismo que del platonismo. Para ello trae a colación un texto de Hertz que dice así:

> Uno(a) no puede escapar del sentimiento de que estas fórmulas matemáticas tienen una existencia independiente de nosotros(as), que son más sabias que nosotros, incluso que sus descubridores, que tienen más en ellas de lo que originalmente se pusieron en ellas (Hertz, cit. por Steiner, 2000: 135).

A continuación, Steiner comenta que tanto Descartes como Hertz sostienen que «los conceptos matemáticos contienen una «información latente» de manera que van más allá de la mera deducción lógica» (Steiner, 2000: 135).

---

su ingenio a los razonamientos matemáticos, también será apto para investigar otras verdades, porque la razón es en todas partes una y la misma (Descartes, 2011: 457).

Si bien la exposición es correcta, la crítica de fondo vuelve a ser incorrecta. Aunque Penrose esté de acuerdo con lo defendido por Descartes y Hertz (que lo está), ¿en qué medida ello lo sitúa fuera del platonismo? La idea subyacente en la frase citada es la independencia de la existencia de los números con respecto al mundo físico y esto, como hemos visto anteriormente, es algo que se encuentra dentro de las características del platonismo. Si hay un aspecto en el que Penrose no vacila al admitir su adherencia al platonismo es este. Por tanto, utilizar dicho rasgo como definitorio del error de Penrose es, cuanto menos, inapropiado.

Dejando a un lado la crítica de Steiner, por otro tenemos la crítica que Feferman hace del platonismo penroseano, la cual también está incluida en el artículo que pudimos ver más arriba.

Siguiendo el mismo tono de su crítica anterior, Feferman reconoce como un hecho que la gran mayoría de los matemáticos(as) puedan sustentar su modo de concebir las matemáticas en el platonismo. Sin embargo, también entiende que esto no puede ser la única garantía de su conocimiento:

> Si bien los matemáticos pueden concebir de lo que están hablando en términos teóricos platónicos, estos resultados muestran que tal concepción no es necesaria para garantizar la confianza en el cuerpo de la práctica matemática (Feferman, 1995: 10).

A pesar de que en un principio dé parte de la razón a Penrose, más tarde muestra su disconformidad, o más bien cierto escepticismo, con respecto a entender la teoría de conjuntos a través del platonismo. El mismo Gödel, gran defensor de esta postura, también en un momento recularía en sus especulaciones:

> […] Pasando a las cuestiones filosóficas planteadas por el platonismo en la teoría de conjuntos, Penrose tiene razón al identificar a Gödel como uno de los principales defensores de esta posición. Sin embargo, creo que es justo decir que tiene pocos seguidores entre los filósofos de las matemáticas. Es cierto que toda filosofía general de las matemáticas tiene sus dificultades, pero Penrose hace que parezca que la posición platónica es una cuestión de consenso común. Este no es el caso para aquellos que han dado estas preguntas más que atención simbólica. Si bien uno puede estar de acuerdo en que las cuestiones de verdad en los números naturales tienen un carácter determinado, ya la suposición de una supuesta totalidad definida de conjuntos arbitrarios de números naturales es muy problemática. De hecho, el propio Gödel, al menos durante un período en la década de 1930, encontró esto profundamente preocupante. En una conferencia previamente inédita […], dijo: «El resultado de la discusión anterior es que nuestros axiomas [para la teoría de conjuntos], si se interpretan como declaraciones significativas, necesariamente presuponen un tipo de platonismo, que no puede satisfacer ninguna mente crítica y que ni siquiera produce la convicción de que son consistentes». Gödel continuó tomando en serio los enfoques teóricos proactivos de la coherencia a lo largo de su vida (Feferman, 1995: 10).

En opinión de Feferman, esta posición es tan poco sostenible que ello sólo puede conducir a Penrose hacia una ruta en solitario:

> […] Por cierto, en la pág. 116 de SOTM[130], Penrose dice que Paul Cohen, en la última sección de su libro de 1966 sobre la independencia de AC y CH de la teoría de conjuntos de ZF «se revela como Gödel [y Penrose] un verdadero platónico para

[130] Siglas de *Shadows of the mind*.

> quien importa la verdad matemática son absolutos y no arbitrarios». Si bien eso es una inferencia razonable de lo que dijo Cohen allí, poco después de eso, en una conferencia de 1967, declaró: «A estas alturas puede haber sido obvio que he elegido la posición formalista [en oposición al realista platónico] para la teoría de conjuntos» (Cohen 1971, p. 13). Que yo sepa, esa sigue siendo su opinión (Feferman, 1995: 10).

Si bien la terea principal de alguien que presenta sus ideas es convencer con ellas y, por tanto, tener más adeptos que detractores, no por ello deja de existir la posibilidad de que ocurra lo contrario. Y no es que Penrose haya emprendido ese camino solitario en el que lo ve Feferman, pero sí que es evidente que sus planteamientos son proclives a la discusión. De todos modos, no creo que Penrose tenga problemas por ver la situación de discusión continua en la que se encuentra su postura. Me atrevería a decir, incluso, que ocurre lo contrario, ya que la actitud de nuestro autor no se caracteriza por querer dar respuestas definitivas[131].

Que las críticas de Steiner o la de Feferman me parezcan que, en principio, no ponen en demasiados apuros al platonismo de Penrose no significa que la postura de nuestro autor me tenga totalmente convencido. Por otra parte, tampoco me parece del todo inadecuada. Las críticas que van más en la dirección de poner en entredicho la postura de Penrose por considerarla demasiado metafísica me parecen que tienen un tono que podría ser más acorde con el mío propio. Pero no en el mismo sentido. Creo que si se le puede achacar que sus ideas con respecto al platonismo se vuelvan metafísicas (normalmente se utiliza de forma peyorativa) no es porque se meta en jardines innecesarios, sino, más bien, por no acabar de meterse del todo en ellos. Y no es que considere que Penrose sea poco atrevido en sus planteamientos (ni mucho menos), pero en lo referente a este asunto en particular sí que echo de menos la audacia de la que hace gala nuestro autor en la mayor parte de su obra.

6. ¿Es el pensamiento matemático definitorio de la consciencia humana?

Para Penrose el papel del pensamiento matemático es, cuanto menos, «especial». Hemos podido ver (§4.2, §4.3) que partes de las matemáticas que en principio pueden parecer completamente alejadas del mundo físico en el fondo están relacionadas de modo tal que ofrecen una explicación muy fiel de este.

[131] Esto es algo que intenta dejar muy claro Penrose al inicio de NME: […] Pedir respuestas definitivas a preguntas tan fundamentales estaría fuera de lugar. Yo no puedo proporcionar tales respuestas: nadie puede, aunque algunos puedan tratar de impresionarnos con sus conjeturas. Mis propias conjeturas jugarán un papel importante en lo que sigue, pero trataré de distinguir claramente tales especulaciones de los hechos científicos brutos, y trataré también de dejar claras las razones que subyacen a mis especulaciones. No obstante, mi principal propósito aquí no es el de conjeturar respuestas sino el de plantear algunos temas aparentemente nuevos concernientes a la estructura de la ley física, la naturaleza de la matemáticas y del pensamiento consciente, y de presentar un punto de vista que no he visto expresado hasta ahora (Penrose, 1991: 24).

Ahora bien, ¿hasta qué punto es definitivo el pensamiento matemático a la hora de definir la consciencia humana? Si bien Penrose parece convencido de que el pensamiento matemático es indudablemente esencial para comprender la consciencia de nuestra especie, también se muestra cauto. Aún estamos lejos (o al menos no estamos muy cerca) de poder ofrecer una respuesta final acerca de la naturaleza de la consciencia humana, así que intentar dar dicha respuesta contundente se antoja una tarea tanto infructuosa como frustrante. Esto, sin embargo, no impide que nuestro autor sea partidario de buscar «pistas» que nos acerquen cada vez más a la realidad, y el pensamiento matemático puede situarnos en un muy buen lugar para dar con esas pistas. Este aspecto, como hemos visto más arriba, ha sido malinterpretado en varias ocasiones, entendiéndose que Penrose deposita exclusivamente en las matemáticas la capacidad de garantizar la explicación sobre la consciencia humana. Penrose no lo entiende así. Si lo hiciese de tal modo se limitaría a permanecer en el plano de las explicaciones matemáticas, pero en lugar de ello extiende los campos de investigación de la consciencia hacia la física y la neurociencia. El motivo por el que lo hace tampoco está relacionado con restringir la tarea a estos saberes, sino a que estos son campos que le resultan familiares (de hecho es físico y ha realizado investigaciones sobre neurociencia junto a su hermana). En ningún caso se detecta que Penrose niegue que otros saberes puedan contribuir a la labor que él se propone, sino más bien todo lo contrario.

Ante la pregunta de si el pensamiento matemático es definitorio de la consciencia humana cabe responder por Penrose que sí, aunque matizando también que no es definitivo. Es necesario hacer participar otras ciencias para seguir avanzando, por usar las palabras de nuestro autor, en nuestro camino a la realidad.

En los dos siguientes capítulos podremos ver el modo en el que Penrose encuentra necesarias tanto la física como la neurociencia[132].

[132] Siendo más extensa y específica la parte de la física con respecto a la neurociencia. Ello se debe, como anteriormente, a que Penrose tiene formación física y no neurocientífica.

Capítulo 3

Argumentos físicos

1. La fuerza de la física clásica y sus límites

1.1. Necesidad de una explicación física

Antes de entrar a considerar los argumentos físicos que Penrose pone encima de la mesa, entendamos por qué él considera que estos deben ser tenidos en cuenta en primer lugar[133].

Recordemos que el punto de vista en el que Penrose se sitúa a sí mismo (el C *fuerte*) defiende que la respuesta a los enigmas del debate de la computabilidad o no-computabilidad de la mente y la consciencia humana debe proceder de la ciencia. Este es un rasgo, como vimos, que comparte con el punto de vista A. Ello no impide, sin embargo, que en este apartado también exista una discrepancia fundamental entre ambos puntos de vista. Mientras que para A las matemáticas son las encargadas de resolver los misterios del debate, para C estas no pueden tener la única y última palabra. Decir esto después del capítulo que acabamos de dejar atrás puede resultar confuso, pero en realidad no lo es en absoluto. Por supuesto que la postura de Penrose aboga por dar un lugar importante en el debate a las matemáticas. No obstante, nuestro autor piensa que la sola descripción matemática no se correspondería con una descripción *completa* de la realidad.

[133] Para la exposición de la física clásica que veremos he seguido el guion de NME, ya que considero que es la obra más adecuada para captar aquello que Penrose quiere decir acerca de dicha física (en SM ignora su explicación, mientras que en EcalR resulta muy extensa para lo que pretendo tratar en este capítulo).

Recordemos que para el punto de vista A la explicación del fenómeno de la consciencia depende, en última instancia, de un procedimiento algorítmico (o de varios), por lo que atener la investigación de manera exclusiva a las matemáticas es plenamente lícito.

Por su parte, Penrose cree, como hemos visto en varias ocasiones, que es necesario encontrar, precisamente, un procedimiento no-algorítmico. Esto no supone, ni mucho menos, arrebatarles competencias a las matemáticas. Como vimos más arriba, Penrose se encarga de intentar demostrar precisamente que las matemáticas en sí mismas van más allá de los procedimientos algorítmicos, para hacer manifiesto que si bien todos los procedimientos de este tipo son matemáticos, no todas las matemáticas son algorítmicas.

En lo concerniente a la búsqueda que Penrose anhela, por un lado surge la pregunta de cómo podemos dar con ese procedimiento no-algorítmico tan fundamental para el debate. Y por otro, tratándose de una propuesta, existe la legítima posibilidad de preguntar en qué medida *existe* dicho procedimiento. Para Penrose estas preguntas no resultan tan abismales como en un principio puedan parecer, ya que tiene respuesta para ambas. Es cierto que tales repuestas no son tan contundentes como a Penrose le gustaría que fuesen, pero ello no es óbice para que nuestro autor defienda que estas no obedecen a simples *creencias* internas. Penrose defiende que la Naturaleza *realmente* nos da pistas sobre el susodicho procedimiento no-algorítmico.

Ahora bien, ¿cuál es el medio más adecuado para traducir las pistas que la Naturaleza nos brinda de tal procedimiento no-algorítmico? Para Penrose es imprescindible que el campo elegido tenga una rigurosidad matemática importante y que dicha rigurosidad sirva para describir los acontecimientos que experimentamos. Estas condiciones nos llevan de manera necesaria al terreno de la física.

La física, por tanto, es la materia plenamente necesaria para los planteamientos del debate que Penrose pretende abarcar. Pero, ¿de veras la física puede dar la última respuesta? Si así fuese, ¿no estaríamos un poco más cerca de dar con una solución al problema, el cual realmente parece estar estancado desde hace ya no pocos años? Con respecto a la primera pregunta, nuestro autor no comparte, de manera obvia, el escepticismo con respecto al alcance de la física. Pero la incredulidad de la segunda pregunta sí que le convence, aunque no por las mismas razones (ya que la pregunta sigue la actitud del planteamiento de la primera). Cierto, Penrose cree que el debate se encuentra en un momento de estancamiento, pero defiende que ello no es debido a la impotencia de la física, sino a que esta necesita ser replanteada:

> […] Tal vez sea realmente relevante el modo detallado en que estamos constituidos, como lo sean las leyes físicas que realmente gobiernan la substancia de la que estamos compuestos. Quizá necesitamos comprender qué cualidad profunda subyace en la naturaleza misma de la materia y determina el modo en que esta materia debe comportarse. La física no ha alcanzado aún este punto. Todavía quedan muchos misterios que desentrañar y muchas intuiciones directas que obtener (Penrose, 1991: 195).

Cabe decir que Penrose no visualiza una revolución en la ciencia actual. Como bien señala Rubén Herce, quien sigue de un modo flexible la terminología de Kuhn, la ciencia para Penrose se encuentra en un período que se conoce como *normal*. En esta fase la ciencia se mueve dentro de un

paradigma[134] consensuado, el cual sirve de base y es susceptible de cambios que bien pueden ser importantes, aunque en raras ocasiones sustanciales (Herce, 2014: 43). No se trata, por tanto, de romper drásticamente con lo establecido, sino de introducir cambios que permitan explicar aspectos de la Naturaleza que por ahora permanecen ocultos.

De hecho, Penrose reconoce la capacidad explicativa que tiene la ciencia clásica, en concreto la física. Por ello encuentra necesario hacer un repaso de teorías físicas que han contribuido de manera notable al conocimiento de los entresijos de la Naturaleza. En ese repaso, Penrose trata la geometría, la dinámica, la mecánica, el electromagnetismo y la relatividad, señalando las virtudes de estas, aunque también sus limitaciones. Es en esta tarea llevada a cabo por Penrose en la que se concentrará el resto de puntos del apartado, además de otros estudios aparte que estarán conectados directamente con dicha tarea.

1.2. La geometría como descripción física

De manera general, es fácilmente reconocible que la geometría pertenece a las matemáticas. Sin embargo, considerar a esta como teoría física no es un asunto que resulte tan trivial, a pesar de ser aquello que fundamentalmente *es*. Penrose se encarga de que este aspecto quede de manifiesto con su análisis de la geometría.

Antes de pasar a la explicación penroseana (en mi opinión bastante filosófica) de la geometría, es de rigor señalar que en su pensamiento este campo tiene una importancia capital en el plano personal, ya que, como dice en varias ocasiones, su pensamiento es marcadamente geométrico (Penrose, 1991: 377), (Penrose, 1991: 529). Así que si bien nuestro autor considera que la geometría es *objetivamente* fundamental, no menos cierto es que ello también estriba de su interés personal. Este no es un aspecto que pueda reprochársele, aunque sí considero que es necesario tenerlo en cuenta. Sin más, pasemos a ver qué nos dice Penrose acerca de la geometría.

Cuando hablamos de geometría de manera general lo hacemos en referencia a una geometría concreta, esta es, la euclídea. El motivo de esta asociación lo encontramos en su etimología, en el sentido de que esta geometría es aquella que durante más tiempo ha permitido una *medida* «fiel» del *mundo*.

No obstante, casi sobra decir que hoy en día está bastante extendido el conocimiento de las geometrías no-euclidianas. De hecho, Penrose nombra algunas de ellas, en concreto de la geometría *lobachevskiana* (que debe su

[134] Como señalo arriba, Herce utiliza estos términos porque toma de manera general el uso de Kuhn: «ciencia normal» significa investigación basada firmemente en una o más realizaciones científicas pasadas, realizaciones que alguna comunidad científica particular reconoce, durante cierto tiempo, como fundamento para su práctica posterior […] Voy a llamar, de ahora en adelante, a las realizaciones que comparten esas dos características*, «paradigmas», término que se relaciona estrechamente con «ciencia normal» (Kuhn, 2004: 33-34).

* i) Su logro carecía suficientemente de precedentes como para haber podido atraer a un grupo duradero de partidarios, alejándolos de los aspectos de competencia de la actividad científica; ii) Simultáneamente, eran lo bastante incompletas para dejar muchos problemas para ser resueltos por el redelimitado grupo de científicos (Kuhn, 2004: 33).

nombre al matemático ruso Nikolái Lobachevski). De esta geometría nuestro autor destaca su similaridad con la euclídea, aunque también indicando aquello que las diferencia a ambas[135].

Pero la geometría que realmente interesa a Penrose es la euclídea. Su perspectiva sobre este tipo de geometría es la que nos interesa, ya que con ella nuestro autor deja entrever su postura en materia de teoría del conocimiento.

Tal y como se encarga de expresarlo el mismo Penrose, de nuevo el pensamiento de Platón es importante en lo que concierne al conocimiento de la geometría. Recordemos que para el filósofo griego las figuras geométricas[136] se encuentran en el mundo de las Ideas, siendo la geometría, por tanto, un estudio imprescindible para el verdadero conocimiento (Platón, 1992: 355-359). Si la geometría pertenece al mundo de las Ideas, esta nunca debe ser considerada como *exactamente expresada* en el mundo de la experiencia física. Pues bien, Penrose lo entiende de un modo bastante similar, aunque, en mi opinión, con una diferencia fundamental. Nuestro autor (como veremos con más detalle en el Capítulo 4) comparte la existencia de un mundo aparte del físico[137], habiendo una interrelación entre ambos. También suscribe la idea de la *superioridad* del contenido del mundo matemático con respecto al mundo físico. Y como consecuencia de este tipo de concepción, entiende que la geometría (como ya sabemos, Penrose habla en concreto de la euclídea) no debe su existencia a una necesidad lógica, sino a lo que observamos en el mundo físico:

> El hecho de que la geometría euclídea parezca tan precisa para reflejar la estructura del «espacio» de nuestro mundo nos ha engañado (o a nuestros predecesores) haciéndonos pensar que esta geometría es una necesidad lógica […] El que la geometría euclídea se aplique de forma tan precisa –aunque no suficientemente exacta- a la estructura de nuestro espacio físico, lejos de ser una necesidad lógica ¡es un *hecho de observación empírica*! (Penrose, 1991: 206).

A pesar de que Penrose tenga buena parte de razón al reconocer la doctrina platónica en este aspecto, en mi opinión, creo que su análisis es incompleto al no reconocer su deuda aristotélica[138].

En su idea de Euclides como seguidor de Platón y su doctrina, Penrose trae a colación la figura de un miembro de la Academia que fundó el filósofo griego, ya que, considera que tal figura contribuyó de manera notable en la geometría ulterior (¡llegando su influencia incluso hasta nuestros días!). El filósofo al que se refiere Penrose es Eudoxo y la aportación que nuestro autor

---

[135] Para los detalles, véase (Penrose, 1991: 204-206).

[136] Platón no habla en términos de la geometría euclídea, ya que los *Elementos* de Euclides se publicaron unos 50 años después del tratado sobre geometría de Platón (Penrose, 1991: 206).

[137] En realidad, como veremos en el Capítulo 4, habla de dos mundos más.

[138] Este argumento lo sostengo conforme a una aclaración que Ferreirós hace en su artículo sobre platonismos, que dice así: Aristóteles rechazó la distinción de su maestro entre un universo de auténticas realidades y un universo de apariencias. Según él, la unidad o la circularidad se abstraen de los objetos materiales, pero no tienen existencia independiente de esos objetos. Conviene recordarlo aquí, porque algunas posiciones filosóficas que se califican de 'platónicas' están, en realidad, más cerca del aristotelismo (Ferreirós, 1999: 458). Aunque Penrose crea en ese mundo matemático platónico, ya hemos visto que no profundiza a la hora de intentar sostenerlo, sino que opta por defender la idea de construcción que lo acerca más a esta corriente, a pesar de que siga siendo decididamente platónico.

destaca es la introducción de los *números reales*[139]. Generalmente la comunidad matemática considera este apartado más como un asunto de análisis que de geometría (Penrose, 1991: 207), algo con lo que Penrose no está plenamente de acuerdo. Nuestro autor explica que con la utilización de los números reales propuesta por Eudoxo se conseguía dar solución a problemas geométricos que habían surgido en las matemáticas griegas, concretamente con los pitagóricos[140]:

> […] Había sido importante para los griegos el poder formular sus medidas (razones) geométricas en términos de (razones de) enteros para que las magnitudes geométricas pudieran ser estudiadas de acuerdo con las leyes de la aritmética. Básicamente la idea de Eudoxo fue proporcionar un método de describir razones de longitudes (esto es, ¡números reales!) en términos de *enteros*. Él fue capaz de dar criterios, establecidos en términos de operaciones enteras, para decidir cuándo una razón es mayor que otra, o si las dos deben considerarse exactamente iguales (Penrose, 1991: 208).

Esto es lo que se conoce comúnmente como teoría de las proporciones de Eudoxo. Esta propuesta permite sortear el problema del valor de la diagonal del cuadrado, porque no se basa en la relación entre números enteros, sino por medio de la relación entre magnitudes homogéneas. Y no sólo eso. Con la teoría de las proporciones se pudo «cubrir simultáneamente el caso de proporciones entre magnitudes conmensurables y el de proporciones inconmensurables, haciendo posible el estudio sistemático de ambos casos desde un punto de vista unificado» (Corry, 1994: 4). La teoría de las proporciones fue esencial para el desarrollo de los *Elementos* de Euclides, que básicamente era una exposición de la teoría en cuestión[141] (Penrose, 1991: 210). Bien sabido es que *Elementos* es una obra [por no decir «la obra»] fundamental para el ulterior desarrollo de la geometría, hasta el punto de que sigue influyendo en la geometría de nuestros días.

Al tratarse la geometría de una teoría física, como vimos arriba, ¿hasta qué punto la geometría euclídea permanece en la física posterior a la desarrollada en la Antigua Grecia? Penrose defiende ya no sólo la geometría de Euclides es un ingrediente esencial para cualquier teoría física actual, sino incluso que, por aquello que hemos visto, los planteamientos de Eudoxo también lo son (Penrose, 1991: 210). La vigencia de la geometría euclídea se sostiene gracias a su innegable capacidad explicativa del mundo empírico, a pesar de que existan otras geometrías que explican aquello en lo que la euclídea guarda silencio.

---

[139] La utilización de estos números, como bien se sabe, tuvo un mayor desarrollo en el siglo XIX, sobre todo con los trabajos de Dedekind, Weierstrass (Penrose, 1991: 208) y Cantor. Es por ello que de forma general se considera el trabajo de Eudoxo como el precedente [muy temprano] de los trabajos del siglo XIX. A pesar de que exista un consenso muy extendido acerca de esta relación directa, no por ello queda exenta de polémica. Autores como, por ejemplo, Unguru o Rowe ponen en tela de juicio este aspecto. Para los motivos que ofrecen estos autores, véase (Corry, 1994). En lo que respecta a la visión que seguiremos, corresponderá, obviamente, a la manejada por Penrose (es decir, dar por sentada la relación directa).

[140] Recuérdese el ejemplo del capítulo anterior, en particular el presentado en §4.1, que es al que Penrose se refiere.

[141] Para ver más sobre la importancia del papel de Eudoxo, tanto por las ventajas de sus contribuciones como por sus desventajas, véase (Kline, 1972: 48-50).

La física clásica, sin duda, tiene una deuda más que palpable con la geometría de Euclides. Sin embargo, también es evidente que la capacidad de esta para describir con *fidelidad* el mundo empírico tiene sus limitaciones. De tales limitaciones se harán cargo (antes que las citadas geometrías no-euclidianas) diferentes campos que surgieron posteriormente, los cuales que acabarían de edificar la física clásica. Los campos con mayor importancia, en gran medida, fueron la dinámica y la mecánica, las cuales veremos en el punto que sigue.

1.3. ¿Cómo de dinámica y mecánica es la naturaleza?

Es reconocido extendidamente que la cultura griega contribuyó de manera notable en la tarea de describir la Naturaleza. Pero si bien los griegos nos legaron una soberbia descripción *estática* de la Naturaleza, no menos cierto es que no podemos decir lo mismo acerca de la descripción *dinámica* (Penrose, 1991: 211). Pero entendamos esto de manera adecuada. Por supuesto que se llevaron a cabo estudios en los que se pretendía dar cuenta del movimiento en la Naturaleza, como bien puede comprobarse, por ejemplo, en la *Física* de Aristóteles. No obstante, fue la ciencia del siglo XVII aquella que verdaderamente sentó las bases de la dinámica y que en cierto modo sigue teniendo vigencia hasta nuestros días. Numerosas fueron las figuras que con sus trabajos consiguieron edificar el campo que hoy en día se conoce como dinámica. Sin embargo, las dos personalidades más influyentes fueron Galileo y Newton[142]. Así lo entiende Penrose y por ello se encarga de destacar algunas de las contribuciones más relevantes para, así, hacer manifiesta la influencia de ambos.

Con motivo de no desviarnos en exceso, no veremos en detalle todas las descripciones que destaca nuestro autor[143]. En lugar de ello podremos reparar en aquellos detalles que hicieron de Galileo y Newton los principales precursores de la dinámica tal y como la conocemos. Uno de ellos es, sin duda, la introducción de conceptos, o mejor dicho, la redefinición de conceptos, tales como *aceleración*, *fuerza*, *celeridad* o *velocidad* (Penrose, 1991: 212). En un análisis, en mi opinión, acertado de Penrose, este considera que el cambio de perspectiva de este tipo de conceptos fue clave en un hecho fundamental para el desarrollo de la ciencia (al menos la occidental): la defensa del heliocentrismo.

Durante siglos, el sistema aristotélico-ptolemaico había logrado explicar los movimientos de los astros. Sin embargo, si bien el sistema geométrico-matemático ptolemaico de epiciclos explicaba la estructura del universo, también es cierto que este se volvía cada vez más y más complejo. Ya desde la Antigüedad existieron teorías diferentes sobre la disposición geocentrista del universo de Aristóteles y Ptolomeo. Conocido es el caso de Aristarco de Samos (310 a. C.- 230 a.C.), quien propuso una teoría heliocéntrica, después de haber estudiado el tamaño y la distancia del Sol. No obstante, también es

[142] No es mi intención ignorar a otros grandes aportadores en el campo de la mecánica (como pueden ser Descartes o Leibniz –por citar algunos). Sencillamente sigo aquellos autores en los que se centra y expone Penrose.

[143] Para conocer en detalle a este apartado concreto véase (Penrose, 1991: 211-220).

sabido que su propuesta no tuvo una gran influencia en su tiempo. La verdadera puesta en duda en referencia al geocentrismo tendría lugar en los siglos XVI y XVII, de la mano de los trabajos de Galileo Galilei[144]. Galileo fue un gran estudioso de la teoría heliocéntrica de Copérnico (quien vivió entre 1473- 1543), sobre todo porque esta ciertamente le había convencido de que el sistema aristotélico-ptolemaico no conseguía explicar los fenómenos observados. A pesar de esta convicción, lo cierto era que el sistema copernicano tampoco lograba resolver de un modo contundente los enigmas de la estructura del universo. Es más, ponía encima de la mesa aún más problemas, para los cuales, para más inri, no se encontraban respuestas satisfactorias. Pese a ello, Galileo consideraba que el heliocentrismo de Copérnico con los ajustes debidos otorgaría resultados más fieles a la realidad[145]. Al fin y al cabo el nuevo modelo ofrecía consecuencias armoniosas y explicaciones simples para hechos que hasta ese momento resultaban misteriosos (Solís, Sellés, 2009: 363). Dotar de una mayor armonía y sencillez a la estructura del universo era una de las principales motivaciones del proyecto copernicano:

> Así, suponiendo los movimientos que atribuyo a la Tierra más adelante en esta obra, encontré al cabo de muchas y largas observaciones que, si se transferían los movimientos de los otros planetas a la rotación de la Tierra y si se tomaba por base a esta en la revolución de los astros, no sólo los fenómenos de los otros astros se seguían de esto, sino también el orden y las magnitudes de todos los astros y esferas y el cielo mismo resultaban conectados de tal modo que no era posible cambiar ninguna de sus partes sin producir una confusión en todo el universo. Es por esta razón que en el curso de esta obra he seguido tal orden (Copérnico cit. por Burtt, 1960: 51).

A pesar de que su renovada armonía y sencillez podían resultar evidentes, el heliocentrismo copernicano presentaba problemas para probar un aspecto fundamental dentro de su esquema. Dicho aspecto era nada más y nada menos que el movimiento de la Tierra. El problema se plantea de una manera muy sencilla: si la Tierra se mueve alrededor del Sol, ¿por qué los movimientos que se dan en ella se corresponden a un escenario inmóvil? La respuesta concreta de Copérnico era apelar a la esfericidad de la Tierra: los movimientos circulares son naturales para la Tierra, y «al no ser violento, lo que surge de la naturaleza se mantiene correctamente y se conserva en su

[144] Si bien el papel de Galileo tiene una importancia capital en el desarrollo de la teoría heliocéntrica, es de rigor destacar que este no fue el único que contribuyó a ello. Entrar en detalles sobre este asunto haría tender el discurso del trabajo hacia un corte historiográfico que no le concierne. Por ello lo mejor ante este tipo de afirmaciones es que se entienda que su uso es debido a la intención de ser práctico y en ninguno de los casos faltar a la verdad y la precisión de los hechos.

[145] Es extendida la idea de que el resurgimiento del heliocentrismo procedía de movimientos exclusivamente (en términos actuales) científicos, quedando relegadas a un segundo plano las corrientes místicas. Sin embargo, se conoce que esto carece de rigor histórico: […] Ya en el siglo siguiente, esta filosofía (misticismo) ayudó a los copernicanos a considerar el heliocentrismo de modo realista, porque una buena teoría matemática expresa la estructura de la creación, y la astronomía copernicana está llena de armonías geométricas, como el aumento de los períodos planetarios con la distancia del Sol. La versión kepleriana de esta filosofía llevó a investigar las causas físicas de los fenómenos celestes mediante el hallazgo de armonías matemáticas, como que los períodos planetarios son proporcionales a la potencia 2/3 de las distancias del Sol, pues revelan la razón de los planes de la creación divina (Solís, Sellés, 2009: 317).

composición óptima» (Solís, Sellés, 2009: 368). Es obvio que esta explicación no ofrece una respuesta satisfactoria, al menos no lo es más con respecto a la que apuesta por la estaticidad de la Tierra.

Pues bien, esto cambiaría, como hemos visto arriba, con la nueva conceptualización de términos dinámicos, concretamente el de *velocidad*. La teoría copernicana encontraba grandes dificultades para explicar el movimiento de la Tierra porque en ella la velocidad era concebida del mismo modo en el que lo hacía el sistema aristotélico-ptolemaico. En dicha perspectiva «aristotélica», la velocidad tiene un valor absoluto, mientras que en el replanteamiento de Galileo este concepto adquiere un carácter relativo. Esto es algo que queda en evidencia con el ejemplo del barco que presenta el pisano y que recoge Penrose:

> Encerraos con un amigo en la cabina principal bajo la cubierta de un gran barco, llevando con vos moscas, mariposas y otros pequeños animales voladores. Llevad un gran recipiente con agua y algún pez dentro; colgad una botella que se vacíe gota a gota en alguna vasija que esté debajo de ella. Con el barco aún en reposo, observad cuidadosamente cómo vuelan los pequeños animales con igual velocidad hacia todos los lados de la cabina. El pez nada indistintamente en todas las direcciones; las gotas caerán en la vasija inferior [...] Cuando hayáis observado cuidadosamente todas estas cosas [...] haced avanzar el barco con la velocidad que queráis, de forma que el movimiento sea uniforme y no haya oscilaciones en un sentido u otro. No descubriréis el menor cambio en ninguno de los efectos mencionados, ni podríais decir a partir de ellos si el barco se mueve o permanece quieto [...] Las gotas caerán como antes en la vasija inferior sin desviarse hacia la popa, aunque el barco haya avanzado mucho mientras las gotas están en el aire. El pez en el agua nadará hacia la parte delantera de su recipiente sin mayor esfuerzo que hacia la parte trasera, y se dirigirá con la misma facilidad hacia un cebo colocado en cualquier parte del borde del recipiente. Finalmente, las mariposas y moscas continuarán su vuelo indistintamente hacia cualquier lado, y no sucederá que se concentren hacia la popa como si se cansaran de seguir el curso del barco, del que hubieran quedado separadas una gran distancia de haberse mantenido en el aire (Galileo, cit. por Penrose, 1991: 213).

Desde luego que este principio de relatividad galileano hacía que la idea del movimiento terrestre fuese más aceptable. Sin embargo, es bien sabido los problemas que acarrearon a Galileo por su defensa del heliocentrismo. Pero este es otro tema. Aquello en lo que nos conviene centrarnos es en las ideas que Penrose destaca de la dinámica de Galileo. Por seguir con el principio de relatividad galileano, nuestro autor apunta que en este se sigue concibiendo el espacio y el tiempo de un modo intuitivo. Mientras que el tiempo es considerado como algo que *fluye*, el espacio es algo que *permanece*. Esto resulta conflictivo con un principio de dinámica que estableció el mismo Galileo, relacionado con el ejemplo del barco que hemos visto. Dicho principio es aquel que establece que el movimiento uniforme y rectilíneo no se distingue del «estado de reposo». Esto es algo que más tarde será compartido por Newton, ya que este entendió de un modo similar (que no igual) a Galileo[146] el espacio y el tiempo. Dejemos a un lado este último asunto por ahora y volvamos al problema que plantea Penrose.

[146] De hecho, la intuición de Galileo acerca del espacio y el tiempo se acerca más a la concepción de Einstein que la de Newton. Es por ello que suele entenderse que el verdadero precedente de la idea de relatividad einsteniana es dicha intuición galileana. Obviamente existen diferencias sutiles que diferencia ambas concepciones. Un ejemplo con el que

Nuestro autor propone que intentemos pensar en un objeto que en un instante se mueve de un punto del espacio a otro en un instante diferente, o incluso que ese objeto no se mueve y permanece en el mismo punto en dos instantes diferentes. El problema que surge es el siguiente: ¿cómo podemos entender que el espacio *permanezca* si no existe un significado absoluto para el «estado de reposo»? Pues *creando* un espacio nuevo por cada instante dado:

> […] ¡Parece como si debiéramos tener un espacio euclídeo completamente *nuevo* para cada instante de tiempo! La forma de que esto tenga sentido es considerar una imagen de la realidad física en un *espacio-tiempo tetra-dimensional*. Los espacios euclídeos tridimensionales correspondientes a los diferentes instantes de tiempo se consideran realmente como independientes uno de otro, pero todos estos espacios están unido para formar la imagen completa de nuestro espacio-tiempo tetra-dimensional. Las historias de las partículas que se mueven con movimiento rectilíneo uniforme se describen mediante líneas rectas (llamadas líneas de universo) en el espacio-tiempo (Penrose, 1991: 214).

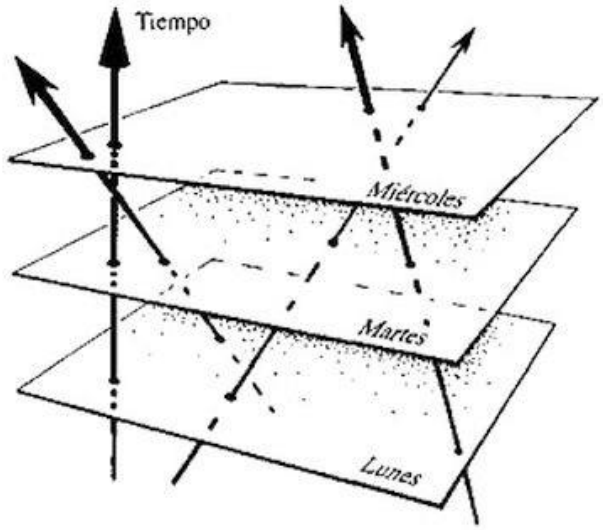

Figura 6. Este es el modo en el que Penrose representa el espacio-tiempo galileano, donde las partículas en movimiento uniforme se corresponden con las líneas rectas de la ilustración y en donde los nuevos instantes *dan lugar* a nuevos espacios.

Queda manifiesto que dicha resolución al problema no *encaja* con nuestra percepción de la realidad. La respuesta que más nos satisface en la actualidad

---

podemos ver la aproximación y a la vez la diferencia entre las concepciones de Galileo y Einstein es el principio de equivalencia, tal y como apunta Herce:
El principio de equivalencia de Galileo se puede formular como: «El movimiento de cualquier partícula de prueba en caída libre es independiente de su composición y estructura». Y en términos de Einstein: «El resultado de cualquier experimento no gravitacional en un laboratorio desplazándose en un sistema de referencia inercial es independiente de la velocidad del laboratorio o de su localización en el espacio-tiempo» (Herce, 2014: 77).
Son manifiestas las diferencias y es muy fácil incurrir en errores contextuales si se toma esta comparación al pie de la letra. Sin embargo, también es evidente que en ambas ideas puede verse aquello en lo que coinciden.

se la debemos a la nueva concepción del espacio y el tiempo por parte de Einstein de su teoría de la relatividad.

El *quid* del asunto no es que Penrose destaque los principios de la dinámica de Galileo para insinuar que estos contengan graves errores y que a fin de cuentas no tuvieron utilidad para el desarrollo ulterior del campo que estaba gestándose, sino más bien todo lo contrario. El rasgo de la dinámica galileana en particular que hemos visto es una prueba de ello. A pesar de que la doctrina que estaba implantando Galileo poseía postulados que podían chocar tan de frente con la realidad, ello no impedía que tuviera la capacidad de explicar una gran cantidad de fenómenos dinámicos (¡consiguiendo permanecer vigentes durante más de tres siglos!). La exposición de este problema está relacionado con el interés de Penrose por mostrar que existen diferentes tipos de teorías y qué tipo de categorías pueden otorgárseles[147]. Los estudios de Galileo ya no sólo hicieron posible la aparición y el asentamiento de la dinámica como campo explicativo de la naturaleza, sino que también sirvieron como base fundamental para el mecanicismo que desarrollaría Newton.

La dinámica[148] conseguía describir el aparente movimiento de los cuerpos físicos, pero la mecánica newtoniana tendría un alcance mayor, en lo que a trasfondo filosófico se refiere:

> Con una ley de fuerzas específica (como la ley de la inversa del cuadrado para la gravitación) el esquema newtoniano se traduce en un preciso y determinado sistema de ecuaciones dinámicas. Si se especifican las posiciones, velocidades y masas de las diversas partículas en un instante, entonces sus posiciones y velocidades (y sus masas, pues estas se consideran *constantes*) están matemáticamente determinadas para todos los instantes posteriores. Esta forma de *determinismo*, satisfecha por el mundo de la mecánica newtoniana, tuvo (y aún la tiene) una profunda influencia sobre el pensamiento filosófico (Penrose, 1991: 218).

Precisamente el determinismo es el tema que se encuentra en el fondo de todo el trabajo de Penrose, ya que nuestro autor defiende que la reforma que ansía para la física debería apuntar hacia el determinismo, tal y como en su día lo propusiera Einstein. Esto, sin embargo, lo veremos más adelante. De momento sigamos viendo aquello que Penrose expone acerca de la mecánica newtoniana.

Nuestro autor intenta hacer *visible* el mundo que presupone la mecánica newtoniana. Para ello invita que imaginemos a las partículas de las que están formadas el mundo físico como bolas de billar (ya que considera que es una imagen muy recurrente cuando hablamos de partículas). Con su exposición,

---

[147] Este tema concreto será tratado con un poco más de profundidad en el siguiente capítulo.

[148] En este paso de la dinámica a la mecánica, en mi discurso se puede ver que no se incluye el papel de grandes contribuidores al mecanicismo como Descartes, innegablemente importante (o el de Wren, Wallis, Huygens o Mariotte). El motivo de ello es porque Penrose los excluye en su obra. Esta nota no es reproche a nuestro autor, ya que entiendo que Penrose también desea economizar en sus explicaciones. Entrar a considerar a estos autores trae consigo la introducción del polémico concepto polémico de «materia» (concepto en el que no profundizan ni Galileo ni Newton)*. Además, es obvio que su exposición no pretende ser rigurosa en términos históricos.

*Aquello que se tuvo que decir sobre este concepto ya fue expuesto en el Capítulo 1 (§2.2).

a la cual va añadiendo propiedades que debe cumplir el mundo newtoniano[149], Penrose pretende hacer dejar claro el determinismo al que este está sometido. Y he aquí un asunto importante en el planteamiento penroseano, este es, cómo el determinismo no implica computabilidad:

> […] El problema del *determinismo* en la teoría física es importante pero creo que es sólo una parte de la historia. El mundo podría ser, por ejemplo, determinista pero *no-computable*. En tal caso, el futuro podría estar determinado por el presente de un modo que en *principio* es no-calculable (Penrose, 1991: 221).

Un ejemplo de elemento que cumple la condición de ser determinista y no-computable es la máquina de Turing. Recordemos que esta estaba programada (es decir, determinada a actuar de una manera concreta), pero de la cual no podía calcularse en términos algorítmicos si podría pararse, tal y como nos decía la solución al problema de la parada. Al fin y al cabo Penrose debe acudir a ejemplos que constituyen experimentos mentales, ya que está abordando temas referidos a lo más profundo de la naturaleza, es decir, a aquello de lo que no tenemos medios *concretos* para saber *ciertamente*. Es por este motivo por el que decide seguir con el ejemplo de las bolas de billar, aunque también porque este es innegablemente ilustrativo. El asunto final es que el mundo de las bolas de billar newtoniano puede ser computable siempre y cuando se condicione de un modo concreto el modo de calcular sus movimientos[150]. Esto quiere decir que mientras que las bolas de billar no *abandonen* el mundo ideal podrán seguir siendo computables. Pero una vez las *trasladamos* al mundo concreto su computabilidad se tambalea, ya que el factor de conocer las condiciones iniciales en nuestro mundo siempre estarán limitadas. Penrose aclara que es por esta razón precisamente por la que la meteorología puede hacer predicciones muy útiles a corto plazo, pero tales predicciones a largo plazo se vuelven cada vez más débiles. A pesar de todo, esta no es la no-computabilidad que le interesa a nuestro autor, ya que se basa en la simple introducción de un elemento aleatorio y *sólo* acabaríamos hablando de impredecibilidad, algo que para Penrose tiene una connotación negativa (Penrose, 1991: 226). Aquella no-computabilidad que aspira encontrar nuestro autor está relacionada, como hemos visto en algunas ocasiones, con un proceso no-calculable, algorítmicamente hablando. Penrose no cree que ese proceso deba ser no-calculable en términos absolutos. Es más, su discurso está encaminado en la dirección opuesta, quedando evidente en la reforma que plantea en la física actual, la cual veremos a lo largo de este capítulo.

Así que, ante la pregunta de hasta qué punto la naturaleza es dinámica y mecánica, toca responder que los estudios dinámicos de Galileo y mecanicistas de Newton nos permiten conocer aspectos de la naturaleza que nos resultan tremendamente prácticos. Aunque, por otro lado, debemos tener en cuenta que las descripciones que obtenemos con tales herramientas están centradas en partes concretas de la realidad. Si bien en la naturaleza encontramos ciertos rasgos dinámicos y mecánicos al modo en el que son descritos en sus estudios, ellos no permiten una descripción definitiva. Por ello Penrose encuentra conveniente seguir describiendo la física clásica, ya

[149] No a modo de hacer trampa, sino con la intención de hacer una explicación lo más clara y sencilla posible.
[150] Para consultar tales condiciones véase (Penrose, 1991: 225).

que aun cuando esta tiene sus límites bien demarcados su papel es fundamental para entender la física moderna. El siguiente paso al que procede Penrose es seguir definiendo la mecánica, pero abandona la newtoniana para centrarse en la hamiltoniana.

1.4. El complemento de la mecánica hamiltoniana

Desde muy tempranamente se acordó entre las élites intelectuales que el poder de la física newtoniana a la hora de describir la naturaleza era muy amplio y legítimo. No obstante, también fue evidente que el trabajo de Newton no era un trabajo plenamente acabado, siendo esto algo que él mismo reconocía[151]. El proyecto de Newton fue perfilándose a partir de los trabajos de Euler, Lagrange, Liouville, Jacobi, Laplace y Hamilton, entre otros, quienes construyeron lo que hoy se conoce como mecánica clásica. Si bien Penrose reconoce el papel de todos los *contribuyentes* al proyecto newtoniano, encuentra oportuno centrarse en uno en particular, este es, Hamilton, en concreto en su mecánica.

¿Qué tiene de especial la mecánica hamiltoniana para que Penrose se centre sólo en ella? Siendo una constante en su obra, nuestro autor intenta sintetizar (sobre todo en NME y SM, ya que en EcalR se permite extender sus explicaciones), y en ese ejercicio elige aquello que ayude a entender mejor la exposición ulterior. Para Penrose, la mecánica hamiltoniana permite captar lo esencial de la ciencia anterior y también de dejar claro cómo influye en la que le sigue[152]. Pasemos a ver los aspectos que Penrose destaca de esta mecánica.

La mecánica hamiltoniana, al igual que la mecánica anterior y contemporánea a esta, había heredado ya no sólo la aplicabilidad al mundo físico que había conseguido la mecánica newtoniana, sino también su potente teoría matemática. Pero obviamente no toda ella se limitaría a seguir de manera religiosa todos los pasos marcados por la física newtoniana.

La principal diferencia entre la mecánica hamiltoniana y newtoniana es sutil pero a la vez significativa. Dicha diferencia consiste en centrar la atención, en las descripciones de los sistemas físicos, en variables distintas a las que hasta ese momento habían sido principales. Mientras que en la mecánica newtoniana las variables imprescindibles en el estado inicial de un proceso mecánico son la *posición* y la *velocidad* de las partículas (por seguir con el

[151] Burtt, en su análisis de la metafísica newtoniana, destaca cómo la concepción de realidad manejada por Newton repercute decididamente en esta idea de que su contribución no es definitiva. Esto no hay que entenderlo indebidamente. Cuando se lee a Newton es fácil comprobar que este cree firmemente que está ofreciendo una descripción fiel de la naturaleza. Sin embargo, también es evidente que el sabio inglés reconoce los límites del entendimiento humano. Su postura viene a decir que siempre podremos seguir añadiendo un conocimiento más certero, a pesar de que estemos en el camino correcto.

[152] Tal y como apunté en la nota 131, el guion seguido es el de NME. Sin embargo, Penrose no abandona esta consideración en sus obras posteriores. En EcalR puede leerse:

La visión hamiltoniana proporciona un ejemplo maravilloso. Aunque la mecánica clásica que encarna está contradicha por algunos hechos brutos del mundo cuántico, la estructura hamiltoniana nos ofrece un camino importante hacia la teoría real de la mecánica cuántica. Más aún: las versiones cuánticas de los hamiltonianos proporcionan ingredientes esenciales para el formalismo cuántico estándar (Penrose, 2006: 656).

ejemplo que utiliza Penrose), en la hamiltoniana la variable fundamental pasará a ser el *momento*[153]. Pero, ¿tan radical resulta este cambio de atención en las variables? Definir esta diferencia como *radical* tal vez no se corresponda con la realidad del cambio, pero, por otro lado, detectar que el cambio es *considerable* sí que resulta un hecho innegable. Sobre todo, el cambio es relevante porque en la mecánica hamiltoniana la posición y el momento son considerados variables independientes la una de la otra, teniendo ambas, además, un rango similar (Penrose, 1991: 227). Este hecho permite que la mecánica hamiltoniana tenga dos conjuntos de ecuaciones, lo cual, en principio, sirve para que los resultados sean más precisos.

Otros aspectos sutiles que definen la mecánica hamiltoniana podemos encontrarla en esta descripción de Belot, en la obra de Butterfield y Earman, *Handbook of Philosophy of Science*:

> La idea básica detrás del enfoque hamiltoniano es trabajar con el espacio de datos iniciales de las ecuaciones de la teoría en lugar de con el espacio de soluciones a las ecuaciones; en términos generales y heurísticos, esto significa trabajar con el espacio de estados instantáneos de la teoría en lugar de con su espacio de mundos posibles.
>
> Las ecuaciones deterministas de movimiento nos dicen cuál debe ser el estado del sistema en tiempos anteriores y posteriores si está en un estado inicial dado. Entonces, al menos para las ecuaciones de movimiento con buen comportamiento, el contenido dinámico de las ecuaciones de movimiento debe ser codificable en un flujo en el espacio de datos iniciales, con las curvas integrales de este fluir siendo las trayectorias dinámicamente posibles a través del espacio de estados instantáneos […].
>
> Intuitivamente hablando, un estado instantáneo del campo es una especificación en cada punto del espacio del valor del campo y su tasa de cambio de tiempo; y al dar una secuencia de tales estados instantáneos, describimos cómo los valores de estas variables evolucionan a través del tiempo en cada punto del espacio. Entonces, para construir una formulación hamiltoniana de una teoría en la que la historia total de un sistema se describe a través de una trayectoria a través del espacio de datos iniciales, necesitamos efectuar algún tipo de descomposición nocional del espacio-tiempo en espacio y tiempo (Belot, 2007: 165).

La mecánica hamiltoniana, al igual que la newtoniana, es considerada determinista, en el sentido de que en ella los resultados obtenidos están determinados (es decir, que no pueden ser otros) por los datos iniciales. A pesar de que esta es una consideración ampliamente aceptada, Penrose se pregunta: ¿hasta qué punto se puede dar cuenta de este determinismo? ¿Y de la computabilidad[154]?

Una de las características principales del enfoque hamiltoniano es que el esquema formal en el que describe los sistemas físicos (espacio de fases[155]) tiende hacia cierta estabilidad. Penrose destaca este hecho porque con él considera que podemos detectar algún tipo de determinismo en tales sistemas

---

[153] Según lo señala Penrose, el momento de una partícula es el producto de su velocidad por su masa (Penrose, 1991: 227), siendo la velocidad una magnitud vectorial y no escalar. Este último detalle es lo que diferencia el momento de la «cantidad de movimiento» cartesiana.

[154] Recordemos que una de las tareas que Penrose se propone es intentar esclarecer la diferencia entre determinismo y computabilidad.

[155]Es conveniente que nos saltemos parte de la exposición de Penrose, porque ella acaba derivando en un terreno técnico (a pesar del esfuerzo de nuestro autor por evitarlo) que no nos concierne aquí. En su lugar veremos las conclusiones filosóficas a las que llega Penrose. Para dicha exposición, véase asimismo (Penrose, 1991: 229-239).

físicos. Esto, sin embargo, no puede considerarse una respuesta definitiva. Por su parte, Penrose conoce suficientemente bien lo problemático que resulta intentar demostrar que en última instancia los procesos naturales son deterministas. De hecho, la estabilidad que muestra el esquema formal no es plena; y no sólo eso, sino que cuanto mayor sea el intervalo de tiempo en el que se desarrolle el movimiento estudiado, más impreciso se volverá el resultado:

> Podemos preguntar […] ¿cómo es posible hacer alguna predicción en mecánica clásica? Esta es, en verdad, una buena pregunta. Lo que esta difusión nos dice es que, independientemente de la precisión con que conozcamos el estado inicial de un sistema (dentro de límites razonables), las imprecisiones tenderán a crecer con el tiempo y nuestra información inicial puede hacerse casi inútil. En este sentido, la mecánica clásica es esencialmente *impredecible* (Penrose, 1991: 237).

Ante este escenario, Penrose se hace la pertinente pregunta acerca de cómo es posible que la mecánica clásica pueda hacer predicciones y, además, de un modo tan satisfactorio. Nuestro autor defiende que el éxito de la mecánica clásica se debió a la capacidad que tiene esta para explicar los movimientos celestes (recordemos que desde Galileo la idea aristotélica de que los movimientos celestes y terrestres eran distintos comenzó a ser abandonada). No obstante, Penrose también apunta que es en este motivo donde podemos encontrar sus insuficiencias como teoría explicativa de los procesos naturales. En mecánica celeste el estudio se centra en un número concreto [y, por tanto, limitado] de cuerpos y, además, las leyes que se aplican a las partículas individuales no son tenidas en cuenta (Penrose, 1991: 238). Es decir, que la mecánica clásica ofrece una descripción *general* muy acorde a los fenómenos observados. Pero, casi de manera paradójica, cuando se intenta *precisar* en las descripciones, teniendo en cuenta más aspectos del movimiento, más imprecisa se vuelve esta mecánica.

Y he aquí la importancia de la mecánica hamiltoniana: a pesar de que esta pertenezca a la mecánica clásica, también es aquella que estableció las deficiencias de la misma (¡precisamente por el grado de precisión que puede alcanzar!). Esto, según Penrose, es lo que hace que esta mecánica sirva de complemento (él no utiliza estos términos, pero sí que expresa la idea) tanto para la física a la que pertenece como para aquella que se desarrollará después de esta.

La mecánica hamiltoniana nos dice que la física clásica no puede describir en un modo preciso el mundo en el que vivimos. Y este es el motivo por el que Penrose defiende que esta no nos puede ofrecer una respuesta acerca de la distinción entre el determinismo y la computabilidad.

1.5. La física clásica, incapaz de distinguir determinismo y computabilidad

¿Cómo una ciencia tan soberbia como la física clásica encuentra en la distinción entre determinismo y computabilidad un límite que sobrepasa sus capacidades? La respuesta es relativamente sencilla. Aquello que hacemos con dicha distinción es plantear una cuestión acerca de los principios últimos

de la naturaleza. Entonces, ¿cómo esperar una respuesta definitiva de la física clásica, que no deja de ser una *simple* teoría? Lejos quedan ya los días en los que los principios donados por Newton se concebían como garantes de las respuestas a todas las preguntas sobre los procesos naturales. Y es que la precisión que requerían estos principios para dar tales respuestas sólo condujo a la certificación de que la unidad anhelada se desvanecía en detrimento del azar, a pesar de que era este aquello que se pretendía eliminar. Resultó un craso error buscar esa unidad, ya que si tenemos en cuenta, precisamente, la naturaleza que observamos esta nos aleja de esa realidad (en cierto modo) parmenídea:

> […] La eliminación de este reducto final de lo azaroso podría vislumbrarse si la suprema norma se autoconstituyese a sí misma –se diese sus propios valores iniciales- o si al menos el problema del comienzo se diluyese en un indefinido regreso hacia atrás en el tiempo. Pero todavía quedaría por explicar la presencia de constantes dentro de ella que debieran ser determinadas fácticamente y en último término cabría considerar *casual* la forma misma de dicha ley. En definitiva, es como si la única alternativa válida a la aceptación del azar –aparte de recurrir a una trascendencia que nos obligaría a hablar no tanto de azar como de libertad o providencia-, fuera la reintroducción más o menos disimulada de una filosofía de la identidad de rasgos parmenídeos: un ser que no difiriese de su propia e interna necesidad. Cualquier otra cosa tendría siempre algo de gratuito que no cuadra con la exclusión del azar (Arana, 2012: 140).

Un escenario tal en el que el azar se abre paso, tiene como principal consecuencia que la prueba del determinismo en los procesos naturales a través de la física clásica, al menos, quede entredicho[156]. Aunque Penrose defienda que esta aparición del azar está relacionada con deficiencias de la física clásica en sí misma, aquello que nos dice la física que la sucedió es que existen ciertos grados de azar que son parte esencial de la naturaleza. Y he aquí una de las bases fundamentales de unos de los asuntos con mayor calado filosófico en el pensamiento de Penrose: su defensa del determinismo.

En las célebres discusiones que mantuvieron Einstein y Bohr acerca de la mecánica cuántica[157], el asunto del determinismo ocupó un plano principal. Como bien se sabe, en ese choque fue Bohr (contrario al determinismo) quien salió victorioso. Pues bien, Penrose de algún modo vuelve a retomar este debate intentando inclinar la balanza del lado de Einstein. Veamos a continuación en qué sentido nuestro autor es seguidor de Einstein.

[156] Este aspecto me ha recordado a la idea, con sus diferencias ya no pertinentes sino fundamentales, de Popper acerca de que el determinismo no puede ser explicado en la física clásica. La diferencia entre lo que decía Popper y lo que dice Penrose es que el primero piensa que el determinismo no puede explicarse porque en última instancia la física clásica también es indeterminista; y el segundo que el determinismo no puede explicarse porque la física clásica es incompleta. Son dos posturas antagónicas que se encuentran por la idea común de la incompletitud de las teorías. Para la postura de Popper sobre este asunto en particular véase (Popper, 1950: 117:133).

[157] De las cuales veremos algún que otro detalle más adelante en este mismo capítulo.

1.6. La relación Penrose-Einstein

En la primera página de la que es una de sus obras más filosóficas, *Mi visión del mundo,* Einstein hace una declaración de intenciones de cuáles son sus convicciones metafísicas del universo en los siguientes términos:

> No creo en absoluto en la libertad del hombre en un sentido filosófico. Actuamos bajo presiones externas y por necesidades internas. La frase de Schopenhauer: «Un hombre puede hacer lo que quiere, pero no puede querer lo que quiere», me bastó desde la juventud. Me ha servido de consuelo, tanto al ver como al sufrir las durezas de la vida, y ha sido para mí una fuente inagotable de tolerancia. Ha aliviado ese sentido de responsabilidad que tantas veces puede volverse una traba, y me ayudó a no tomarme demasiado en serio, ni a mí mismo ni a los demás. Así pues, veo la vida con humor (Einstein, 2013: 11).

A pesar de que en ninguna parte del texto citado aparezca la palabra «determinismo», este concepto se encuentra de un modo totalmente explícito en él (cuando niega la libertad y apelando a la necesidad). Como dice, la idea del determinismo lo convenció prontamente y fue algo que permaneció en su esquema intelectual.

Einstein era perfecto conocedor de las consecuencias [ya no sólo físicas, sino también metafísicas] que traía consigo la teoría cuántica que se empezó a forjar a partir de los estudios de Planck. Tales consecuencias no le convencían de ninguna de las maneras, ya que estas implicaban tener que renunciar al determinismo que tan importante era en su esquema intelectual (e incluso vital, como puede verse en la cita de arriba). En este célebre fragmento de una carta que envió a su amigo Max Born en 1924 (unos años antes de los debates que tuvo con Bohr en Solvay -1927), puede observarse la disconformidad de Einstein ante la física que se estaba consolidando:

> […] La opinión de Bohr sobre la radiación me interesa mucho. Pero no me obligarán a renunciar a la causalidad estricta sin defenderla más que hasta ahora. La idea de que un electrón expuesto a radiación elija por su propia voluntad el momento y la dirección en que dará el salto me resulta insoportable. En ese caso, preferiría ser zapatero o empleado de una timba y no físico. Verdad que mis intentos de dar forma tangible a los cuantos hasta ahora me han fallado, pero no pierdo la esperanza. Y aunque no logre nada, siempre me quedará el consuelo de que fue culpa mía (Born, 1971: 108).

Podemos ver que la concepción einsteniana del determinismo es extrema, ya que sigue concibiendo los procesos naturales como deterministas pese a que las pruebas científicas aportadas por la teoría cuántica (cada vez más fuerte) digan lo contrario. Einstein se mantiene en la idea de que si nos es imposible probar el determinismo es por nuestra propia ignorancia, en ninguno de los casos porque este no exista.

Por su parte, Penrose en ninguno de sus escritos es tan claro como Einstein con respecto al determinismo. No obstante, sí que es manifiesta la simpatía que nuestro autor profesa por Einstein y sus ideas[158]. Cuando Penrose habla

[158] De hecho, encuentra en la vida de Einstein un símil con su situación personal, concretamente en la que concierne a su búsqueda del procedimiento no-algorítmico. Aunque, considera, que Einstein tuvo algo más de suerte que él: […] Einstein fue llevado a su punto de vista revolucionario a partir de algunas consideraciones poderosas, algunas matemáticamente complejas, y algunas físicamente sutiles; pero la más importante de estas

de su postura determinista lo hace casi de pasada. Esto es una constante en los planteamientos penroseanos y se debe, en mi opinión, a que nuestro autor es muy prudente a la hora de entrar en debates filosóficos. Dicho aspecto podría entenderse que es porque estima que son temas que no merecen atención. Pero nada más lejos de la realidad. Penrose apunta más bien en la dirección contraria, en el sentido de que es partidario de no posicionarse en un debate del que no puede dar una opinión experta, precisamente por considerarlo importante. Pero volviendo a la relación que mantiene nuestro autor con las ideas de Einstein, esta es más notable si nos centramos en sus argumentos científicos. Como veremos más adelante en este capítulo, cuando Penrose plantea la posibilidad de reformar la física cuántica elige las ideas de Einstein y Schrödinger[159]. Mientras, otras como las de Bohr o Heisenberg, de indudable importancia en la física moderna, casi no son tomadas en cuenta. Penrose prefiere, por tanto, permanecer en aquel campo en el que es experto para fundamentar su apoyo a la perspectiva einsteniana sobre el determinismo.

Otro de los aspectos en los que podemos observar que las ideas de Penrose y Einstein confluyen es en el poder que ambos otorgan a las matemáticas. Ya hemos visto en el capítulo anterior aquello que Penrose piensa acerca de este tema en concreto. La postura de Einstein es llamativamente similar a la de nuestro autor, aunque, evidentemente, con sus diferencias pertinentes.

Einstein entiende que las matemáticas son esenciales a la hora de estudiar los entresijos de la naturaleza. De hecho, entiende que cuanto más puras son las matemáticas más capacidad de explicación de la naturaleza puede otorgarnos:

> […] Creo que a través de una construcción matemática pura es posible hallar los conceptos y las relaciones que iluminen una comprensión de la Naturaleza. Los conceptos usables matemáticamente pueden estar próximos a la experiencia, pero en ningún caso pueden deducirse de ella. Está claro que la experiencia es el único criterio que tiene la Física para determinar la utilidad de una construcción matemática. Pero el principio creativo se encuentra en realidad en la matemática. De algún modo creo que es cierto que a través del pensamiento puede comprenderse la realidad, tal como lo soñaron los antiguos (Einstein, 2013: 140).

En esta cita aparece un rasgo interesante y que en parte también está relacionado con lo defendido por Penrose. El rasgo en cuestión es aquel que comprende la idea de la construcción en las matemáticas. Como vimos en el capítulo anterior, generalmente se entiende que al hablar de construcción en matemáticas se está adoptando una postura afín al intuicionismo. Pero también vimos que esto no tiene que ser necesariamente así, tal y como lo intenta explicar Penrose. Es cierto que la explicación de nuestro autor no llegaba a ser todo lo consistente que a él le hubiese gustado, pero es de rigor advertir que su idea sí que es clara al respecto: construcción sí, pero no al

---

había permanecido patente, aunque inapreciada en su justa medida, desde los tiempos de Galileo (el principio de equivalencia: todos los cuerpos caen a la misma velocidad en un campo gravitatorio). Además, un requisito previo necesario para el éxito de las ideas de Einstein era que deberían ser compatibles con todo lo que se conocía en los fenómenos físicos de su época (Penrose, 2012: 245).

[159] Sobre la influencia de Schrödinger tendremos la oportunidad de ver algún detalle más adelante. Obviamente esta influencia es importante, pero igual de manifiesto es que la de Einstein es mucho más marcada.

modo del intuicionismo. Pues bien, Einstein no entra en debate contra la postura intuicionista en concreto. No obstante, sí que podemos ver manifiesta que su perspectiva se ve enfrentada a esta corriente de un modo similar a como lo hace bajo el punto de vista de Penrose. Ambos, Einstein y Penrose, coinciden en tener en cuenta la experiencia, pero también inciden en que la construcción no se basará nunca en esta, sino en las mismas matemáticas. En la cita de arriba vemos que Einstein sugiere que en este aspecto da la razón a los antiguos (sin temor a equivocarme entiendo que se refiere a la filosofía platónica), mientras que Penrose es mucho más explícito[160].

Siendo sus posturas plenamente coincidentes o no, lo cierto es que guardan rasgos muy similares y Penrose no es tímido a la hora de declarar que es deudor de Einstein en muchas de sus ideas, ya no sólo en el terreno de la física sino también en el de la filosofía.

Son precisamente las ideas filosóficas las que sitúan a Einstein (y en cierto modo también a Penrose) en la perspectiva de la física clásica, porque en lo que concierne al plano científico bien sabido es que este contribuyó de manera definitiva en la construcción de la física que pasará a ser denominada como moderna. Einstein supo de la potencialidad de la física que estaba abriéndose paso, pero también estaba convencido de que esta también tenía en su seno límites que no debían ser obviados. Si bien su tarea por mostrar tales limitaciones no dieron el fruto que él esperaba (algo que podemos ver en la cita de arriba), Penrose encuentra en ella el camino a seguir. Y es en esta idea einsteniana sobre la que se apoya nuestro autor para comenzar a bosquejar la reforma en la física actual que pretende.

2. La fuerza de la física moderna

2.1. Crisis de la física clásica e irrupción de la física moderna

Pero antes de pasar a la propuesta de Penrose y los argumentos en los que sustenta su defensa de una posible reforma en la física moderna, en este apartado veremos el cambio que supuso el asentamiento de esta nueva física. Lo que viene a continuación no será un repaso histórico estrictamente hablando, aunque sí que pretende contextualizar el debate que, en mi opinión, pretende retomar Penrose de alguna forma.

Hoy en día los límites de la física clásica se conocen extensamente, sin embargo, no fue hasta hace relativamente poco cuando estos *saltaron* y dieron paso a lo que actualmente se conoce como física moderna. Fue justo al entrar

[160] De hecho, la postura einsteniana no puede ser considerada del todo platónica, ya que en ocasiones parece defender ideas más acordes al formalismo:

[…] El avance logrado por la axiomática consiste precisamente en que a través de ella se trazó una frontera nítida entre lo lógico-formal y el contenido práctico. Únicamente lo lógico formal constituye, con arreglo a la axiomática, el objetivo de las matemáticas. No así la intuición ni cualquier otro tema vinculado a lo lógico-formal (Einstein, 2013: 150).

Sin embargo, Einstein no deja de decir lo que dice. Cuando dice que mediante el pensamiento podemos entender la realidad, está apelando a un idealismo propio de la filosofía platónica. Pero debido a que Einstein no se extiende en exceso en este tema lo conveniente es identificar en qué modo coincide con unas corrientes y con otras.

el siglo XX, en concreto en el año 1900, el momento en el que la física comenzaría un nuevo camino en sus explicaciones de los procesos naturales. Dicho año queda marcado, como bien se sabe, por los estudios de la radiación del cuerpo negro llevados a cabo por Max Planck. Este físico alemán dio el primer golpe significativo que haría temblar los cimientos de la física clásica al introducir la discontinuidad en el experimento mencionado. Lo llamativo del asunto es que la solución que propuso Planck no era aquella con la que pretendía dar. De hecho, Planck en no pocas ocasiones recalcó que se vio *obligado* a admitir la discontinuidad, ya que con esta sí que quedaban explicados los fenómenos. En la siguiente cita veremos esta actitud y además las razones que forzaron su decisión:

> […] resumido brevemente, se puede describir lo que hice como un acto de desesperación. Por naturaleza soy pacífico y rechazo toda aventura dudosa. Pero por entonces había estado luchando sin éxito durante seis años (desde 1894) con el problema del equilibrio entre radiación y materia y sabía que este problema tenía una importancia fundamental para la física; también conocía la fórmula que expresa distribución de la energía en los espectros normales. Por consiguiente, había que encontrar costase lo que costase, una interpretación teórica. Tenía claro que la física clásica no podía ofrecer una solución a este problema, puesto que con ella se llega a que a partir de cierto momento toda energía será transferida de la materia a la radiación. Para evitar esto se necesita una nueva constante que asegure que la energía se desintegre. Pero la única manera de averiguar cómo se puede hacer esto es partiendo de un punto de vista definido. En mi caso, el punto de partida fue el mantener las dos leyes de la termodinámica. Hay que observar, me parece, estas dos leyes bajo cualquier circunstancia. Por lo demás, estaba dispuesto a sacrificar cualquiera de mis convicciones anteriores sobre las leyes físicas. Boltzmann había explicado cómo se establece el equilibrio termodinámico mediante un equilibrio estadístico, y si se aplica semejante método al equilibrio entre materia y radiación, se encuentra que se puede evitar la continua transformación de energía en radiación suponiendo que la energía está obligada, desde el comienzo, a permanecer agrupada en ciertos cuantos. Esta fue una suposición puramente formal y en realidad no pensé mucho en ella (Planck, cit. por Sánchez Ron, 1995: 199-200).

El motivo principal por el que le costó tanto aceptar la idea de las respuestas estadísticas y la discontinuidad fue su convicción filosófica determinista. Planck deseaba desdeñar por todos los medios la posibilidad de que en los procesos naturales las leyes estadísticas tuviera cabida, porque realmente creía en las leyes absolutas. En realidad esta fue una actitud que no abandonó nunca, ya que (como hemos visto en la cita) la discontinuidad y la estadística no dejaba de ser una mera «suposición formal»[161]. Pero con lo que probablemente no contaba Planck era que dicha suposición había venido para quedarse.

Pronto comenzaría a verse que las ideas de Planck tendrían un tremendo alcance y cómo ellas desplazaban a la física clásica, ya no sólo en lo que respectaba al aspecto práctico del campo en sí mismo, sino también en la concepción filosófica. De repente, la visión del mundo determinista, continua y absoluta que se había forjado con la física clásica parecía volverse borrosa, o al menos no era tan clara como se creía. Esta situación, sin embargo, no provocó que la idea del indeterminismo fuese aceptada de forma instantánea. Existían sectores más conservadores, los cuales se negaban a admitir el

[161] La cita pertenece a una carta que Planck envía a un colega físico en 1931, es decir, 31 años después de su descubrimiento.

indeterminismo que traía consigo la física emergente. En ese lado conservador permanecía Planck, y Einstein, quien curiosamente también había contribuido al asentamiento de la teoría cuántica en 1905 con sus estudios acerca de la naturaleza de la luz. Ellos dos, sin duda, fueron los grandes estandartes de la «resistencia al indeterminismo» desde una perspectiva determinista al modo clásico. Pero pasemos a ver de un modo más detallado a las principales personalidades que formaron los bandos determinista e indeterminista.

2.2. Bandos determinista e indeterminista

Hemos visto que tanto Planck como Einstein fueron imprescindibles para el nacimiento y el asentamiento de la teoría cuántica. Pero es de rigor señalar que la teoría cuántica[162], obviamente, no surgiría de forma exclusiva desde bando determinista.

Comenzando por el bando indeterminista, la figura de Niels Bohr tiene un peso importante. Tal es así que generalmente es considerado, no sin razón, como el mayor representante de esta postura. No es ningún secreto que desde sus inicios en el campo de la investigación física la idea de los saltos cuánticos fue objeto de su interés. Muchos(as) consideran que el tutelaje de Rutherford sin duda tendría una influencia capital en los intereses en tales intereses. Prueba de ello fue una de sus primeras grandes aportaciones al campo de la física, como lo es su modelo atómico (1913), que seguía las ideas de su mentor. Más adelante veremos la importancia del modelo atómico para el pensamiento de Bohr. De momento, sigamos con qué otras influencias recibió (y reconoció) el físico danés.

Bohr no vacilaba a la hora de admitir a Einstein y Planck como determinantes en su convicción acerca del indeterminismo. El físico danés siempre exteriorizó su admiración por Einstein, a quien, incluso, una vez escribiría: «Conocerle y hablar con usted fue una de las mejores experiencias de mi vida. Jamás olvidaré nuestra conversación de vuelta del Dahlem a su casa» (Lindley, 2008: 92).

Con respecto a la influencia de Planck, en la colección de artículos recogidos en la obra *La teoría atómica y la descripción de la Naturaleza*, Bohr se expresa en los siguientes términos:

> Los artículos que la componen se escribieron en una época en la que el programa de desarrollar un tratamiento general de los problemas atómicos sobre la base del descubrimiento original de Planck del cuanto de acción alcanzó un sólido fundamento al establecerse el adecuado formalismo matemático (Bohr, 1988: 50).

La influencia de la que habla Bohr está relacionada, obviamente, con las ideas más revolucionarias de ambos científicos. Estas, como ya sabemos, no

[162] Un apunte que es necesario tener en cuenta antes de pasar a ver los dos lados del debate es que la física emergente contaba con un campo en el que mejor se traducían sus ideas, este es, la mecánica cuántica (tal y como sucedía con la física clásica, que encontraba en la mecánica su mejor valedor). Por ello, cuando más adelante me refiera a teoría, física, o mecánica cuántica, en realidad estaré haciendo mención a la misma cosa.

son aquellas que identifican el pensamiento de Einstein ni de Planck, que se encuentran en el bando contrario.

Otra de las grandes figuras del indeterminismo es, sin duda, Werner Heisenberg. La contribución por antonomasia por parte de Heisenberg es el principio de incertidumbre. Antes de elaborar dicho principio el físico alemán fue alumno de Sommerfeld en Múnich y más tarde vivió bajo la influencia de Pauli y Born en Gotinga. Tener como padrinos intelectuales a tales personalidades incuestionablemente marcaría el devenir de las investigaciones de Heisenberg, que lo convertirían en uno de los físicos más importantes de su época. Pero no sería hasta su estancia en Copenhague (a partir de 1924), con Bohr como su valedor, donde el trabajo de Heisenberg supusiera un hito en el desarrollo de la mecánica cuántica.

Si Bohr era aquella personalidad al que rodeaban aquellos que sentían simpatía por las ideas indeterministas, el papel de Max Born en esta faceta no era menos importante. De su influencia se beneficiaron grandes defensores del indeterminismo como Wolfgang Pauli, Werner Heisenberg o Pascual Jordan. Aunque si es reconocida la labor de Born, es cierto que, para muchos(as) la magnitud de la misma no tiene la notoriedad que merece. Born fue indudablemente uno de los estandartes de la conocida como Interpretación de Copenhague. Fue el padre de la concepción de la función de onda como probable y de la mecánica matricial (junto a Jordan y Heisenberg). Mantuvo amistad y correspondencia con Einstein, con quien discutía acerca de los problemas que surgían a partir de los logros de la nueva física[163]. Born fue un baluarte fundamental en el bando indeterminista.

Ahora pasamos a ver aquellos que defendían una postura determinista. De ellos ya conocemos los casos de Einstein y Planck. Estos, sin embargo, no serían los únicos. El quehacer de dos científicos en concreto ayudó a mantener el espíritu determinista de la física clásica. Estos eran Louis de Broglie y Erwin Schrödinger. Lo cierto es que la postura de estos no era (siendo el caso de Schrödinger algo más evidente) estrictamente determinista al modo clásico de Einstein o Planck. De hecho, se suele considerar el pensamiento de Schrödinger y de Broglie como semiclásico. Esta concepción reconoce en cierta medida lo que mostraba la teoría cuántica. Pero con la diferencia de que sigue defendiendo la idea de los procesos naturales como continuos, propia del determinismo.

De Broglie propondría en el debate acerca de la naturaleza de la luz una respuesta que refleja la clara postura de conciliación entre la física clásica y la moderna. Esta respuesta es la que se conoce como de «doble solución». Desde la irrupción de la teoría cuántica se pretendía conocer la naturaleza de

[163] Entre ambos existía un respeto y una admiración mutua. Y aunque estuvieran en las antípodas con respecto a sus convicciones filosóficas, siempre existió un espacio para el reconocimiento. Born llegó a decir sobre este hecho: Creo que no debe pasarse por alto la opinión del físico más grande del momento, quien, además, ha hecho más que nadie para establecer ideas modernas. Einstein no comparte la opinión de la mayoría de nosotros de que hay una evidencia abrumadora para aceptar la mecánica cuántica. Sin embargo, reconoce "éxito inicial" y "considerable grado de verdad". Obviamente está de acuerdo en que actualmente no tenemos nada mejor, pero espera que esto se logre más tarde, ya que rechaza al "dios que juega a los dados" [...]. El principio de Einstein de la existencia de un mundo real objetivo es, por lo tanto, bastante académico. Por otro lado, su afirmación de que la teoría cuántica ha renunciado a este principio no está justificada, si la concepción de la realidad se entiende correctamente (Born, 1949: 123).

la luz, habiendo dos posturas que prevalecían: la de quienes entendían dicha naturaleza como ondas y la de quienes la entendían como partículas. Sin entrar en detalles (ya que están reservados para un poco más adelante), la propuesta del físico francés, como el nombre de la solución indica, comprende que la naturaleza de la luz está compuesta tanto de ondas como de partículas. Esta sería una idea que mantendría con fuerza al bando de los deterministas.

El deseo de los deterministas de que apareciera una evidencia consistente de la naturaleza continua de los procesos naturales tiene que ser situado en la aparición en escena de la función de onda de Erwin Schrödinger. Esta función permitía que se concibiera a la naturaleza formada por ondas, lo cual hacía permanecer la noción de continuidad en los procesos naturales. De repente, la función de onda daba una respuesta satisfactoria para el bando determinista cuando este se estaba viendo acorralado por los incesantes logros de la física moderna.

El trabajo en física y las ideas filosóficas de Schrödinger guardan una estrecha relación. Y no es que el resto de científicos fueran contradictorios o poco claros con respecto a uno y otro aspecto, sino que en el trabajo de Schrödinger se nota el esfuerzo por ofrecer una visión interdisciplinar de aquello que desarrolla a través de la física.

Una vez visto a grandes rasgos los principales representantes que he decidido destacar de un bando y de otro, ahora veremos algunas de las ideas que son clave para entender las distintas aportaciones y en qué medida enriquecieron el debate entre ambas perspectivas.

2.3. Cambios puntuales [pero significativos] en la física

En esta parte veremos las principales contribuciones de las personalidades destacadas anteriormente. Esto servirá para entender en qué medida tales contribuciones influyeron para los argumentos de un bando y otro.

En primer lugar, podremos ver cómo Bohr comienza a formar su postura indeterminista a partir de su modelo atómico. En segundo lugar veremos la función de onda de Schrödinger. Este paso significa dar un salto de un bando al otro. La decisión de este modo de exponerlo tiene que ver con un motivo en parte cronológico y en parte epistemológico. Si decidiese explicar en primer lugar todo el bando indeterminista podría no entenderse del todo aquello que desarrollaré acerca de Born. Del físico alemán, en tercer lugar, examinaremos la concepción de la función de onda como probable. Más tarde conoceremos las relaciones de incertidumbre de Heisenberg. Y por último, veremos a de Broglie y su semiclásica «doble solución».

2.3.1. El modelo atómico de Bohr

El modelo atómico de Bohr supuso un antes y un después dentro del estudio de la naturaleza de los átomos. Este asunto fue objeto de interés del físico danés desde muy temprano en su carrera investigadora. Dicho interés y la

indiscutible capacidad de la que gozaba Bohr le llevaron a convertirle en un joven prolífico. Hasta tal punto era prometedora la carreta de Bohr que pudo conseguir que la figura por excelencia en ese momento en lo que a estudios atómicos se refiere le aceptara como su pupilo. Dicha personalidad no era otra que J. J. Thomson.

Los estudios de este científico británico consiguieron cambiar la concepción del átomo, asestando un duro golpe a aquella concepción que había predominado hasta entonces, la de Dalton. Es cierto que varios estudios anteriores habían puesto en tela de juicio el modelo de Dalton, pero fue Thompson quien acabó de condenarlo. La gran aportación de Thomson fue el descubrimiento del electrón dentro del átomo. El estudio de los rayos catódicos fue fundamental para la elaboración el nuevo modelo. Para Thomson, el átomo estaría formado por una carga positiva, la cual tendría, a su vez, pequeñas cargas negativas (electrones).

Este modelo, si bien fue innovador y útil, también tendría sus carencias. La introducción de los electrones como parte del átomo permitía dar respuestas a cuestiones que antes estaban veladas. Pero en lo concerniente a la distribución de la carga positiva, este modelo resultaba insuficiente para poder seguir dando respuestas.

Fue entonces el discípulo de Thomson, Ernst Rutherford, quien daría con otra clave acerca de la naturaleza del átomo. Rutherford conservó la idea del electrón dentro del átomo, pero con la diferencia de que la carga positiva tendría una localización concreta: el núcleo. Esta sería la primera vez que se utilizara el concepto de núcleo atómico. El modelo dice que el átomo está formado por unas cargas positivas (protones), que componen el núcleo y sobre el cual giran las cargas negativas (electrones). La propuesta de Rutherford conseguía solventar las carencias del modelo de su maestro, pero ello no evitaría que también tuviera rasgos que fueran susceptibles de corrección.

Bohr sería el encargado de poner encima de la mesa el modelo que solucionara las deficiencias de aquel planteado por Rutherford. En ese momento el físico danés ya había abandonado Cambridge y a Thomson, para unirse en Manchester a Rutherford, quien tendría una personalidad más afín que su antecesor con Bohr.

El modelo atómico de Bohr supondría un hito en el estudio del átomo porque este sería el primero en el que se contemplasen ideas de la física que surgía a partir de los estudios de Planck y Einstein.

Bohr seguía concibiendo el átomo formado por un núcleo de carga positiva, mientras que la carga negativa giraría a su alrededor. Una de las diferencias más significativas con respecto al modelo de Rutherford estribaría en que la carga negativa (que girara alrededor de la positiva –protón-) estaría compuesta de un solo electrón, el cual tendría una energía muy poco significante[164]. A priori esta característica no logra distinguir de un modo claro el modelo de Bohr y el de Rutherford, pero en ella existe un matiz que permite localizar dicha diferencia. Según el electromagnetismo clásico, los electrones que giraran en la órbita más próxima al núcleo acabarían colapsando. Esto supone un conflicto para el modelo de Rutherford, ya que

[164] Dicha composición está referida al átomo de hidrógeno, que es el átomo más simple. En el caso de átomos más complejos, es decir, con más electrones, estos girarían entorno al núcleo

la disposición de las cargas en dicho modelo no conseguiría explicar el comportamiento de los átomos. Por su parte, para el modelo de Bohr esto tal problema no existe, ya que este plantea que las órbitas siguieran siendo circulares, pero específicas, sin la proximidad al núcleo que consideraba el modelo de Rutherford. Este sería uno de los puntos clave para entender la introducción de las ideas cuánticas a la naturaleza del átomo y la diferencia con el modelo anterior. Si bien el modelo de Bohr daba cuenta del problema del no-colapso del átomo de tal forma, la dispersión de la carga eléctrica negativa no quedaría explicada. El electrón (en el caso del átomo de hidrógeno, que es aquel en el que se centra) tendría un recorrido de la órbita regular (en forma circular, como hemos visto), por lo que la dispersión de la carga tendría que ser necesariamente regular (¡condición que no se da!). Para explicar las irregularidades de la dispersión, Bohr propone el cambio de órbita del electrón. Este cambio respondería a la misma idea de Planck del salto cuántico. El electrón al emitir cierta cantidad de radiación salta hacia otra órbita, lo cual permite comprender no solo el no-colapso, sino también la estabilidad del átomo.

Este modelo haría patente la insuficiencia de la física clásica a la hora de querer describir la naturaleza del átomo:

> […] Sin embargo, un examen más detallado revela de inmediato que existe una diferencia fundamental entre un átomo y un sistema planetario[165] […]. En un modelo clásico del átomo la frecuencia de la radiación emitida está determinada por el período característico del movimiento, que, a su vez, depende de la energía del átomo; como consecuencia de la pérdida de forma continua y así la radiación prevista por la teoría clásica no tendrá similitud alguna con las líneas espectrales de los elementos (Bohr, 1988: 78-79).

La física clásica era insuficiente para explicar lo que el átomo estaba mostrando. Por ello, era el momento de introducir la física moderna. Pero la irrupción de las ideas de Planck en el modelo atómico de Bohr tampoco garantizaría una respuesta definitiva. Si bien la propuesta de Bohr conseguía explicar muchas de las características del comportamiento del átomo, no menos cierto era que su aplicación se tambaleaba cuando se analizaban átomos más complejos que el del hidrógeno. Este problema fue solventado por Sommerfeld, quien corrigió el modelo de Bohr. La diferencia más llamativa fue no entender las órbitas de forma circular, sino elípticamente. Ello no sólo permitiría un mejor análisis de los demás átomos, sino también explicaría de un modo más adecuado otras características del átomo de hidrógeno. Algo reconocido por Bohr años después en su célebre lectura en Como (1927):

> Las reglas de cuantificación han permitido explicar en gran parte la estructura fina de los espectros. Particularmente interesante fue la demostración de Sommerfeld de que las pequeñas desviaciones del movimiento kepleriano que resultan de la modificación de la mecánica newtoniana, exigida por la teoría de la relatividad, proporcionan una explicación natural de la estructura fina de las líneas del hidrógeno (Bohr, 1988: 85).

[165] En referencia al modelo de Rutherford.

2.3.2. La función de onda de Schrödinger

Lo que había quedado claro con la propuesta del físico danés era que las nuevas ideas venían para quedarse. Pronto aparecerían seguidores del indeterminismo que se observaba en los estudios atómicos. En Alemania, Max Born sería uno de los principales representantes de esta corriente. Este, a su vez, estaría rodeado de jóvenes físicos que se convertirían en grandes contribuidores del desarrollo de la mecánica cuántica. No sería casualidad que uno de los pilares de este campo fuera aportado por Born y dos de sus ayudantes, Pascual Jordan y Werner Heisenberg. El papel de Heisenberg sería definitivo para la formación de lo que se conocería como mecánica matricial. Esta mecánica trataba de dar cuenta matemáticamente de los nuevos resultados, entendidos dentro del marco de los saltos cuánticos. Si bien la solución que ofrecía fue positiva ello no impidió que apareciera una respuesta opuesta [en parte] casi a la par. Esta era la función de onda dentro de la mecánica ondulatoria de Erwin Schrödinger. La diferencia fundamental entre ambos estudios estaría en que el primero se pretendía contemplar los saltos cuánticos, mientras que para el segundo esta condición era pasada por alto. El papel de Schrödinger entonces se volvería central en la física a partir de entonces.

Las contribuciones científicas más importantes por parte de Schrödinger se sitúan en sus estudios en electromagnetismo. Sin duda la introducción de la función ψ [166] para la explicación de la ecuación de onda constituye un paso importante en toda la física posterior. En esta explicación de la función (aun en ese momento «campo mecánico escalar») el objeto de estudio era el átomo de hidrógeno, lo cual era lo común (por su simplicidad, como apunté más arriba).

Para Schrödinger, que no se pueda derivar de un modo consistente la intensidad y la polarización de las ondas electromagnéticas que el átomo de hidrógeno emite, tiene que ver con el significado que se le da a la función ψ. Según el físico austríaco es necesario atribuir a dicha función un significado electromagnético (Jammer, 1974: 24).

Dejando a un lado las diferentes cuestiones que tiene en cuenta Schrödinger para llevar a cabo el desarrollo de sus investigaciones (porque ello nos llevaría a un terreno técnico, el cual podría ser perjudicial más que beneficioso), destacaré el aspecto más importante en el plano filosófico que entraña la solución de la ecuación de onda.

Schrödinger plantea que la radiación emitida puede entenderse en términos de física clásica. Ello tiene sentido si entendemos que lo que sucede (como, por ejemplo, la densidad de carga eléctrica, etc.) como producto de ondas. Es decir, que la realidad física está compuesta exclusivamente por ondas. La característica fundamental de la naturaleza de las ondas es que esta es continua. Con lo cual podemos permanecer en el plano de la física clásica.

Visto de este modo la cuestión parece una predilección de gustos (en este caso filosóficos). Pero no es del todo correcto. El éxito de la función de onda vino cuando se pudo demostrar (de la mano del mismo Schrödinger) que el formalismo de esta y el de la mecánica matricial son equivalentes. Aunque

[166] Su nombre es «psi», por la letra del alfabeto griego. Esta función era denominada primeramente «campo mecánico escalar», para serlo luego como «función de onda».

esto no era todo. La función de onda tenía la ventaja de ser menos compleja y más intuitiva que la mecánica matricial. Este hecho realzaba la actitud de aquellos que no pretendían abandonar la idea de continuidad propia del determinismo:

> Estoy convencido de que ha realizado un avance decisivo… como también estoy convencido de que el rumbo que han tomado Heisenberg-Born es equivocado (Einstein cit. por Lindley, 2008: 130).

> (La mecánica de matrices)[167] era extremadamente intrincada y amenazadoramente abstracta. Schrödinger ha venido a rescatarnos (Sommerfeld cit. por Lindley, 2008: 134).

La función de onda permitía una exposición más clara que la mecánica matricial, pero su alcance no era ilimitado. De esto dieron cuenta incluso partidarios del determinismo. Tal y como apunta Jammer, el caso de Hendrik Antoon Lorentz es esclarecedor. Lorentz sentía entusiasmo por los resultados que ofrecía la mecánica ondulatoria, pero siempre y cuando el objeto de estudio de esta se limitara a lidiar con un número determinado de coordenadas.

> […] Sin embargo, «si hay más grados de libertad», escribió Lorentz, «entonces no puedo interpretar físicamente las ondas y las vibraciones, y por lo tanto debo decidir a favor de la mecánica matricial» (Lorentz cit. por Jammer, 1974: 32).

Schrödinger llegó para rescatar a aquellos que se negaban a aceptar los saltos cuánticos. Pero dicho rescate, como hemos visto, no sería definitivo. No mucho tardó el bando indeterminista en reaccionar también. La función de onda generaba probabilidades y las principales podían ser explicadas en cuatro vertientes[168].

La primera compendia la inexactitud que puede llegar a suponer representar una partícula mediante un paquete de ondas, tal y como exigía el esquema de Schrödinger. Dicha inexactitud se debe a que la concentración de carga eléctrica de partículas, la cual se da en pequeñas zonas de espacio no concuerda a la que corresponde a los paquetes de ondas, la cual se dispersa en amplias zonas. Schrödinger era consciente de este problema y tenía una respuesta, que finalmente resultaría inadecuada. Dicha respuesta sería el conocido como caso del oscilador armónico. Tal caso del oscilador armónico no dejaba de ser una respuesta insatisfactoria porque este remitía a un caso muy particular, algo que, precisamente, probó su oponente, la mecánica matricial (Rioja, 1995: 255).

La segunda tiene que ver con una incongruencia relativa a la idea de continuidad. Cuando se está realizando un proceso de medida ψ cambia discontinuamente a una nueva configuración, que es aquello que se conoce como reducción del paquete de ondas. Esto no puede ser explicado si, como suponía Schrödinger, entendemos las ondas expandiéndose de forma continua por el espacio (Rioja, 1995: 255).

El tercer problema reside en que las implicaciones de la función ψ son supuestas dentro de un espacio abstracto y no de uno real (Rioja, 1995: 255).

[167] Paréntesis añadido por mí.

[168] Para la exposición de estas cuatro vertientes me he servido de la explicación ofrecida por Ana Rioja (1995).

De hecho, dicho espacio es un espacio con tres dimensiones y no cuatro (como era aceptado desde los estudios de Einstein, quien introdujo el espacio de Minkowsky). Esta característica abre la posibilidad de acusar a la función de ondas de sus resultados son más producto de una suposición que de una descripción fiel de la naturaleza.

Y en último lugar tenemos que ψ es una función compleja, es decir, que se expresa mediante números complejos y no a través de números reales (Rioja, 1995: 256). Esto, tal y como pudimos ver en el anterior capítulo, limita la capacidad de explicación de la función de ondas.

A pesar de las limitaciones que posee la función de ondas de Schrödinger, más adelante veremos cómo Penrose encuentra en ella un argumento convincente para volver hacia una perspectiva determinista en la reforma que plantea nuestro autor. Pero, de momento, sigamos con el asunto que nos ocupa en este apartado.

2.3.3. Born y la función de onda como probable

Max Born reconoció las ventajas que tenía el modelo de Schrödinger con respecto a la mecánica matricial, pero también fue capaz de exponer las limitaciones de la función de onda, teniendo en cuenta las limitaciones anteriormente citadas y añadiendo una más. Born sostiene que la función de onda puede tener dificultades a la hora de ser explicada y entendida en la realidad, porque llega a entrar en conflicto con el conocimiento que podemos tener de ella. Jammer lo explica del siguiente modo:

> […] la dependencia de la función ψ de la elección de las variables utilizadas para su formación o, en resumen, su dependencia de representación tiene que esperarse ya que el conocimiento sobre la posición obtenida de la «representación de posición» es naturalmente diferente del conocimiento sobre el impulso obtenido de la «representación de momento» (la función ψ en el espacio de momento) (Jammer, 1974: 43).

El científico alemán realiza un estudio muy técnico[169] acerca de la función de onda de Schrödinger. Por nuestra parte, no será este el aspecto en el que nos centremos, sino, como hemos podido entrever ya, en las conclusiones más filosóficas de Born acerca de la función de onda.

Born expone que la función de onda de Schrödinger (aplicada en concreto al caso de un electrón disperso) da una respuesta más amplia y general siempre y cuando esta sea interpretada en términos de probabilidad. Se da, por tanto, la situación en la que una suposición menos absoluta lograba tener mayor poder de explicación. Es por ello por lo que Born plantea que la mecánica ondulatoria de Schrödinger no ofrece una respuesta absoluta sino

[169] De los autores tratados, Born es el más técnico de todos a la hora de explicar sus argumentos. Con técnico me refiero a que la mayor parte de sus explicaciones están expresadas en lenguaje formal matemático (algo natural teniendo en cuenta su campo de investigación). Este hecho ha provocado que su exposición sea más reducida. En ninguno de los casos tiene que ver por considerarlo menos relevante.

probabilística[170]. Esta consideración es adoptada por Born más como un apoyo a lo que decía la función de onda que como un argumento en contra de la misma:

> […] Se deduce, dijo Born, que la mecánica ondulatoria no da una respuesta a la pregunta: ¿Cuál es, precisamente, el estado después de la colisión? En su lugar, solo responde a la pregunta: ¿Cuál es la probabilidad de un estado definido después de la colisión? En el primero de sus documentos más detallados sobre colisiones, describió la situación de la siguiente manera: «El movimiento de las partículas se ajusta a las leyes de probabilidad, pero la probabilidad misma se propaga de acuerdo con la ley de causalidad»[171] (Jammer, 1974: 40).

El hecho de que Born acepte buena parte las implicaciones de la función de onda, sobre todo aquellas que tienen en cuenta el rasgo que hemos visto, ello no impide que la visión de este y de Schrödinger estén alejadas. Born reconoce y acepta el formalismo de la función de ondas, pero el determinismo al que parece intentar volver no le convence de ningún modo.

Born, como vimos más arriba, es indeterminista. Ello provoca un conflicto evidente y nos lleva a una cuestión de manera obligatoria: ¿con qué perspectiva, entre la de Born o la de Schrödinger, nos quedamos según aquello que demuestran? Según lo visto, puede dar la sensación de que la de Born, al poder ofrecer una respuesta más general, es la triunfante en el pleito. Sin embargo, es la de Schrödinger la que consigue ser más precisa y abarcadora. Un ejemplo que hace manifiesta la superioridad de la función de onda schrödingeriana es el experimento de la doble rendija, en el cual la función de onda como probable de Born no resulta satisfactoria[172].

---

[170] Esto no significa que Born defienda que Schrödinger plantea la función de ondas en términos absolutos (porque, de hecho, no lo hace). Sin embargo, como hemos visto más arriba, cuando la función de onda apareció en escena habían quienes veían en ella el resquicio para poder seguir dando respuestas absolutas al modo al que aspiraba la física clásica. Es cierto que Schrödinger llegó a entender las implicaciones de su planteamiento como inclinadas hacia una vuelta a la perspectiva determinista, empero, esta convicción nunca fue plena ni duradera.

[171] Cuando Born habla de causalidad lo hace teniendo en cuenta diferentes conceptos, ya que considera que todos ellos están íntimamente relacionados, los conceptos son los siguientes:
El *determinismo* postula que los eventos en diferentes momentos están conectados por leyes de tal manera que se pueden hacer predicciones de situaciones desconocidas (pasadas o futuras).
Por esta formulación se excluye la predestinación religiosa, ya que supone que el libro del destino solo está abierto a Dios.
La *casualidad* postula que existen leyes por las cuales la ocurrencia de una entidad *B* de cierta clase depende de la ocurrencia de una entidad *A* de otra clase, donde la palabra «entidad» significa cualquier objeto físico, fenómeno, situación o evento. *A* se llama la causa, *B* el efecto.
Si la casualidad se refiere a eventos únicos, se deben considerar los siguientes atributos de la casualidad:
El *antecedente* postula que la causa debe ser anterior, o al menos simultánea con el efecto.
La *contigüidad* postula que causa y efecto deben estar en contacto espacial o conectado por una cadena de cosas intermedias en contacto (Born, 1949: 9).

[172] Jammer lo explica del siguiente modo: […] la interpretación original de Born implicaba que el ennegrecimiento en la pantalla de grabación detrás de la doble rendija, con ambas rendijas abiertas, debería ser la superposición de los dos ennegrecimientos individuales obtenidos con un solo deslizamiento abierto por turno. El hecho muy experimental de que hay regiones en el patrón de difracción que no están ennegrecidas en absoluto con ambas ranuras abiertas, mientras que las mismas regiones exhiben un fuerte ennegrecimiento si solo una ranura está abierta, refuta la versión original de Born de su interpretación probabilística.

Efectivamente las esperanzas que los deterministas tenían depositadas en la función de ondas para rescatar la postura filosófica que defendían no eran infundadas. No obstante, la postura indeterminista volvería a contar con un argumento que tendrá una fuerza innegable: las relaciones de incertidumbre de Heisenberg.

2.3.4. Las relaciones de incertidumbre de Heisenberg

Sin duda la función de ondas de Schrödinger constituyó un bastión importante para el bando determinista, pero tal defensa se vería relativamente pronto en aprietos, teniendo gran parte de culpa el quehacer de Werner Heisenberg. El joven físico alemán había presenciado en Múnich una conferencia de Schrödinger en la que este último expuso sus estudios de mecánica ondulatoria. En dicha conferencia Heisenberg criticó a Schrödinger diciendo que «la ley básica de la radiación de Planck no se podía entender dentro del marco de la interpretación de Schrödinger». Por aquel entonces (1926) Heisenberg ya era miembro del grupo de investigación de Bohr en Copenhague. Bien se sospecha que el interés y la crítica de Heisenberg movieron a Bohr a invitar a Schrödinger a Copenhague[173] (Jammer, 1974: 56).

Schrödinger aceptó la invitación y allí debatió acerca de la función de onda, su inclinación determinista e intentó aclarar la relación entre la mecánica cuántica (tal y como la entendían Heisenberg y Bohr) y los datos de la experiencia. Se sabe que los debates entre Heisenberg y Bohr contra Schrödinger fueron largos e intensos, hasta el punto de que este último declaró que si bien no se fue convencido de Copenhague sí que se fue vencido en energías.

Lejos de quedar satisfecho con lo conseguido en la reunión de Schrödinger en Copenhague, Heisenberg seguía buscando una respuesta que fuera capaz de tumbar al determinismo. Y si bien no puede considerarse que consiguió tumbar de una manera absoluta con el determinismo, sí que se reconoce que dio con una de las claves que cambiaría ya no sólo su carrera sino también el rumbo de la física moderna: las relaciones de incertidumbre.

Según se desarrollaban los resultados a través de la perspectiva de la función de ondas, Heisenberg seguía planteándose dos cuestiones que, considera, son esenciales, porque con estas se aborda la problemática conceptual de un modo más directo. Tales cuestiones son las siguientes: i) ¿El formalismo permite el hecho de que la posición de una partícula y su velocidad son determinables en un momento dado solo con un grado limitado de precisión?; ii) ¿Esta imprecisión, si la teoría lo admite, sería compatible con la exactitud óptima

---

Dado que este experimento de doble rendija se puede llevar a cabo a intensidades de radiación tan reducidas que solo una partícula (electrón, fotón, etc.) pasa al aparato a la vez, queda claro, en el análisis matemático, que la onda ψ asociada con cada la partícula interfiere consigo misma y la interferencia matemática se manifiesta por la distribución física de las partículas en la pantalla. Por lo tanto, la función ψ debe ser algo físicamente real y no simplemente una representación de nuestro conocimiento, si se refiere a partículas en el sentido clásico (Jammer, 1974: 44).

[173] De hecho, Heisenberg en su autobiografía lo cuenta de tal modo.

que se puede obtener en mediciones experimentales? (Jammer, 1974: 61). Aquello que quiere seguir planteando Heisenberg es que si el conocimiento que nos ofrece la física es, de algún modo, impreciso (*Ungenau*), esto se debe a que la naturaleza en última instancia tiene que *ser* de tal modo. El caso es que Heisenberg logra explicar[174] esta imprecisión (indeterminación, incertidumbre) alegando que es imposible obtener un resultado preciso del momento de una partícula a la par que queremos obtener su velocidad. No sólo se trata de una alegación interesada (en términos filosóficos) por parte de Heisenberg, sino que esto pudo demostrarse, provocando que allí donde guardaba silencio la interpretación de Schrödinger, las relaciones de incertidumbre de Heisenberg tuvieron algo más que decir:

> […] Se demostró que no es posible determinar a la vez la posición y la velocidad de una partícula atómica con un grado de precisión arbitrariamente fijado. Puede señalarse muy precisamente la posición, pero entonces la influencia del instrumento de observación imposibilita hasta cierto grado el conocimiento de la velocidad; e inversamente se desvanece el conocimiento de la posición al medir precisamente la velocidad; en forma tal, que la constante de Planck constituye un coto inferior del producto de ambas imprecisiones (Heisenberg, 1976: 33).

Por tanto, este planteamiento hacía manifiesto que el indeterminismo era propio de la naturaleza, y no el resultado de una mera falta de conocimiento. Es más, aquello que nos dice la naturaleza es precisamente que cuanto más conocimiento tenemos de ella más indeterminada se nos muestra. Al fin la respuesta que buscaban los indeterministas estaba encima de la mesa. Frente a aquellos que *suponían* una naturaleza determinista, Heisenberg defendía estar *mostrando* lo que la naturaleza nos permite conocer:

> […] es posible preguntarse si todavía queda oculto detrás del universo estadístico de la percepción un universo «verdadero» en el que la ley de causalidad sería válida. Pero tal especulación nos parece carecer de valor y sin sentido, ya que la física debe limitarse a la descripción de las relaciones entre las percepciones (Heisenberg cit. por Jammer, 1974: 76).

A pesar de que las relaciones de incertidumbre suponían un argumento muy favorable para la postura indeterminista, ello no significó que se aceptara de manera automática como tampoco impidió que surgiera un debate dentro de este bando. Dicho debate estaría protagonizado por el mismo Heisenberg y por Bohr. Si bien, por un lado, el físico danés compartía la totalidad de la conclusión a la que había llegado Heisenberg a través de las relaciones de incertidumbre, por el otro disentía de la forma en la que había llegado a ello.

Para Heisenberg era imprescindible no abandonar el formalismo matemático del que se había servido para construir las relaciones de incertidumbre. Tal formalismo matemático era abstracto, por lo que servía para cualquier tipo de suposición con respecto a si la naturaleza está compuesta de partículas o de ondas. Sería el experimento de la cámara de Wilson[175] aquel que adquiriría una importancia capital a la hora de entender

[174] Para una explicación formal suficientemente detallada del razonamiento de Heisenberg para llegar a las relaciones de incertidumbre véase (Jammer, 1974: 61-71).

[175] También conocida como cámara de niebla, es un experimento propuesto por el físico escocés Charles Thomson Rees Wilson, con el cual se pretende estudiar partículas de radiación ionizante. La ionización se da cuando la carga de la partícula llega a un valor

lo que nos dicen las relaciones de incertidumbre. Sobre todo porque dicho experimento permite que el fenómeno que se observa en él puede ayudar a comprender la relación entre el formalismo y la experiencia. Es decir, que con este experimento el formalismo matemático no correría el peligro de ser considerado como una *simple* abstracción. La importancia del experimento venía de lejos, ya que aun cuando Heisenberg y Bohr acabaron sus discusiones con Schrödinger estos no podían dar con la solución del mismo. El hecho es que Heisenberg pudo resolverlo matemáticamente a través de lo postulado en las relaciones de incertidumbre, es decir, atribuyendo grados de incertidumbre (imprecisión) a la velocidad y a la posición de las partículas de la cámara.

La discrepancia entre Heisenberg y Bohr residía en el formalismo utilizado por el primero. Bohr consideraba que un formalismo *neutro*, como el empleado por Heisenberg (ya que, recordemos, contemplaba la naturaleza compuesta tanto como de partículas como por ondas), no conduce a una descripción fiel de la naturaleza. Según el físico danés, la solución pasaba por presuponer que en la naturaleza se da la dualidad onda-partícula, ya que ello permite entablar la discusión desde los conceptos clásicos y no fuera de estos:

> El punto de partida de Bohr fue la fundamental dualidad onda-partícula que encontró su expresión en la individualidad de los procesos atómicos y que, en consecuencia, llevó a la pregunta sobre los límites dentro de los cuales los objetos físicos de tal naturaleza pueden describirse en términos de conceptos clásicos; la limitación de la mensurabilidad confirma la limitación de la definibilidad pero no la precede lógicamente. La posición de Bohr podría apoyarse en el argumento de que, como se mencionó previamente, cualquier derivación de las relaciones de indeterminación de los experimentos de pensamiento, esto es, la formulación de Heisenberg de las limitaciones de la mensurabilidad, debía basarse en las ecuaciones Einstein-de Broglie, que, a su vez, conectan las características de la descripción de la onda y las partículas, y por lo tanto presuponen implícitamente la dualidad onda-partícula (Jammer, 1974: 69).

En un principio Heisenberg no entendía por qué las relaciones de incertidumbre no podían quedarse tal y como él las había presentado. Este hecho casi llegó a ser un problema en su relación personal con Bohr, pero finalmente, gracias al papel intermediario de Pauli, el tema fue resuelto. De hecho, Heisenberg posteriormente agradecería la labor de Bohr de hacerle corregir su trabajo, ya que este le aseguraba que ello serviría para la introducción de nuevos conceptos en la física que se estaba asentando[176]. Heisenberg adaptó conceptualmente las relaciones de incertidumbre a la ciencia clásica y Bohr lo pudo apoyar en su concepto de complementariedad, término que veremos con algún detalle en el apartado del debate que sostuvo con Einstein.

concreto y se mezcla con el vapor contenido en la cámara. Es en esta mezcla o interacción cuando se produce la ionización, siendo el vapor aquello que se ioniza. Lo importante de este experimento es poder determinar las trayectorias de las partículas.

[176] Estoy muy agradecido al profesor Bohr por haber tenido la oportunidad de ver y discutir sus nuevas investigaciones que pronto se publicarán como un ensayo sobre la estructura conceptual de la teoría cuántica (Heisenberg cit. por Jammer, 1974: 69).

2.3.5. Doble solución de Louis de Broglie

Pero antes de pasar al intenso debate entre Einstein y Bohr haré una breve mención a la contribución de Louis de Broglie, ya que ha sido citada anteriormente y es de rigor que al menos tengamos en cuenta de qué trata en concreto.

La aportación de este físico francés se encuentra incluida ya en su tesis doctoral, muestra de su enorme capacidad para la innovación en el campo.

En realidad, el tema que pone encima de la mesa pertenece a un debate ya planteado desde la Antigüedad: la naturaleza de la luz. Por otro lado, la perspectiva de Louis de Broglie acabó constituyendo una postura muy útil y sutil, y acabó diferenciándose del resto. Pero volviendo al debate referido, este no es otro que aquel en el que se discute si la luz está compuesta de ondas (al igual que el sonido, por ejemplo) o, en cambio, lo hace de partículas. El debate planteado de tal forma era legítimo, ya que en un principio la suposición de la luz como compuesta de una u otras debía darse de modo exclusivo. Pero ello cambiaría con la propuesta de Louis de Broglie.

Si bien de Broglie toma la idea principal de las investigaciones sobre la luz que Einstein había realizado en 1905[177], no menos cierto es que su aportación personal constituirá un antes y un después en la concepción de la luz en la física. Veamos, pues, su propuesta.

La tesis de Louis de Broglie dice que la naturaleza de la luz es doble, es decir, de onda y partícula. El modo en el que de Broglie reconcilia la naturaleza opuesta de estas dos también está basado en estudios de Einstein. El estudio einsteniano concreto es en el que se incluye la célebre ecuación $E = mc^2$, la cual nos dice, como sabemos, que la masa y la energía son dos dimensiones de una misma cosa. Esto le servía porque si la onda solía ser considerada, de algún modo, opuesta a las partículas es porque mientras que la primera carece de masa la segunda sí que la posee. Por lo que entender que la masa puede expresarse de otro modo (es decir, como la energía) hace que la suposición de una no implique el descarte de la otra. Esto, sin embargo, no es todo lo esclarecedor que requiere una suposición de este calibre. De un modo más concreto, aunque no menos complejo, aquello que defiende de Broglie es que un electrón (es decir, una partícula) se caracteriza porque en sí mismo es una onda y no porque se comporte como tal. Concebir de manera clara esta idea no es una tarea sencilla, pero sí que puede ser bosquejada mentalmente. Aunque ello no logre explicarla puede resultar útil para que se entienda de algún modo:

> […] De ser cierta la hipótesis que plantea de Broglie, si yo lanzo un solo electrón hacia una pared, este se desplazaría como onda, expandiéndose hacia todas direcciones. Sin embargo, al llegar a la pared el electrón incidiría sólo en un punto. Al fin y al cabo, lancé sólo un electrón. Pero el electrón puede incidir en cualquier lugar que esté permitido por su onda electrónica, ¡aunque haya obstáculos de por medio! Si la onda puede llegar entonces el electrón también puede. […] Es como si

[177] En dicho experimento, Einstein concluyó y demostró que la luz en ocasiones se presenta como un conjunto de partículas, en oposición a la visión generalizada de que la luz está compuesta exclusiva de ondas. Es de rigor recordar que los estudios sobre la naturaleza y el comportamiento de la luz de Einstein tuvieron tal importancia que estos fueron aquellos que le permitieron obtener el premio Nobel de Física en 1921.

> las olas del mar llegaran a la orilla de la playa y rompieran sólo en un punto (Fuentes Fernández, 2015: 17).

Todo esto, sin duda, tiene tintes muy filosóficos, incluso metafísicos, pero, ¿qué nos dicen los resultados físicos? Obviamente de Broglie expuso sus argumentos científicos y el resultado de sus cálculos le daban la razón. Además, experimentos posteriores también demostraron dichos resultados, hechos que le valieron el Nobel de Física sólo cinco años después de haber acabado su tesis doctoral, en 1929.

Habiendo visto cómo se consolidaron los dos bandos con sus argumentos actualizados con respecto al desarrollo de la teoría cuántica, veamos ahora el culmen de las discusiones sostenidas por ambos: el debate entre Einstein y Bohr.

2.4. Debate Einstein-Bohr

Hablar de debate entre Einstein y Bohr de forma singular es, históricamente hablando, incorrecto, ya que ambos pensadores intercambiaron ideas en numerosas ocasiones, tanto en persona como por correspondencia e incluso respondiéndose a través de artículos. Sin embargo, el tema principal de dichas discusiones sí que fue invariable: determinismo vs indeterminismo.

Las dos grandes ocasiones en las que Einstein y Bohr entablaron discusiones directas sobre el determinismo y el indeterminismo fueron en los congresos de Solvay, concretamente en el quinto (1927) y el sexto (1930).

Si bien estas fueron las más significativas, no menos cierto es que no fueron las únicas. La rivalidad intelectual entre Einstein y Bohr comenzó en 1920. Ese año Bohr visita Berlín con motivo de reunirse con Max Planck, James Franck y Albert Einstein y sería con este último con quien intimaría más en dicha estancia. Pronto quedó claro que ambos tenían convicciones filosóficas alejadas, pero ello no sería obstáculo para que ambos profesasen un respeto y una admiración mutua. En esta ocasión la discusión se centraba en la defensa de Einstein acerca de que la naturaleza de la luz resulta de una combinación entre partículas y ondas, mientras que Bohr optaba por la opción «clásica» de la luz como onda. A pesar de que la concepción de Bohr era «clásica», el modo de defender su postura incluía la introducción del salto cuántico, mientras que Einstein se negaba a ello. En esta discusión en concreto, más tarde se demostraría con la aportación de Louis de Broglie, precisamente en el quinto congreso de Solvay, que era Einstein quien tenía razón.

Otro punto en el que ambos pensadores chocaban se dio en 1924, con motivo del descubrimiento del efecto Compton. Si en el debate de 1920 Einstein acabó teniendo razón al otorgar una doble naturaleza a la luz, en 1924, Bohr sería quien obtendría su parte de razón, ya que el efecto Compton vino a confirmar la naturaleza cuántica de la luz.

Estas discusiones no tuvieron la relevancia que sí tendrían las mantenidas en los congresos quinto y sexto de Solvay, pero, sin duda, fueron un preludio significativo de los debates que allí se dieron.

Antes de pasar a las discusiones entre Einstein y Bohr de manera particular es conveniente aclarar que los congresos de Solvay supusieron la mayor

concentración personalidades influyentes dentro del mundo científico que hasta el momento se había producido. El motivo de estas reuniones era la discusión acerca de la teoría cuántica y sobre cómo esta se asentaba o, por el contrario, mostraba sus limitaciones. Es decir, tales congresos eran el escenario en el que se daban de un modo directo las discusiones entre los dos bandos que hemos visto. Aparte de los anteriormente citados, nombres como Paul Langevin, Maurice de Broglie (hermano de Louis), Marie Curie, Henri Poincaré, Max von Laue, William Lawrence Bragg, Paul Dirac, Peter Debye, Arthur Holly Compton, Owen William Richardson, Enrico Fermi o Paul Ehrenfest (por citar sólo a los más ilustres) hicieron acto de presencia en algunos de estos congresos celebrados entre los años 1911 y 1933, año del último congreso al que acudiría Einstein[178].

El quinto congreso, que tenía como asunto a tratar los electrones y los fotones, comenzó con la intervención de Louis de Broglie, quien expondría el problema de la naturaleza de la luz introduciendo su teoría de la doble solución. Pero en su presentación no sólo se limitaría a explicar su teoría, sino también cómo a través de ella se puede entender de un modo más claro la función de onda ψ. Para de Broglie la función de onda schrödingeriana tiene una doble tarea: i) es una onda de probabilidad, pero ii) también es una onda piloto, que a través de la fórmula de guía determina la trayectoria de la partícula en el espacio (Jammer, 1974: 110). Esta primera exposición constituía una defensa del determinismo y según la opinión general de quienes participaban en el congreso no provocó demasiados conflictos. Sin embargo, Wolfgang Pauli sí que tuvo algo que decir al respecto. Pauli aseveraba que lo defendido por de Broglie es semejante a la teoría de las colisiones elásticas de Born, la cual es insostenible una vez son consideradas las colisiones inelásticas[179]. De Broglie respondería a Pauli, pero su contrarréplica no convenció ni a los presentes ni a él mismo en el fondo (Jammer, 1974: 114).

Las dos siguientes exposiciones corrieron a cargo de Born y Heisenberg, por un lado, y de Schrödinger por el otro. La primera se centró en la explicación de la mecánica matricial, mientras que la segunda fue la mecánica ondulatoria. Ya vimos más arriba que la mecánica matricial y la función de onda eran equivalentes, pero la mayor simplicidad del trabajo de Schrödinger hizo que la balanza se inclinara a su favor. Pero este asunto no sería el único que se tratase a partir de tales exposiciones. También fue objeto de debate las relaciones de incertidumbre de Heisenberg, ya que habían salido a la luz hacía relativamente poco tiempo. Pero fue precisamente en este congreso donde se dieron las discusiones, las cuales fueron muy intensas y en ellas intervinieron un gran número de quienes estaban presentes. Entre las participaciones más destacadas está la llevada a cabo Einstein. Pero antes de Einstein intervino Bohr.

La exposición de Bohr estuvo motivada por una cuestión que Lorentz lanzaba a aquellos que defendían la postura indeterminista, aunque de un modo más concreto al físico danés. Tal cuestión fue planteada con las siguientes preguntas: « ¿Podría una mente más profunda no ser consciente de

[178] Más tarde se seguirían celebrando, hasta un undécimo en 1958, pero, como hemos visto, nuestro interés está centrado en el quinto y el sexto, que son aquellos en los que las posturas de Einstein y Bohr fueron expuestas y sometidas a debate.

[179] Para una explicación técnica del argumento de Pauli, véase (Jammer, 1974: 111-113).

los movimientos de estos electrones? ¿No podríamos mantener el determinismo siendo objeto de una creencia? ¿Deberíamos exigir necesariamente el indeterminismo en principio?» (Jammer, 1974: 114). Como se puede ver en estas preguntas, Lorentz pretende seguir apelando a la ignorancia que podemos tener con respecto al movimiento de los electrones en lugar de otorgar a la naturaleza una parte indeterminista. La respuesta de Bohr fue básicamente la misma que ofreció en su célebre lectura en Como un mes antes, ya que en ella exponía por primera un concepto que él consideraba fundamental para entender el indeterminismo de la naturaleza: la complementariedad. Tal y como era habitual en Bohr, si bien defendía que dicho concepto es clave para entender su postura, ello tampoco implica que la definición que puede ofrecer sobre ella sea cerrada o completa[180]. De hecho, cuando presenta tal idea lo hace con la intención de que se perfile entre los presentes, ya no sólo en Como sino también en Solvay. Como vimos más arriba, la idea de complementariedad surge en el contexto de las relaciones de incertidumbre de Heisenberg. Es por ello precisamente por lo que para la idea de complementariedad son indispensables conceptos tales como *medida*, *observador* u *observado* (Saunders, 2005: 423). La principal idea de la complementariedad de Bohr va en contra de la concepción de Louis de Broglie, de manera tal que la primera noción entiende que la naturaleza de la luz puede entenderse de un modo (partículas) u otro (ondas) y nunca a la vez. El modo en el que resuelve Bohr que la luz se comporte a veces como formada por partículas y otras por ondas es argumentando que aunque ambas sean incompatibles entre sí ello no impide que sean complementarias. En realidad esta postura de Bohr, aunque científica, es marcadamente filosófica, siendo en ocasiones considerado como antirrealista e incluso perteneciente a una forma de subjetivismo propio de un neokantismo singular (Saunders, 2005: 426-427). Que podamos dar dos tipos de descripciones que se excluyen pero que acaban aportando informaciones complementarias es resultado de la incapacidad que tenemos los humanos de poder hacer una descripción exacta de los procesos naturales, debido a la naturaleza cuántica de los mismos. A pesar de que la complementariedad (y con ella las relaciones de incertidumbre de Heisenberg) decrete(n) la imposibilidad de la descripción de la naturaleza ello se basa en la idea de que la teoría cuántica es una teoría completa. Es decir, que la completitud de la teoría es la que permite definir la inconmensurabilidad de la naturaleza.

Una vez finalizada la exposición de Bohr siguieron varias intervenciones a propósito de lo explicado por el pensador danés, como las de Brillouin y Born. Pero la que tuvo más peso fue la de Einstein, ya que Bohr se había dirigido a él de manera directa y porque su respuesta no se hizo esperar.

Einstein respondió reafirmándose en su escepticismo con respecto a la naturaleza como cuántica y de que la teoría de los cuantos sea una teoría

[180] Esta era una consideración que, lejos de ser personal (ya que no he tenido la oportunidad de estudiar a fondo el pensamiento de Bohr) procedía obviamente de sus detractores, pero también de quienes en momentos determinados llegaron a ser sus seguidores. Estas palabras de Paul Dirac hacen manifiesto tal sentimiento: […] los argumentos [de Bohr] eran principalmente de naturaleza cualitativa y no era capaz de señalar realmente los hechos subyacentes. Lo que yo quería eran afirmaciones que pudieran expresarse en términos de ecuaciones y el trabajo de Bohr muy raramente ofrecía tales afirmaciones (Dirac cit. por Lindley, 2008: 147).

completa. Como argumento utiliza un experimento, con el cual pretende explicar los dos diferentes puntos de vista de la teoría de los cuantos.

El experimento que propuso Einstein consiste en concebir el comportamiento de una partícula, colocando una pantalla fotográfica con forma semiesférica detrás de un diafragma con hendidura (la cual denota 0). El experimento, obviamente, se calcularía a través de operaciones de probabilidad. La probabilidad que se intentaría obtener sería mediante la intensidad de las ondas difractadas a partir del punto de la hendidura. Pasemos a ver los dos puntos de vista de la teoría cuántica según nos permite ver este experimento.

El primer punto de vista es aquel en el que las ondas propias de los estudios de Louis de Broglie, Schrödinger e incluso del mismo Einstein, no representan una partícula individual, sino un conjunto de partículas distribuidas en el espacio. Esta condición nos lleva a pensar que no se está atendiendo a un proceso individual, sino de un conjunto, por lo que el resultado que obtengamos será probable (Jammer, 1974: 115).

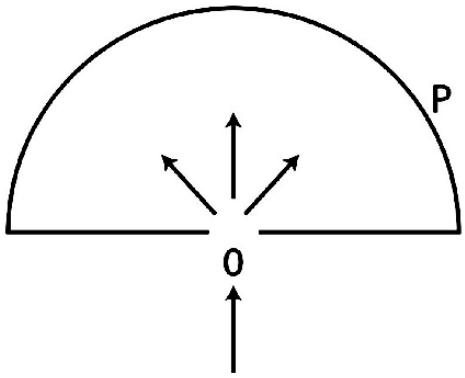

Figura 7. Representación del experimento de Einstein

El segundo punto de vista parte de la consideración de que la mecánica cuántica es una teoría completa de los procesos individuales, algo con lo que Einstein, tal y como vimos más arriba, no está de acuerdo. Atendiendo al experimento que propone Einstein, podemos ver que cada partícula (recordemos que el experimento nos permite estudiar, porque sigue el modelo de Broglie-Schrödinger, un conjunto de partículas y no una partícula individual) que se mueve hacia la pantalla se describe como un paquete de ondas. Este paquete de ondas después de la desviación (en la Figura 7 esta desviación se representa con las flechas que surgen del punto 0) llega a un cierto punto (en la Figura 7, P, como ejemplo) en la pantalla. Los resultados que prueban este hecho también se expresan en términos de probabilidad. En ese marco que permite la probabilidad Einstein supone, erróneamente[181], que cuando no se realiza una localización de las partículas, estas tienen la posibilidad potencial de estar presente en toda el área de la pantalla fotográfica. En la suposición de Einstein también cabe que una vez se realiza la localización, deber ser tomada en consideración una acción a distancia peculiar (*peculiar action-at-a-distance*) en dos lugares de la pantalla.

[181] Esta no es una consideración propia, sino que está tomada de Jammer (1974: 116), quien se basa en resultados físicos posteriores que le dan la razón y se la quitan a Einstein.

A pesar de los esfuerzos de Einstein, sus argumentos no llegaron a convencer a los presentes, incluido, como vimos más arriba, Louis de Broglie, quien había sido defensor de la perspectiva determinista. El indeterminismo de Bohr y su defensa de la mecánica cuántica como una teoría completa habían hecho posible la victoria del pensador danés frente al empeño de Einstein por rescatar la perspectiva determinista.

Este, sin embargo, no sería el final de los debates, ya que el sexto congreso de Solvay, celebrado tres años después, volvería a ser el escenario de las discusiones entre Bohr y Einstein.

En esta ocasión, Einstein fue el primero en lanzar su argumento. Tan bien armado estaba dicho argumento que en realidad Einstein no esperaba que Bohr pudiese ofrecer un contraargumento, algo que finalmente acabó por suceder. Veamos cómo se dieron los sucesos.

Comenzando con el envite de Einstein, en esta ocasión este presentó otro supuesto con una diferencia sutil pero que cambiaría sustancialmente el resultado, al menos en principio. La propuesta de Einstein consistía ahora en un diafragma estacionario con una hendidura y de otro diafragma con otra hendidura, poniéndose en movimiento este último mediante un mecanismo equipado con un reloj. Dicho movimiento del segundo diafragma provoca que parte de la luz se corte a través de las dos rendijas durante el intervalo de tiempo determinado que dura el movimiento del diafragma. Esto tiene como consecuencia que la precisión que podamos obtener de la luz sea arbitraria, al contrario de la defensa de Bohr de la objetividad de la teoría cuántica. No obstante, esta objeción a la perspectiva de Bohr no es lo suficientemente importante, ya que el resultado obtenido en el ejemplo de Einstein es –digámoslo así- a posteriori y en principio no serviría para poner en entredicho a la teoría cuántica en su faceta predictiva. Einstein era consciente de esto y por ello su ejemplo contenía matices que merecen la pena ser revisados.

El caso es que Einstein logra criticar con éxito[182] la faceta predictiva de la teoría cuántica, por lo que su postura cobraba una fuerza que había perdido hasta entonces. Por primera vez un argumento de Einstein ponía en verdaderos apuros a la perspectiva de Bohr y Heisenberg.

[182] Para los detalles del argumento de Einstein véase Jammer (1974: 132-134).

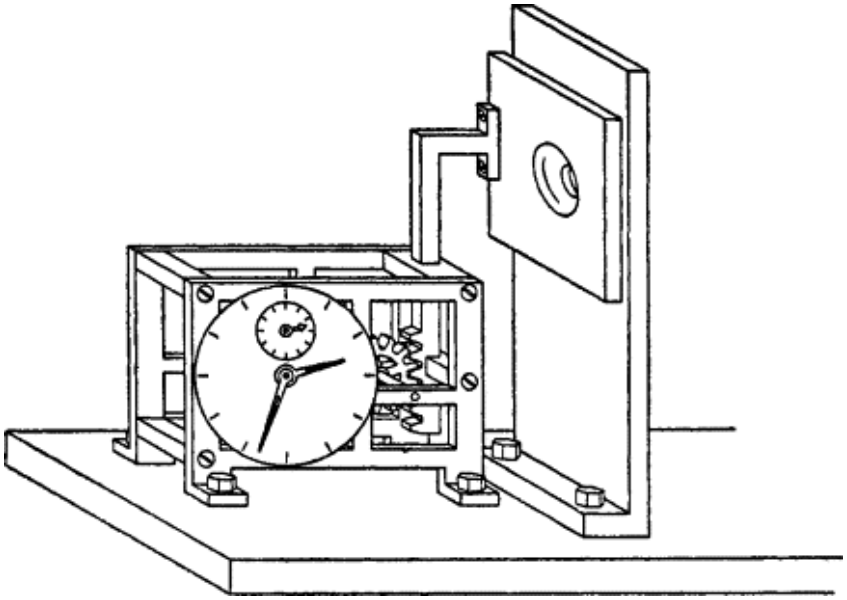

Figura 8. Representación del ejemplo utilizado por Einstein

Ello, sin embargo, no impidió que Bohr elaborase un contraargumento que, además, le diese la razón. Una vez Einstein presentó su ejemplo Bohr dispuso de una noche para intentar replicar. Tal era la fuerza del argumento de Einstein que el pensador danés tuvo que pasar despierto toda esa noche fabricando el argumento que presentaría al día siguiente.

Lo llamativo del caso es que Bohr acabaría contraargumentando gracias a la utilización de conceptos relativistas, algo que, curiosamente, no había hecho hasta entonces (Jammer, 1974: 134). Einstein volvería a ver que sus propias ideas se volverían en su contra también en este asunto. Bohr se había percatado de que Einstein también había ignorado la utilización de los conceptos de su propia teoría de la relatividad. El pensador danés se lo hizo ver suponiendo que el aparato propuesto por Einstein estuviese suspendido en una especie de balanza.

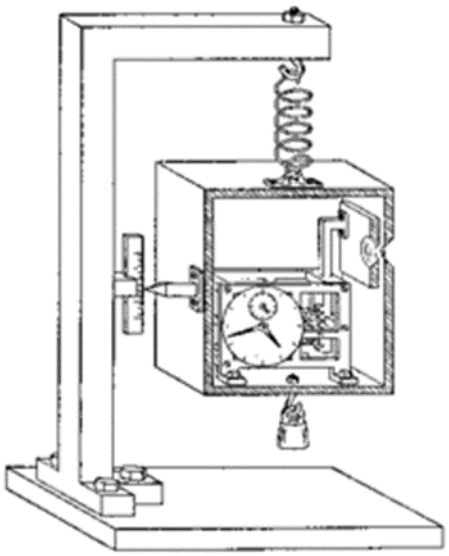

Figura 9. Representación del ejemplo utilizado por Bohr

¿En qué medida este cambio, en apariencia insignificante, da la razón a Bohr? Según la relatividad general, el proceso de pesaje (al que es sometido el aparato al situarlo en esa especie de balanza) se vería afectado con la velocidad. La inconsistencia de la teoría cuántica que Einstein pretendía poner encima de la mesa no se llegó a concretizar con la respuesta de Bohr. Precisamente la derrota en este aspecto provocó que Einstein cambiase su actitud con respecto a la inconsistencia de lo que defendían Bohr y Heisenberg, centrando su posteriores críticas en la incompletitud de la teoría cuántica.

Definitivamente este último argumento fue el que puso fin a las discusiones en Solvay entre Einstein y Bohr en torno al asunto del determinismo contra el indeterminismo. No obstante, el resultado victorioso por parte de Bohr no supondría el punto final a las discusiones entre ambos pensadores. De hecho, cinco años después (1935), Einstein presentó un experimento mental, junto a Nathan Rosen y Boris Podolsky, que venía a poner en entredicho de nuevo al indeterminismo. Este es el célebre experimento conocido como EPR, acorde a las siglas de los apellidos de los tres científicos que lo firmaron[183] y que fue publicado en un artículo bajo el sugerente título "¿Puede la descripción de la

[183] El caso de Einstein es peculiar, porque si bien suscribe aquello que se dice en el artículo en el que se presenta el experimento EPR, ello no significa que entienda del mismo modo la noción de realidad que se infiere en el artículo, la cual se corresponde en mayor medida al punto de vista de Rosen y Podolsky. El modo en el que entienden la realidad estos, como veremos a continuación, conlleva a aceptar una perspectiva que casa perfectamente con un realismo al uso. Esto, sin embargo resulta un problema si tenemos en cuenta que Einstein difícilmente puede ser considerado realista. Si bien no es claramente contrario a esta corriente, no menos cierto es que su postura dista de ser manifiestamente realista. El posible realismo einsteniano ha sido objeto de debate y en la mayoría de los casos se le considera como tal, pero no en un sentido estricto. No obstante, hay quienes niegan toda relación de Einstein con el realismo. Un estudio que apunta en esta dirección es de Don Howard (1993), cuyo artículo se titula provocativamente "Was Einstein Really a Realist?

realidad mecánica-cuántica ser considerada completa?". Veamos a grandes rasgos qué dice dicho artículo, el cual se puede dividir en cuatro partes[184].

La primera parte establece unas condiciones fundamentales que las que las teorías físicas deben cumplir. Una de ellas nos dice que «cada elemento de la realidad física debe tener una contraparte en la teoría física» (Jammer, 1974: 181). A esta condición la denominan «de integridad». Otra concierne a la realidad física, a la cual dan el nombre de «condición de realidad», y no consiste en otra cosa que poder predecir el valor de una cantidad física, siempre y cuando el sistema en el que se encuentre no sea alterado. Si bien intentan dejar de lado las expresiones filosóficas, lo cierto es que con estas condiciones se está tomando parte en una postura realista en su sentido estricto. Los resultados de una teoría que cumpla estas condiciones dependerán exclusivamente de las mediciones y los experimentos que se lleven a cabo en el mundo físico.

La segunda parte corresponde a una descripción de la mecánica cuántica. Esta es definida en términos de funciones de onda. Al mismo tiempo, se entiende que de dos cantidades físicas, estas están representadas por operadores que no se conectan, es decir, que el conocimiento preciso de una de ellas impide un conocimiento similar de la otra, tal y como nos dicen las relaciones de incertidumbre de Heisenberg. De esto se puede deducir que la descripción que podemos obtener de la teoría cuántica no es completa, como creen Einstein, Rosen y Podolsky. Aunque también existe la posibilidad de entender que entre los operadores que no se conectan (como dice la teoría cuántica) la realidad no puede ser simultánea, que es la postura oficial de Bohr y Heisenberg.

Como podemos ver, hasta entonces el artículo está dedicado a exponer los puntos esenciales del punto de vista al que pretende criticar, pero esto cambiará de tono a partir de la siguiente parte, donde se expone el ejemplo que intenta esclarecer la pregunta que se hace en su mismo título.

El ejemplo de la tercera parte consiste en un sistema concreto, que se compone de dos partículas (I y II). Este ejemplo supone que al medir el momento de I permite predecir con certeza el momento de II sin alterar a esta última. Sucedería lo mismo con respecto a la posición[185]. Esta condición tiene como consecuencia, siguiendo la «condición de realidad» establecida en la primera parte, que ambas partículas pertenecen a la misma realidad física.

La cuarta parte corresponde a la conclusión del artículo. Dicha conclusión a la que llegan es, a estas alturas ya obvia, que la mecánica cuántica no ofrece una descripción de la realidad completa y su defensa es una respuesta basada en la lógica, teniendo como premisas y argumentos lo visto en las anteriores partes.

La acogida al artículo fue extendidamente positiva, ya que lograba volver a poner en tela de juicio la infalibilidad de la que hasta el momento había gozado la teoría cuántica. Esto, sin embargo, lejos de desanimar a sus contrarios provocó una respuesta casi inmediata, como no podía ser de otro modo, de la mano de Bohr.

---

[184] Esta división también está tomada de Jammer (1974), ya que el artículo original está dividido sólo en dos partes.

[185] En el artículo se explica el modo en el que se daría tales condiciones (Einstein, Podolsky, Rosen, 1935: 779-780).

Antes de publicar el artículo que sería la respuesta oficial al artículo de Einstein, Rosen y Podolsky, Bohr ya había escrito en otro lugar su disconformidad con respecto al concepto de realidad que estos habían usado en su trabajo[186]. Pero esta sería una crítica menos profunda que la llevada a cabo en el artículo que presentó a *Physical Review*, el cual tendría el mismo título que el de Einstein, Rosen y Podolsky. Sin embargo, el contenido de la crítica en dicho artículo gira en torno al tema del concepto de realidad que *se decida* tomar en consideración.

El grueso del argumento de Bohr consiste en que dado un experimento concreto (habla de un diafragma que contiene dos ranuras muy estrechas), los resultados que podemos llegar a obtener dependerán no sólo de aquello que se estudie, sino también del aparato con el que se estudie. Este hecho hace que el cálculo exacto del momento y la posición no sean posibles. Es decir, que las relaciones de incertidumbre y la complementariedad siguen teniendo vigencia. Para Bohr, el error que comete EPR, como hemos visto ya, se encuentra en el concepto de realidad manejado. El pensador danés piensa que esta concepción de realidad más que aclaratoria es difusa. Una afirmación que hace manifiesta dicha ambigüedad es aquella que dice que «al medir el momento de la partícula I se puede predecir *con certeza* el momento de la partícula II sin alterar a esta última». ¿Cómo podemos asegurar, se pregunta Bohr, que no se alteran los momentos? La pregunta es pertinente porque el supuesto de EPR pasa por alto la posible influencia que puede tener el experimentador (ya sea un aparato o un ser humano).

Bohr remarca este asunto y decide hacer una distinción entre la física clásica y la moderna, ya que considera que este es el aspecto que no tienen en cuenta EPR y ello precisamente es lo que les hace errar.

Comenzando por la física clásica, Bohr destaca que para que sus leyes posean un sentido experimental, debemos poder determinar el estado exacto de todas y cada una de las partes relevantes del sistema[187]. Otra de las características de la física clásica, y que está relacionada con lo visto hasta ahora de la crítica de Bohr, es que en esta puede hacer una distinción mediante un análisis conceptual apropiado entre el objeto y el aparato de medición (Jammer, 1974: 196). Es decir, que podemos tomar *partes* de la realidad y tratarlas como un *todo*. Esto, sin embargo, resulta conflictivo para la teoría cuántica, ya que el objeto y el dispositivo de medición forman una unidad que no puede analizarse. Es decir, que en la física clásica la interacción entre el objeto y el dispositivo de medición puede descuidarse o compensarse, cuando en la teoría cuántica estos son «inseparables» (Jammer, 1974: 197). La teoría cuántica, por tanto, no ofrece una descripción del estado del objeto de estudio, sino de la situación experimental en la que está involucrado (Jammer, 1974: 197). De un modo casi paradójico, la física clásica queriendo describir de un modo más realista la naturaleza deja de lado la totalidad que sí tiene en cuenta la teoría cuántica.

Con este artículo, Bohr había conseguido de nuevo que el consenso de la comunidad científica se decantara por su propuesta. Pero esto no sería así para

[186] El artículo era del mismo año (1935), se tituló "Mecánica cuántica y realidad física" y fue publicado en la revista *Nature*.

[187] Es por esta razón por la que la física clásica acabó derivando en un determinismo fuerte, aunque grandes representantes de la física clásica no habían sido deterministas, teniendo como gran ejemplo a Newton.

siempre. Bien se sabe que existen casos de científicos que no cejan su empeño en seguir rescatando al determinismo[188]. Es en esta corriente donde se sitúa Penrose, quien retomará los argumentos de Einstein y Schrödinger para reinterpretarlos, intentando con ello que la física cuántica se reestructure desde un enfoque determinista. De hecho, la reinterpretación del experimento EPR tendrá una importancia notable dentro de la reforma que propone Penrose. Pasemos a ver, pues, en qué medida dicho experimento es una de las columnas de la reforma penroseana.

3. Argumento a favor del determinismo dentro del mundo cuántico I: experimento EPR

3.1. ¿Satisface EPR un criterio de realidad serio?

Una vez Bohr presentó su contraargumento a EPR, una de las preguntas que surgieron fue si el experimento presentado en el artículo de Einstein, Podolsky y Rosen presenta un concepto de realidad satisfactorio. Pues bien, Penrose piensa que dicha idea no corre tanto peligro como pretende hacer ver Bohr y sus partidarios, siempre y cuando se haga un análisis adecuado de esta.

Entrar de lleno en querer definir el concepto de realidad puede llevarnos hacia un terreno farragoso, pero es indispensable ofrecer una noción al menos general para poder abordar el asunto que nos concierne en este apartado.

Según el diccionario Akal de filosofía, por ejemplo, la realidad la podemos entender del siguiente modo:

> […] según su uso filosófico normal, el modo en que las cosas son, en oposición a su mera apariencia. La apariencia tiene que ver con cómo ve las cosas un perceptor o grupo de perceptores determinado. En ocasiones se dice que la realidad es doblemente independiente de la apariencia (2004: 828).

Esto tiene como consecuencia necesaria que la apariencia no determina la realidad, ya que estamos hablando de una realidad *objetiva*, o como dice en la definición, independiente de quien la observa y [o] estudia. Tanto Einstein como Penrose son partidarios en su mayor medida de una idea de realidad de esta naturaleza y es por ella por lo que sienten cierta simpatía por el realismo. Ahora bien, ¿a qué tipo de realismo en concreto se adscriben sino lo hacen a un realismo al uso? Penrose en este tema, como en todos los que conciernen a un debate filosófico profundo, prefiere ser cauto y ofrece una respuesta escuetamente general. Sin embargo, también suscribe el modo en el que Einstein defiende su particular punto de vista del realismo.

Ya hemos visto anteriormente que el realismo de Einstein es, cuanto menos, peculiar. Para Einstein, la realidad *objetiva* es susceptible de ser conocida a partir de estudios científicos, principalmente aquellos que involucran a la física (ya que una parte importante de esa realidad *objetiva* es física). Al tener tanta relevancia el aspecto físico de la naturaleza es acertado pensar que el peso del conocimiento de la realidad recaiga en la experimentación y la

[188] Científicos no muy posteriores a Bohr, como David Bohm, son un claro ejemplo de esta corriente que no acabó por desaparecer nunca, a pesar de la victoria del físico danés.

observación de la misma, tal y como fue con Galileo, por ejemplo. No obstante, Einstein no procede de tal modo, como muy bien señala Penrose:

> Hay una moraleja en esta historia: las motivaciones de Einstein para dedicar ocho o más años de su vida al desarrollo de la teoría general de la relatividad no eran ni producto de la observación ni de la experimentación (Penrose et al., 1999: 24).

Einstein entiende que un desarrollo adecuado de las matemáticas en la física nos puede aportar el deseado conocimiento real de la naturaleza. Es decir, que si bien la experimentación es imprescindible, no menos cierto es que el éxito de una teoría no depende en último término de esta, sino de consistencia y la adecuación de la teoría:

> La teoría fue desarrollada originalmente sin tener que responder a ninguna observación: la teoría matemática es muy elegante y está físicamente muy bien motivada. La cuestión es que la estructura matemática está precisamente allí, en la naturaleza, y la teoría existe realmente allí, en el espacio: no ha sido impuesta a la naturaleza por nadie […] Einstein revelaba algo que ya estaba presente. Más aún, no era simplemente algún elemento menor de física lo que descubrió: es lo más fundamental que tenemos en nuestro Universo, la propia naturaleza del espacio y del tiempo (Penrose et al., 1999: 25).

A pesar de que en cierto sentido Einstein se aleje de Galileo y sus métodos de búsqueda de la verdad, no deja de ser patente que no abandona plenamente al pisano, ya que su actitud sigue siendo la de descifrar el libro de la Naturaleza a través de las matemáticas. Esta es una idea que seduce completamente a Penrose y es por ello por lo que entiende que no existe un concepto de realidad más adecuado que el defendido por Einstein, llegando al punto de aceptar en gran medida la realidad expuesta en EPR. Ya vimos más arriba que el concepto de realidad utilizado en EPR podía diferir en algunos aspectos con los de Einstein, pero ello no le impide aceptar aquello que se deriva de tal idea.

Penrose piensa que una reinterpretación de las consecuencias de EPR podrían ser mejor visualizadas si se tienen en cuenta ejemplos diferentes de la misma naturaleza. De tal forma, defiende, EPR adquiere una amplitud mayor y es por ello por lo que él mismo aporta varios ejemplos. Pasemos a ver algunos de ellos.

3.2. Experimentos tipo EPR y la idea fundamental en estos: el entrelazamiento

La propuesta alternativa de experimento EPR más relevante es la desarrollada por David Bohm en 1951. El experimento de Bohm parte de la suposición de la desintegración de una partícula de espín cero en un punto concreto (central) y de esa desintegración surgen dos partículas de espín ½, las cuales se alejan en direcciones opuestas de forma exacta. Lo destacable de este caso en concreto es que una vez nos decidimos a medir el espín de una de las dos partículas producidas (a las cuales Penrose identifica como electrón una y positrón la otra) en una dirección, tenemos que la otra partícula tiene un espín en la dirección opuesta. Esto hace concluir que tal *elección* de

la medida de una de las partículas parece haber fijado (determinado) de manera simultánea el eje de giro de la otra (Penrose, 1991: 357).

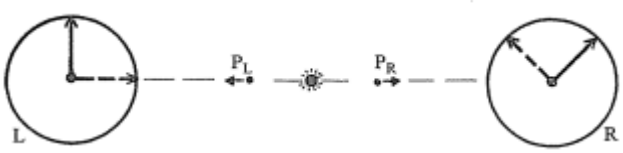

Figura 10. Representación del experimento de Bohm

La conclusión de este experimento no está basada en una inferencia lógica simple como la que acabamos de ver, sino que podemos dar con ella dentro del formalismo de la teoría cuántica[189]. Con esta conclusión se está poniendo de relieve un debate importante, este es, concretar si las partículas se comportan independientemente, tal y como lo enfocaría la física clásica o si, por el contrario, lo hacen como dicta la teoría cuántica, es decir, de manera entrelazada. Penrose no es ambiguo en este aspecto y defiende el entrelazamiento y rechaza la independencia.

Penrose destaca que los resultados de los experimentos que se han realizado en concordancia a lo expuesto en este experimento parecen decirnos que la naturaleza se comporta de manera entrelazada, es decir, como nos dice la teoría cuántica. Pero si Penrose acepta todo esto, ¿ello no lo sitúa en un bando contrario a las conclusiones de EPR? Siendo precisos la respuesta es un rotundo no. Penrose no pretende abandonar la teoría cuántica y volver hacia una física que cumpla las directrices de la clásica. Aquello que busca nuestro autor es reformar la física cuántica, encontrando aquello que no *funciona* para, así, desarrollarla de un modo más adecuado. Y el entrelazamiento es una idea que conviene tener presente, ya que este logra explicar esa *misteriosa* conexión que parece existir en la naturaleza. De hecho, Penrose defiende que a pesar de que a todas luces dicho entrelazamiento existe[190] también reconoce el misterio que supone demostrarlo[191].

El modo en el que Penrose se desmarca de la visión cuántica que enfrenta la perspectiva de EPR es optar por un realismo objetivo, es decir, un realismo en el que el papel del observador(a) es casi secundario:

> […] Creo que algo de la naturaleza de una «medida» es siempre una parte esencial del *montaje* de un experimento cuántico, para asegurar que el estado no esté contaminado por enjambres de estos entrelazamientos indeseados. Con esto no quiero decir que el experimentador monte deliberadamente una «medida» para conseguirlo. Mi idea es que la propia naturaleza está activando continuamente efectos de procesos **R**, sin que haya ninguna intención deliberada por parte de un experimentador ni ninguna intervención de un «observador consciente» (Penrose, 2006: 797).

[189] Para detalles de tal formalismo, véase (Penrose, 1991: 357-360) (Penrose, 2006: 787-789).
[190] Gracias a los experimentos que se han llevado a cabo y no a una opinión personal de él.
[191] Concretamente Penrose habla del doble misterio al que tiene que hacer frente la demostración del entrelazamiento: i) con respecto a cómo interpretar los fenómenos y ii) encontrar la razón por la que no percibimos el entrelazamiento de un modo más frecuente (Herce, 2014: 96).

Estos son los motivos filosóficos, o metafísicos, en los que se sustenta Penrose, pero no son los únicos. De hecho, nuestro autor se apoya en mayor medida en argumentos científicos. El principal es aquel con el que rechaza las desigualdades clásicas de Bell, las cuales, a grandes rasgos, niegan el entrelazamiento en favor de la independencia[192]. Para tal rechazo, Penrose se basa en los resultados de un experimento concreto, este es, el llevado a cabo por Alain Aspect y sus colegas. Los experimentos de Aspect acaban dando la razón a EPR y poniendo en tela de juicio las desigualdades de Bell. En los experimentos de Aspect se llegó a observar cierto tipo de entrelazamiento. Tales resultados, sin embargo, no son definitivos y hay quienes rechazan el abandono de las desigualdades de Bell, apuntando que el desarrollo de los aparatos de los experimentos (detectores) no nos permite llegar a un resultado concluyente. No obstante, nuestro autor es de los que sí dan crédito al trabajo de Aspect:

> […] A mi modo de ver, sería extraordinariamente *im*probable que el excelente acuerdo entre teoría cuántica y experimento que se manifiesta en el experimento de Aspect sea de alguna forma un artificio –un artificio de la *poca* sensibilidad de los detectores- y que con detectores más perfectos el acuerdo con la teoría desaparecería, en la medida precisa y considerable que sería necesaria para que se recuperaran las relaciones de Bell (Penrose, 2012: 266).

Otro experimento tipo EPR que Penrose trae a colación es el propuesto por Lucien Hardy[193]. Este comparte con el de Bohm la emisión de dos partículas de espín ½ en direcciones opuestas desde un punto central. Ambas partículas son dirigidas hacia detectores de espín situados en puntos muy alejados (denotados como L – ya que se dirige hacia la izquierda- y R –su dirección es hacia la derecha). La diferencia con respecto al experimento de Bohm es que el estado inicial del punto central no es de espín 0, sino un estado de espín 1 particular (Penrose, 2017: 242).

Penrose ofrece una explicación técnica de este experimento tipo EPR en EcalR (Penrose, 2006: 792-794), pero en MFF desarrolla un contraejemplo que, en mi opinión, hace más clara la explicación de aquello que pretende exponer. El citado contraejemplo consiste en intentar ver si un modelo clásico puede dar cuenta de los problemas, sobre todo el relacionado con el entrelazamiento, que suscitan los experimentos tipo EPR. Para ello pide que imaginemos el ejemplo de Hardy, pero con la particularidad de que las partículas estén preprogramadas para dar determinados resultados al llegar a los detectores. A su vez, tales resultados dependerán de cómo esté orientado cada uno de los detectores. Otra característica es que «los componentes individuales del mecanismo que gobierna el comportamiento de cada

[192] Bell ilustra su idea con un sencillo ejemplo, el cual es conocido como *calcetines de Bertlmann*, que Penrose explica del siguiente modo: Bertlmann era un colega suyo que invariablemente llevaba calcetines de distinto color […], si uno pudiera echar una ojeada a su calcetín izquierdo y advirtiera que era verde, entonces sabría instantáneamente que su calcetín derecho *no* era verde. En cualquier caso, no sería razonable inferir que había una misteriosa «influencia» que viajaba instantáneamente desde su calcetín izquierdo a su calcetín derecho […]. El efecto puede conseguirse haciendo simplemente que Bertlmann decida por adelantado que sus calcetines serán de diferente color. Los calcetines de Bertlmann no violan las relaciones de Bell, y no hay influencia a larga distancia que conecte sus calcetines (Penrose, 2012: 265).

[193] No confundir con el anteriormente citado Geoffrey Harold Hardy.

partícula se envíen señales unos a otros una vez que las partículas se han separado de O» (Penrose, 2017: 243), que es como denota el punto central.

Con todo esto, los detectores de las partículas deben estar orientados para poder medir una dirección concreta (Penrose habla de la dirección ←, porque con ella podemos dar cuenta de la ortogonalidad o no ortogonalidad del sistema estudiado). El motivo de ello es que para que al menos alguna de las partículas dé como resultado dicha dirección, obedeciendo a las reglas de ortogonalidad y no ortogonalidad[194]. No obstante, este caso en el que sólo alguna de ellas esté predispuesta de esta manera, como puede requerirlo la física clásica, acaba quebrantando una de estas reglas (en concreto la primera). Esto tiene como consecuencia que si queremos cumplir con las reglas no podemos hacerlo a través de aparatos que obedezcan a la física clásica. Por tanto, la solución pasa por abandonar este modelo y abrazar la idea del entrelazamiento que supone la mecánica cuántica:

> La conclusión de todo esto es que, en muchas situaciones, objetos cuánticos separados, por muy alejados que estén entre sí, siguen interconectados y no se comportan de forma independiente. El estado cuántico de uno de estos pares de objetos está *entrelazado* (en la terminología de Schrödinger) […] Los entrelazamientos cuánticos no son, de hecho, poco habituales en mecánica cuántica. Al contrario, los encuentros entre partículas cuánticas (o sistemas previamente no entrelazados) darán casi indefectiblemente por resultado estados entrelazados. Y, una vez que están entrelazados, es muy improbable que dejen de nuevo de estarlo meramente a través de la evolución unitaria (**U**[195]) (Penrose, 2017: 244).

Si bien esto parece poco rebatible, no menos cierto es que Penrose reconoce que en realidad el entrelazamiento no resulta tan evidente como a él mismo le gustase que fuera. El entrelazamiento es una [más que] posible propiedad que no deja de ser tremendamente sutil y necesita de una visión muy sofisticada para poder intuirla (Penrose, 2017: 244). Esto, sin embargo, no impide que nuestro autor considere que el camino a seguir debe estar orientado a descifrar el entrelazamiento[196].

---

[194] Dichas reglas [necesarias] son tres y Penrose las define del siguiente modo: 1) si los dos detectores para la medición del espín en L y R se configuran para medir ↓, entonces algunas veces (con una probabilidad de 1/12, de hecho) se dará que ambos detectores en efecto obtengan ↓ (esto es, **SÍ**, **SÍ**). […] 2) si un detector se configura para medir ↓ y el otro para medir ←, entonces ambos no pueden obtener estos resultados (es decir, al menos uno de ellos obtiene **NO**). Por último, nos dice que 3) si ambos detectores se configuran para medir ←, no pueden ambos obtener el resultado opuesto →, o, dicho de otro modo, al menos uno de ellos debe obtener ← (esto es, **SÍ**) (Penrose, 2017: 243).

[195] El significado de esto lo veremos más adelante en §4.1.

[196] «En la práctica, la mayoría de los físicos parecen considerar que al realizar estos experimentos, los entrelazamientos previos de las partículas con el mundo exterior pueden ser ignorados, o promediados, sin que influyan en el resultado. Penrose considera que el problema está en la interpretación que se hacen de los resultados. Más que interpretar el formalismo **U**/**R** habría que buscar una teoría más completa que supere la aparente dicotomía. La paradoja del proceso de medida sigue siendo un reto explicativo sin solución» (Herce, 2014: 96).

3.3. El alcance «real» de EPR

¿Por qué es tan importante para Penrose demostrar el entrelazamiento entre partículas? ¿Acaso la física actual no nos dice que dicho entrelazamiento *sólo* puede ser *intuido* del modo en el que lo hace? La segunda cuestión es importante dentro de los planteamientos de Penrose, porque, como hemos visto en varias ocasiones, aquello que precisamente busca nuestro autor es reformar la física actual para dar respuesta, entre muchas otras, a esta pregunta. Una reinterpretación de EPR garantizaría un acercamiento a la realidad más adecuado, lo cual puede entenderse como respuesta a la primera de las preguntas.

Penrose está convencido de que la idea de la realidad *objetiva* defendida en EPR tiene que ser, valga la expresión, real. Y no sólo eso, sino que su naturaleza tiene que ser determinista. La interpretación penroseana del entrelazamiento cuántico es precisamente este: si todo está entrelazado, ¿hasta qué punto es correcto seguir afirmando que los procesos naturales son indeterministas? Esta afirmación-interrogación lleva consigo una idea, cuanto menos, polémica, esta es, relacionar el indeterminismo con la independencia entre los sucesos naturales. De hecho, la idea de entrelazamiento no la habríamos visto nacer si no hubiese sido por la teoría cuántica. Ahora bien, es cierto que una idea en la que la naturaleza en último término se encuentre entrelazada, incluso a nivel de partículas, casa de un modo más adecuado, al menos a grandes rasgos, con las ideas deterministas que con las indeterministas[197]. Sin embargo, todo esto no deja de ser misterioso, como bien se encarga Penrose de aclarar:

> Los entrelazamientos cuánticos son ciertamente sutiles, ya que, como hemos visto, se requiere una sofisticación considerable para detectar cuándo tales entrelazamientos están en realidad presentes. No obstante, los sistemas entrelazados cuánticamente, que deberían ser una consecuencia casi ubicua de la evolución cuántica, nos ofrecen situaciones de comportamiento holístico en que, en un sentido claro, el todo es más que la suma de las partes. Es más de una manera sutil y algo misteriosa, lo que hace que en la experiencia normal no notemos en absoluto los efectos de los entrelazamientos cuánticos. Es sin duda una cuestión desconcertante por qué, en el universo que experimentamos realmente, tales características apenas se manifiestan a pesar de estar presentes (Penrose, 2017: 244-245).

A pesar de que este problema sea misterioso, Penrose cree que aún con lo que tenemos en la ciencia actual se puede llegar un poco más al fondo del asunto. Nuestro autor habla concretamente de otra nueva reinterpretación, pero esta vez sobre la función de ondas de Schrödinger, adoptando, así, de nuevo una postura que lo acerca a Einstein (quien, recordemos, sentía gran simpatía por dicha idea como respuesta determinista).

[197] A pesar, como hemos visto, de que el concepto de entrelazamiento es tenido en cuenta a partir de los estudios cuánticos, que son reconocidos ampliamente como indeterministas.

4. Argumento a favor del determinismo dentro del mundo cuántico II: la función de ondas

4.1. Importancia de la función de ondas

Penrose se niega a creer que el determinismo fuese derrotado en los debates de principios del siglo XX que hemos visto más arriba y por ello defiende que esta corriente aún puede encontrarse en la física cuántica.

Para dicha tarea cree que es imprescindible el buen uso de los números complejos, porque, como vimos en el Capítulo 2, estos números están más relacionados con la realidad de lo que en un principio se cree. Tal relación se hace evidente concretamente a nivel cuántico[198] (Penrose, 2012: 275), mientras que a la física clásica le basta con los números reales (Penrose, 2012: 276). Los números complejos son los más adecuados porque estos poseen la capacidad de explicar las peculiaridades de la naturaleza a nivel cuántico, como lo puede ser, por ejemplo, la superposición. La superposición es un concepto clave para aquello que logra explicar la función de onda de Schrödinger, además de ser poco [o nada] intuitiva en términos de física clásica. Esto, sin embargo, no es óbice para reconocerla como algo real:

> Sencillamente se da el *hecho* de que encontramos que el mundo en el nivel cuántico se comporta *realmente* de este modo misterioso y poco familiar. Las descripciones son perfectamente claras –y nos ofrecen un micromundo que evoluciona de acuerdo con una descripción que es matemáticamente precisa y, además, *¡completamente determinista!* (Penrose, 2012: 277).

Es común relacionar el determinismo con la simplicidad, sobre todo si tenemos en cuenta que la física clásica (marcadamente determinista) es mucho más simple que la física cuántica. No obstante, Penrose sostiene que esta es una idea equivocada, ya que podemos entender el determinismo dentro del esquema de la teoría cuántica.

Un concepto que es fundamental para entender ya no sólo la propuesta de Penrose, sino también la teoría cuántica en general es el conocido como *evolución de Schrödinger* o *evolución unitaria*. La evolución unitaria es la traducción que hace la ecuación de Schrödinger (o función de ondas) del estado de un sistema, el cual se ve alterado cuando se realiza una medida sobre dicho sistema. Otro de los conceptos es el denominado como *reducción del vector de estado* o *colapso de la función de ondas*. Con la reducción del vector de estado se da cuenta del sistema cuando este está siendo sometido a la medida mencionada anteriormente. Según la ciencia que conocemos, por tanto, existe una alternancia entre el estado que describe la unidad unitaria y el que describe la reducción del vector de estado, dependiendo de si se está haciendo una medida o no en el sistema estudiado. Mientras que la evolución unitaria, al ser resultado de la ecuación de Schrödinger, es un proceso determinista, que la reducción del vector de estado es indeterminista. Es conveniente remarcar que la reducción del vector de estado generalmente es sometido a debate, el cual gira entorno a determinar si es un hecho de la «realidad» física o si se trata de una *simple* herramienta de la teoría (Penrose,

[198] […] nivel de los objetos físicos que son «suficientemente pequeños» en algún sentido, tales como moléculas, átomos o partículas fundamentales (Penrose, 2012: 276).

2006: 713). Por su parte, Penrose no entra en tal debate y se limita a explicar dicho proceso según se utiliza en mecánica cuántica. Como último apunte, cabe decir, para posteriores menciones, que la evolución unitaria es denotada como **U**, mientras que el colapso de la función de ondas lo es como **R**.

La pregunta que surge de nuevo es si **U** sólo puede dar cuenta, por su condición determinista, de los acontecimientos simples de la naturaleza. Y de nuevo la respuesta es no. De hecho, **U** contempla de manera no forzada la anteriormente mencionada superposición. Penrose lo explica con el ejemplo del comportamiento de las partículas de la luz:

> Consideremos una situación en la que incide luz sobre un espejo semiplateado, es decir, un espejo semitransparente que refleja solamente la mitad de la luz que incide sobre él y transmite la mitad restante. Ahora bien, en teoría cuántica la luz se considera compuesta de partículas llamadas *fotones*. Podríamos haber imaginado perfectamente que, en una corriente de fotones que inciden sobre nuestro espejo semiplateado, la mitad de los fotones serán reflejados y la mitad transmitidos. ¡Nada de eso! La teoría cuántica nos dice que, en lugar de ello, cada fotón *individual*, cuando incide en el espejo, se coloca por su cuenta en un estado superpuesto de reflexión y transmisión. Si antes de su encuentro con el espejo el fotón está en el estado |A), entonces a partir de ese momento evoluciona de acuerdo con **U** para convertirse en un estado que puede escribirse |B) + i |C), donde |B) representa el estado en el que el fotón se transmite a través del espejo y |C) el estado en el que el fotón es reflejado por él (Penrose, 2012: 279).

En el planteamiento de este problema no se está suponiendo que cada uno de los fotones se desdoble y acaben siendo dos, sino que gracias a los números complejos que permiten los cálculos en la teoría cuántica se contempla esa coexistencia entre las alternativas (Penrose, 2012: 281). Esta es una situación que ya fue expuesta por el mismo Schrödinger con su célebre experimento mental del gato, en el cual nos detendremos en el siguiente punto.

Con respecto a **R**, Penrose defiende que si bien para la física actual este es completamente necesario, no menos cierto es que con ello no podemos aspirar a dar una descripción fiel de los procesos naturales, ¡porque es contradictoria a **U**! Con R podemos medir uno de los dos estados superpuestos que se describen gracias a **U**, con el inconveniente de que esa medida sólo encuentra una explicación en el azar, debido al «salto» que requiere. **R** es muy necesario porque traduce los procesos cuánticos complejos a términos de física clásica más simples. Ahora bien, según Penrose, la física (o, más bien, los/as físicos/as) no puede conformarse con esta descripción y debe seguir en busca de una más completa, que no definitiva.

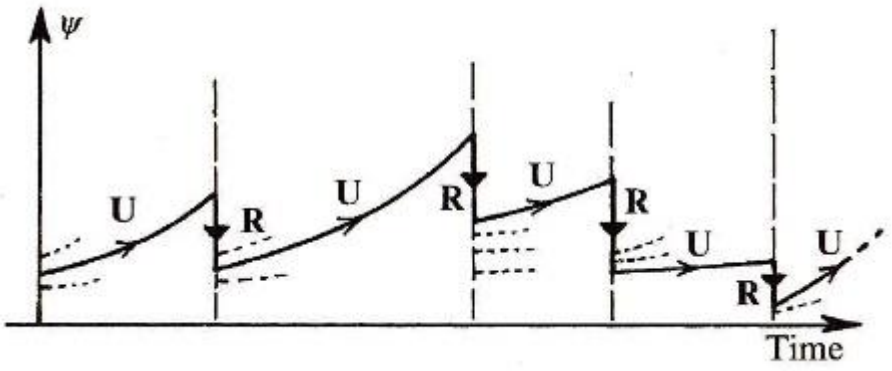

Figura 11. Evolución temporal del estado de un sistema (en el caso concreto de la ilustración, ψ), siendo **U** determinista y continua, y **R** indeterminista y discontinua.

Penrose está convencido de que **R** no tiene la capacidad de describir la «realidad» del estado cuántico que se proponga describir. Es una herramienta muy útil, pero que tiene fecha de caducidad. Tan radical es su postura que defiende que no se puede creer en la mecánica cuántica de forma plena, es decir, aceptando **U** y **R** al mismo tiempo. La clave está en **U** y ello podemos volver a verlo con el gato de Schrödinger.

4.2. Al rescate del gato de Schrödinger

La física está enfocada de forma inadecuada y una relectura del experimento del gato de Schrödinger podría facilitarnos algunas pistas sobre ese enfoque erróneo. Para ello, Penrose modifica levemente las condiciones originales propuesta por Schrödinger. El cambio más significativo es situar a alguien que observe el experimento desde dentro de la habitación, donde originariamente sólo se encontraba el gato. No obstante, este no es el único. Otra de las modificaciones es cambiar el suceso cuántico: mientras que en el original este se correspondía con la desintegración de un átomo radioactivo, en el de Penrose se insiste en que este esté representado por el disparo de una fotocélula por un fotón, que es reflejado en un espejo semi-reflectante (Penrose, 1991: 367). El motivo por el que Penrose decide situar a alguien que observe el experimento es porque esta, en principio, sería la perspectiva que corresponde a **R**, ya que quien observa puede ver al gato o vivo o muerto y no en ambos estados a la vez. Sólo a un observador exterior a la habitación le pertenece una observación correspondiente a **U**. Con estas modificaciones nuestro autor resalta lo que también pretendía enfatizar Schrödinger, esto es, la incapacidad que tenemos de explicar la superposición en el ámbito macrofísico. Pero al contrario que el físico austríaco[199], Penrose no llega a pensar que esta condición se deba a alguna deficiencia del proceso **U**, sino más bien de **R**.

[199] Al menos en la primera etapa de su pensamiento, Schrödinger era más partidario de la perspectiva indeterminista. Pero, siendo justos, lo cierto es que nunca se inclinó de forma férrea a ninguna de las dos corrientes.

¿Qué consecuencias o a qué conclusiones nos lleva la nueva interpretación del experimento de Schrödinger por parte de Penrose? Aquello que nuestro autor pretende poner encima de la mesa es que la perspectiva de la física actual es extremadamente subjetiva. Evidentemente esto no es conveniente si lo que queremos es ofrecer una descripción fiel de la naturaleza. Sin embargo, lo único que puede aportar Penrose a este problema es su disconformidad, apelando a la convicción, filosóficamente legítima, de que nuestra concepción de la experiencia está muy limitada:

> […] Mi opinión es que no veo por qué lo que llamamos «experiencias» deben no estar superpuestas. ¿Por qué un observador no debería ser capaz de experimentar una superposición cuántica? No es a lo que estamos acostumbrados, desde luego, pero ¿*por qué* no lo estamos? Se podría argumentar que sabemos tan poco sobre lo que constituye en realidad una «experiencia» humana que tenemos ciertamente derecho a especular sobre estas cuestiones en un sentido o en otro. Pero podemos sin duda cuestionar por qué habríamos de permitir que las experiencias humanas des-superpongan un estado cuántico dado en dos estados paralelos del universo en lugar de mantener solo *un* estado superpuesto, que es lo que la descripción **U** nos proporciona realmente (Penrose, 2017: 269).

Sin duda, la propuesta de Penrose es muy sugerente y muy probablemente ampliaría la capacidad explicativa de la ciencia. No obstante, también es evidente que la posibilidad real de llevar a cabo un cambio de este calibre se antoja, cuanto menos, poco factible. Esta es precisamente la crítica más recurrente a la que se tiene que enfrentar Penrose cuando expone sus ideas. Por otro lado, ello no le impide seguir defendiendo su postura a capa y espada. Un gato puede ser un objeto de estudio demasiado complejo para que la superposición *pueda verse* reflejada en él. A pesar de todo, ello no debería suponer un problema insalvable para la ciencia, porque, como decía Einstein [de nuevo], lo importante es la construcción *debida* de la teoría. Rescatemos, pues, al gato.

Nuestro autor considera que si su propuesta supone un abismo con respecto a la ciencia actual es por la situación de esta no es propicia para que se contemplen cambios, ya no sólo muy pronunciados como los que él sugiere, sino para cualquier tipo de cambio.

4.3. Situación de la teoría cuántica

Según Penrose, existen varios aspectos en la física actual que no permiten un cambio de enfoque.

El primero de ellos está relacionado con lo que acabamos de ver con el gato de Schrödinger. La teoría cuántica tiene el inconveniente de contemplar todas las posibilidades. Por una parte esto puede ser positivo, pero por otra, que es lo que defiende Penrose, puede ser negativo. Para nuestro autor es negativo porque ello no permite determinar qué cálculos son posibles y cuáles imposibles, ya que no existe una distinción tajante entre las distintas posibilidades (Penrose, 1991: 370). Pero que la teoría cuántica deje abierta todas las posibilidades no significa que en ella todo valga. De hecho, si hay algo que caracteriza a la teoría cuántica es su potente precisión en la

descripción de los procesos naturales. La crítica de Penrose, sin embargo, es plenamente razonable.

Otro de los aspectos es que la teoría cuántica no ofrece una descripción adecuada del entorno de los experimentos (tanto mentales como empíricos). Penrose reconoce que este es un problema real, pero también se encarga de aclarar que intentar abordar una descripción completa del entorno es una tarea imposible (Penrose, 1991: 370). Después de todo, no podemos tener los atributos del demonio de Laplace. El universo no deja de ser tremendamente complejo y nosotros parecemos condenados a ofrecer una respuesta subjetiva acerca de él. ¿Es posible apartarnos de tal perspectiva? Penrose tiene la esperanza de que así sea, aunque reconoce la dificultad que ello supone.

¿Por dónde pasa la solución? Como ya hemos visto en reiteradas ocasiones, Penrose defiende que volver a la perspectiva determinista nos ayudaría a lograr un enfoque distinto, que es lo realmente necesario según la situación actual de la física. La propuesta de nuestro autor no resulta, ni mucho menos, fácil de llevar a cabo. De hecho, el mismo Penrose reconoce la dificultad que supone dejarlo todo en manos del determinismo:

> Podríamos tratar de adoptar la postura de que la evolución real es la **U** determinista, pero que las probabilidades surgen de las incertidumbres envueltas de saber cuál *es* realmente el estado cuántico del sistema combinado. Esto sería adoptar una visión muy «clásica» sobre el origen de las probabilidades –la de que todas surgen de las incertidumbres en el estado inicial. Podríamos imaginar que diferencias pequeñísimas en el estado inicial podrían dar lugar a diferencias enormes en la evolución, como el «caos» que puede ocurrir en los sistemas clásicos […] Sin embargo, tales efectos de «caos» sencillamente no pueden ocurrir con **U** por sí solo, puesto que es *lineal*: ¡las superposiciones lineales no deseadas persisten para siempre con **U**! Para resolver una superposición tal en una alternativa o la otra, se necesita algo *no*-lineal, de modo que **U** sólo no basta (Penrose, 1991: 371).

Es necesario, por tanto, encontrar un procedimiento no-lineal que permita el ansiado cambio de enfoque. Este, para Penrose, es un problema real y es perfectamente perceptible a través del problema de la consciencia (¡siendo este precisamente el argumento principal de sus trabajos!). Si la consciencia no puede ser explicada de un modo adecuado tal que podamos determinar que esta no es computable se debe, precisamente, a que la física actual no lo permite.

Penrose no tiene la soberbia y la audacia necesarias para creerse el único que haya planteado este tipo de problemas desde la perspectiva en la que lo hace. Por ello reconoce los trabajos de Von Neumann, Wheeler o Wigner, a pesar de que ninguno de ellos le convenza del todo. Algo más de crédito concede a la teoría de Everet de los *muchos universos* o *universos múltiples*. Si bien es cierto que no se suscribe de forma plena a este planteamiento concreto, ello no impide que sienta cierta simpatía por algunas de sus ideas fundamentales. Aquella que más le convence es la renuncia al proceso indeterminista **R** por parte de la teoría de Everet[200], a pesar de que no comparta la idea al completo:

[200] La teoría de Everet dice, a grandes rasgos, que existen tantos universos como alternativas puedan existir no dependiendo de medidas, sino de la amplificación macroscópica de estados cuánticos (Penrose, 1991: 373).

> […] Se ha alegado que la «ilusión» de **R** puede, en cierto sentido, deducirse efectivamente en esta imagen pero no creo que estas afirmaciones se sostengan. Cuando menos se necesitan más ingredientes para que el esquema funcione. Creo que la idea de los muchos universos introduce una multitud de problemas propios sin resolver realmente los *auténticos* enigmas de la medida cuántica (Penrose, 1991: 373).

Al fin y al cabo, el problema de la mecánica cuántica es que si bien es idónea para describir procesos y resolver problemas que para la física clásica son imposibles de abordar, no menos cierto es que su capacidad no es tan infalible a nivel macroscópico. De hecho, la física clásica es más práctica y más fiel al ámbito macroscópico que la teoría cuántica. En consideración de Penrose, este es un factor que no hay que dejar de tener en cuenta a la hora de juzgar la necesidad de una reforma en la física actual. La muleta indeterminista y subjetiva de **R** ¡no puede ser la solución definitiva! No obstante, Penrose reconoce que **R** puede llegar a ofrecernos una descripción objetiva del comportamiento de una partícula, siempre y cuando no estén involucradas varias de ellas (Penrose, 1991: 374). El conocimiento que nos proporciona **R**, por tanto, está limitado inevitablemente y es por ello por lo que necesitamos dar con nuevas alternativas, siempre aprovechando lo que ya tenemos:

> […] necesitamos comprender la nueva ley para ver cómo el mundo cuántico enlaza con el clásico. ¡Creo también que necesitamos esta nueva ley si queremos conocer alguna vez las mentes! Por todo esto pienso que debemos buscar nuevas claves (Penrose, 1991: 376).

5. ¿Hasta qué punto es posible una reforma en la física moderna?

5.1. Proyecciones de Penrose sobre la posibilidad de la reforma que pretende

La de Penrose es una postura muy particular, pero esta también pertenece a una corriente muy específica de la física, esta es, la [per]seguidora de la *Gravedad Cuántica* (QG, por sus siglas en inglés), la cual, a su vez, pertenece al movimiento de las *Teorías del Todo* (TOE). Como la gran mayoría de estas teorías, la que presenta Penrose encuentra muchas dificultades a la hora de verse confirmada, debido, como es obvio, a su novedad. Hay quienes son menos optimistas y piensan que esta incapacidad de confirmación se debe a que realmente este tipo de teorías están condenadas al fracaso. Como ya sabemos, nuestro autor se aleja de manera evidente de esta última actitud y piensa que el camino de la QG es estrictamente necesario para poder seguir dando respuestas que cada vez sean más precisas.

En el caso concreto de Penrose, este aboga por una propuesta con un plus de originalidad, ya que sugiere un enfoque diferente con respecto a la relación entre la mecánica cuántica y la relatividad general:

> […] 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 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 integrar 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 del contorno, entonces podría dar una explicación de los fenómenos cuánticos que se aprecian en las singularidades y de ese modo englobar los fenómenos cuánticos ya conocidos (Herce, 2014: 153).

La pregunta pertinente que surge de nuevo es: ¿hasta qué punto es posible llevar a cabo la pretensión de Penrose? Esta cuestión parece estar sometida al mismo problema de la pescadilla que se muerde la cola, ya que si se argumenta que la propuesta está muy alejada de implementarse, se puede contraargumentar (como de hecho lo hace Penrose) que si se encuentra en dicha situación es, precisamente, porque se necesita tal cambio.

Las principales críticas a los planteamientos de Penrose giran en torno a la no-aplicabilidad en la física actual y a que debe aportar algo más que la simple convicción de nuestro autor para que la reforma que pretende pueda llevarse a la práctica. Estas críticas, por otra parte, no dejan de ser injustas, ya que Penrose no se limita a lanzar argumentos al amparo de sus corazonadas. Él mismo ha desarrollado una teoría propia, la de los twistores, la cual está en vías de desarrollo de cara al futuro y que tiene la ambición de ser contemplada como ese ingrediente que falta en la física actual.

Si las ideas de Penrose pueden llegar a ser revolucionarias para la física moderna, ¿por qué no tienen un mayor alcance? Apelar a la novedad de su propuesta nunca ha sido un argumento que haya utilizado Penrose en su favor y lo cierto es que tampoco ha tratado a fondo tal cuestión hasta sus últimos trabajos. Nuestro autor piensa que el factor de la moda dentro de la física es un componente muy a tener en cuenta, ya que en última se obedece a esta más de lo que en un principio pueda llegar a creerse.

5.2. La moda como obstáculo real de una evolución en la física moderna

Penrose no considera que la moda[201] sea lo único que dicta lo que *se debe* investigar o no en la física o en la ciencia en general. Existen aspectos prácticos y la contribución al conocimiento del campo es imprescindible para que una teoría prospere y se estudie de manera extendida. Sin embargo, esta condición tampoco es exclusiva. Hay que valorar el factor moda dentro de la ciencia y Penrose lo expone con una teoría en concreto: la teoría de cuerdas moderna.

Aunque Penrose quiera aclarar que la moda tiene una importancia capital para el devenir de las investigaciones a través del caso concreto de la teoría de cuerdas, también da cuenta de que esta dinámica no es cosa de la actualidad, sino que se ha dado a lo largo de la historia. Para ello hace mención a la teoría de los cuatro elementos con formas geométricas ampliamente aceptada en la Antigua Grecia, los estudios astronómicos de Ptolomeo, que constituyeron la imagen del universo durante siglos, o la teoría del flogisto, que permaneció vigente durante más de un siglo. El éxito de estas teorías de la naturaleza no se debió a una simple aceptación *per se* de las élites

[201] Las siguientes consideraciones están sacadas de la última gran obra de Penrose del 2016, *Moda, fe y fantasía en la nueva física del universo*.

intelectuales de la época. Detrás de ellas hay un encaje complejo y elegante de unas matemáticas (al menos en las dos primeras) que parecen constituir el mundo que nos rodea. Sin embargo, todas finalmente resultaron erróneas o, siendo algo más benévolos, menos precisas de lo que en un principio se creía. Penrose defiende que estas teorías no hubiesen permanecido tanto tiempo en lo alto de la palestra sin el factor moda que las sustentara.

Otros ejemplos con los que nuestro autor intenta destacar el papel de la moda en la ciencia son aquellas teorías que se *abandonaron* pero que fueron *rescatadas* con el paso del tiempo. Tales ejemplos son la teoría de los átomos como nudos de William Thompson, rescatada precisamente por la teoría de cuerdas, o la idea platónica del cosmos como un dodecaedro, la cual se ha visto revitalizada hace escasos años (2003).

Con respecto al argumento principal en el que Penrose habla de la teoría de cuerdas moderna, nuestro autor llega a hacer un repaso del alcance de dicha teoría y de los distintos temas que puede llegar a abordar (supersimetría, mundos de brana, etc.). Pero dicho repaso lo hace con el pretexto de acentuar su idea principal, esta es, que la teoría de cuerdas más que poder explicativo goza del privilegio de estar de moda. Y no hay que entender de un modo equivocado esta postura. Penrose reconoce de muy buena gana aquello de positivo que tiene la teoría de cuerdas, de ahí a que lo enumere. No obstante, también quiere hacer ver que los resultados que se obtienen con la teoría de cuerdas no son aquellos que en un principio se esperaban de esta (siendo menos amplios).

Ahora bien, ¿cuáles son los factores necesarios para que una teoría –o varias- estén de moda? ¿La simple influencia de la comunidad científica marca el camino a seguir? Penrose está muy lejos de ser ingenuo y no deja caer todo el peso en el mandato de la comunidad científica, a pesar de que esta tiene una importancia innegable. Nuestro autor destaca varios aspectos. Como podremos ver, todos constituyen una parte muy importante en continuar la moda más que a participar en la creación en sí misma de la moda.

El primero de tales aspectos es el más evidente: el potencial explicativo de la teoría. Aunque la teoría de cuerdas tiene serias limitaciones, no menos cierto es que su amplitud es considerable. El segundo tiene que ver con la actitud de las nuevas generaciones científicas, las cuales suelen seguir las teorías de moda para intentar revolucionarlas, consiguiendo con ello perpetuar que sigan de moda. El tercero es el concerniente a la parte de la financiación. La teoría de cuerdas es, sin duda, una de las teorías que más recauda en materia de investigación, lo cual la hace tener un prestigio de las que las demás no gozan[202].

El hecho de que Penrose destaque estos aspectos no significa que pretenda denostar a las teorías que están de moda. Nuestro autor siente un profundo respeto por tales teorías, aunque ello no impide que encuentre necesario tener más en cuenta teorías que puedan ofrecer *algo más* en el ámbito experimental.

[202] Este hecho es algo que no es del agrado de Penrose, quien llega a decir: En mi opinión, la teoría de cuerdas ha tenido una representación desproporcionada desde hace muchos años. No cabe duda de que en la teoría hay suficientes aspectos fascinantes y de que merece mucho la pena continuar desarrollándola […]. Pero su dominio de los desarrollos en física fundamental ha resultado empobrecedor y, en mi opinión, ha entorpecido el desarrollo de otras áreas que, en última instancia, podrían haber albergado una mayor promesa de éxito (Penrose, 2017: 126).

La teoría de cuerdas ha llegado a tal punto de sofisticación que casi ha acabado siendo una teoría de matemáticas puras, hasta el punto de poder prescindir de la experiencia. A pesar de que Penrose sienta una gran simpatía por las matemáticas puras, también es cierto que nuestro autor espera que tales matemáticas puedan verse reflejadas en el mundo material.

¿Cómo de acertado o desacertado está Penrose en esta cuestión? Una respuesta seria acerca de este tema conllevaría hacer un estudio que podría ampliarse hasta extremos insospechados (ya que influyen factores sociológicos, científicos, económicos, etc.). Por otra parte, en el tema concreto que aborda nuestro autor sí que parece tener la razón consigo, una vez visto, claro está, el devenir de la teoría de cuerdas, la cual se considera, a grandes rasgos, casi obsoleta.

Estando o no de moda, la postura de Penrose toma contacto con muchos terrenos del conocimiento y las implicaciones filosóficas que nos regalan sus propuestas hacen de la suya una perspectiva realmente abarcadora.

Capítulo 4

Argumentos científico-filosóficos

1. ¿Qué entiende Penrose por consciencia?

1.1. Características de la consciencia [según Penrose]

Ya hemos visto que Penrose no es muy dado a ofrecer definiciones exactas de conceptos. También hemos podido ver que ello ha sido objeto de crítica en numerosas ocasiones. La estrategia de Penrose con respecto a definir los conceptos fundamentales de los debates en los que se sumerge se caracteriza por entender de un modo general tales conceptos y a partir de esas ideas generales añadir aspectos que sólo contribuyen a esa misma generalidad para, de ese modo, no entrar en debates conceptuales. Veamos, pues, cuáles son las características que Penrose encuentra en uno de los conceptos más importantes de su obra: la consciencia.

Para empezar, nuestro autor pone encima de la mesa la complejidad que surge de determinar hasta qué punto es posible saber qué (o quienes) tiene(n) consciencia y qué (o quienes) no. Sin duda este es el *quid* de la cuestión:

> […] ¿Revelaría siempre su presencia la consciencia en algún objeto? Me gustaría pensar que la respuesta a esta pregunta es necesariamente «sí». Sin embargo, mi fe en esto se ve alentada por la falta total de consenso sobre en qué partes del reino animal debe encontrarse la consciencia. Algunos no admiten que pueda poseerla en absoluto algún animal no humano (y, algunos, ni siquiera que la poseyeran los seres humanos antes de alrededor del año 1000 a. C., cfr. Jaynes, 1980), mientras que otros atribuirán consciencia a un insecto, un gusano, ¡o quizá incluso a una roca! Por mi parte dudaría que un gusano o un insecto — aunque ciertamente no una roca— tenga mucho, si tiene algo, de esta cualidad; pero los mamíferos, de modo

> general, me dan la impresión de tener cierta genuina consciencia. De esta falta de consenso podemos inferir, al menos, que no hay un criterio generalmente aceptado para la manifestación de la consciencia. Pudiera ser todavía que exista una impronta de comportamiento consciente, aunque no universalmente reconocida […] (Penrose, 1991: 505).

Lo que sí parece tener claro Penrose es que la consciencia tiene dos partes claramente diferenciadas, que son la *activa* y la *pasiva*, siendo, para él, la *activa* la que permite la identificación de la consciencia. En la continuación del anterior párrafo citado dice con respecto a este punto:

> […] Aun así, esto sólo podría significar el papel *activo* de la conciencia. Es difícil ver cómo la simple presencia de consciencia, sin su contrapartida activa, pudiera ser directamente verificada. Esto tiene un apoyo en el terrible hecho de que, durante algún tiempo en los años cuarenta, el curare fue utilizado como «anestésico» en operaciones realizadas a niños, cuando en realidad el efecto de esta droga es paralizar la acción de los nervios motores sobre los músculos, de modo que la agonía que *realmente experimentaban* estos infortunados niños no tenía modo de percibirla el cirujano en esa época (Penrose, 1991: 505-506).

De hecho, el motivo de la convicción de Penrose es doble. El primero de ellos se basa en el esencial papel que para nuestro autor tienen las intuiciones del sentido común. Esto puede llevar a preguntarnos, ¿cómo algo tan poco fiable (científicamente hablando) como el sentido común es para Penrose una de las garantías para que podamos identificar la consciencia? Si recordamos bien, nuestro autor sabe de las dificultades por las que las máquinas pasan cuando hacen frente a problemas que requieren de sentido común (humano). Obviamente, clamar por el sentido común no es casual. También es cierto que la justificación que ofrece Penrose no es desdeñable (esta es, creer -con sentido común- que es poco probable que alguien que *no esté* consciente *parezca* estarlo).

Por su parte, el segundo es igualmente especulativo, aunque también responde a una convicción firme. Dicho motivo es la selección natural. Para Penrose es inconcebible que la consciencia no tenga ningún propósito evolutivo, y por ello se pregunta « ¿por qué la Naturaleza se tomó la molestia de hacer evolucionar cerebros conscientes cuando parece que hubieran bastado cerebros «autómatas» no-sintientes como los cerebelos[203]?» (Penrose, 1991: 506). Sin duda, la pregunta es pertinente, pero sobre cómo responde Penrose a esta nos centraremos en el siguiente punto.

Otra de las características de la consciencia que Penrose destaca es la «autoconsciencia». Este rasgo parece uno de esos puntos en los que existe un consenso muy generalizado. Sin embargo, Penrose advierte que es plenamente necesario comprender de forma adecuada la «autoconsciencia». Nuestro autor destaca que existe una idea extendida de que un sistema posee «autoconsciencia» cuando este contiene dentro de sí un modelo de *sí mismo*. Entenderlo de esta forma es un error, por la sencilla razón de que cualquier máquina (incluidas las de las primeras generaciones) está constituida de tal forma. Que tales máquinas no tienen consciencia está reconocido incluso por quienes defienden la IA *fuerte* (excepto algunos casos extremos). Por tanto, este tipo de «autoconsciencia» no puede ser definitoria de la consciencia:

[203] Si bien se intuye que la función del cerebelo es automática, en § 3.1 veremos con algún detalle más el papel que tiene esta parte concreta del cerebro.

> […] Pero un programa de ordenador que contenga dentro de sí (digamos como subrutina) alguna descripción de otro programa de ordenador no hace al primer programa consciente del segundo; ni ningún aspecto *auto*-referencial de un programa le hace *auto*-consciente. A pesar de las afirmaciones que parecen hacerse con frecuencia, los temas reales concernientes a la consciencia y autoconsciencia apenas se tocan, en mi opinión, en consideraciones de este tipo. Una vídeo-cámara no es consciente de las escenas que está registrando, y tampoco una vídeo-cámara que esté dirigida hacia un espejo posee auto-consciencia (Penrose, 1991: 508-509).

Al fin y al cabo, aquello que rechaza Penrose es que los procesos mecánicos (como esta forma de «autoconsciencia») sean los que caractericen la consciencia. Por una parte, no tiene problemas en admitir que los seres humanos también actuemos de manera mecánica en numerosas ocasiones (por ejemplo, en las acciones llevadas a cabo por el cerebelo). No obstante, piensa que dichas acciones pertenecen al ámbito de lo inconsciente, de la parte *pasiva*. Es decir, volvemos a la idea de que la consciencia no puede ser algorítmica. ¿A qué debe Penrose su convicción de que la consciencia no responde a procedimientos algorítmicos? Ya vimos en el capítulo 2 que nuestro autor responde diciendo que tal y como él mismo *piensa* las matemáticas le resulta imposible concebir que nuestro pensamiento (ya no sólo matemático, sino en general) responda en última instancia a un algoritmo, por muy sutil que este pueda llegar a ser. Esto no quiere decir que los seres humanos sean ajenos a los procesos algorítmicos. Es más, Penrose reconoce que en el proceso de hacer matemáticas *buscamos* nuevos procedimientos algorítmicos. Aun así, esa *búsqueda* en sí misma no es algorítmica (Penrose, 1991: 512).

Este último asunto guarda relación con otra característica que Penrose rechaza, esta es, la importancia que se le pretende dar a la aleatoriedad en la actividad de la consciencia[204]. No es que nuestro autor obvie la importancia de la aleatoriedad para determinados aspectos, pero sí que considera que no es definitiva para alcanzar un mayor conocimiento de la consciencia. Esto se debe a que sistemas computacionales *pseudoaleatorios* pueden simular casi perfectamente sistemas puramente aleatorios. Es decir, esta característica no puede aportar la no-computabilidad que Penrose busca (Herce, 2014: 135).

Este rechazo a la aleatoriedad por parte de Penrose tampoco es casual. Ya hemos podido ver que nuestro autor se considera a sí mismo como determinista. Un elemento como la aleatoriedad hasta sus últimas consecuencias choca de frente con un punto de vista determinista, y es por ello por lo que Penrose no la acoge de forma plena.

En §2.2 podremos comprobar que el determinismo se encuentra, como pudo vislumbrarse más arriba, en la idea que Penrose tiene acerca de la selección natural.

En el siguiente punto veremos la selección natural, pero en lo que atañe a cómo esta afecta –o no- a los algoritmos.

[204] Esta es una idea que defiende en SM y ahí habla en concreto de la actividad del cerebro. Esto, sin embargo, no supone un conflicto, ya que nuestro autor defiende que la consciencia surge del cerebro.

1.2. La consciencia y selección natural [algorítmica]

¿La consciencia es fruto de la selección natural? Y en el caso de que la consciencia dependa de un algoritmo sutil, ¿dicho algoritmo debe su *razón de ser*, por tanto, a la selección natural? Tanto en NME como en SM[205], Penrose ofrece sus explicaciones con respecto a este asunto concreto desde la perspectiva contraria a la suya propia (es decir, aquella que responde de manera afirmativa a la segunda pregunta) para, de ese modo, acabar por contradecirla.

El modo de buscar la contradicción por parte de nuestro autor es la siguiente. Suponemos que el algoritmo (o los algoritmos[206]) originador de la consciencia ha sido el elegido por la evolución. Tal algoritmo debe tener la capacidad de realizar juicios de *validez* acerca de los otros algoritmos que [tenemos la certeza que] poseemos los seres humanos. Pero, ¡aquí está el problema (y la contradicción)! Penrose no acude a su argumento gödeliano, sino a la máquina de Turing, la cual nos dice que la decisión de parar no se da algorítmicamente (Penrose, 1991: 514). Por tanto, los procesos algorítmicos no pueden tener la eficacia de crear algo tan sutil como lo es la consciencia, la cual sí tiene la potestad de decisión.

Quienes defienden la naturaleza algorítmica de la consciencia aún podrían reclamar que la selección natural da lugar a algoritmos válidos de forma *aproximada*, por lo que sortearían el problema señalado. Obviamente esta contrarréplica no convence a Penrose. Para nuestro autor, la selección natural en este escenario (de algoritmos aproximadamente válidos) actuaría en el modo de respuesta de los algoritmos (*outputs*) y no a las ideas subyacentes de tales algoritmos. Esta tarea no sólo sería inservible en el aspecto evolutivo sino que, además, sería impracticable (basta con intentar verificar un algoritmo a través de su *output*) (Penrose, 1991: 514). ¡La naturaleza es mucho más inteligente y práctica de lo que se pretende insinuar en este supuesto!

Al fin y al cabo, aquello que quiere volver a expresar Penrose es su defensa de que la computabilidad no es la respuesta al problema de la consciencia. Pero su exposición no se limita a repetir lo anteriormente dicho. Más bien aclara algunas de las implicaciones de su alegato, algunas de ellas, curiosamente, no han sido localizadas por sus oponentes. La más importante (en relación a entender de un modo adecuado la perspectiva penroseana) es aquella en la que Penrose reconoce la *posibilidad* de que se pueda *descubrir* aquello que permite la aparición de la consciencia y que ello pueda ser *construido* por nosotros mismos. ¿No es esta la postura contra la que carga Penrose a lo largo [y ancho] de su obra? Realmente no. Cuando nuestro autor habla de la imposibilidad de obtener una consciencia a través de una *construcción* humana, remarca que si esta no puede darse es por el empeño de creer que hay que encaminar dicha tarea en términos de computabilidad. Y he aquí el error:

---

[205] Penrose en SM centra su atención en el pensamiento matemático. Pero, como hemos podido ver en repetidas ocasiones, cuando nuestro autor habla en tales términos no lo hace con afán discriminatorio, sino por concretizar su discurso (ya que considera que el pensamiento en general y el matemático *actúan* del mismo modo).

[206] Ya vimos en su exposición de los diferentes puntos de vista (§2, Cap.1) que entiende igual de imposible que sea un algoritmo o varios los que originen la consciencia.

> Si alguna vez descubrimos en detalle qué permite a un objeto físico llegar a ser consciente, entonces sería concebible que pudiéramos construir tales objetos por nosotros mismos, aunque podrían no calificarse como «máquinas» en el sentido que ahora entendemos. Podríamos imaginar que estos objetos tendrían una tremenda ventaja sobre nosotros, puesto que podrían diseñarse *específicamente* para una tarea, a saber, *alcanzar consciencia*. No tendrían que crecer a partir de una sola célula. No tendrían que arrastrar el «equipaje» de su ascendencia (las viejas e «inútiles» partes del cerebro o del cuerpo que sobreviven en nosotros gracias a los «accidentes» de nuestros remotos ancestros). Podríamos imaginar, a la vista de estas ventajas, que tales objetos podrían tener éxito en superar *efectivamente* a los seres humanos en las tareas en las que (en opinión de gente como yo) las computadoras algorítmicas están condenadas a la subordinación (Penrose, 1991: 515-516).

Sólo de una forma no-computacional podría darse una inteligencia artificial. La consciencia, en tanto que se conciba como algorítmica o computable, está condenada a permanecer en ese rango de *lo misterioso* para el ser humano.

De todos modos, lo que Penrose quiere expresar en el fondo de la anterior cita es su oposición a quienes entienden que la consciencia es computable y no reclamar la creación de la consciencia por parte de los humanos. Eso parece estar muy lejos de nuestro alcance, al menos a medio y corto plazo. Como partidario del punto de vista C, tiene fe en que la ciencia pueda dar cuenta de este problema hasta sus últimas consecuencias (en este caso tal *creación*). Pero ya vimos que su punto de vista también es partidaria de esperar a que lleguen cambios en la ciencia para emprender el nuevo camino que llevaría a respuestas más definitivas[207].

1.3. ¿Tan incapaces son los algoritmos?

Una de las críticas más recurrentes contra la postura de Penrose es aquella en la que se le acusa de no reconocer el mérito de los procesos algorítmicos. Esto generalmente se ha relacionado con una especie de deslegitimación hacia los algoritmos para cualquier tarea inteligente. Sin embargo, en el caso de Penrose en particular esta conclusión no se sigue de sus argumentos.

Penrose reconoce que los procesos algorítmicos tienen la capacidad de crear cierto tipo de inteligencia. Lo que sucede es que ¡las inteligencias de los ordenadores y de los humanos son diferentes! Esto no quiere decir que los algoritmos sean incapaces de realizar cualquier tarea que los seres humanos somos capaces de llevar a cabo. De hecho, existen muchos quehaceres en los que las computadoras han mostrado un dominio equivalente y en algunos casos superior al de los seres humanos. El *quid* de la cuestión, sin embargo, reside en si el modo en el que los ordenadores y los seres humanos llevan a

[207] Este tipo de ideas también quedan reflejadas en SM: […] yo no estoy defendiendo en absoluto que sea necesariamente imposible construir un *dispositivo* genuinamente inteligente, siempre que tal dispositivo no sea una «máquina» en el sentido específico de estar controlada computacionalmente. En lugar de ello, tendríamos que incorporar el mismo tipo de acción física que es responsable de provocar nuestra propia consciencia. Puesto que no tenemos aún ninguna teoría física de dicha acción, es ciertamente prematuro especular sobre si o cuándo podría construirse un dispositivo semejante (Penrose, 2012: 414).

cabo los mismos procesos para llegar a los mismos resultados. Y es ahí donde Penrose responde con un rotundo «no».

El ejemplo del ajedrez es recurrente en la exposición de Penrose, ya que los ordenadores *muestran* una habilidad muy similar a la del ser humano, pero, en su opinión (y yo personalmente la comparto), el modo en el que *ejecutan* sus estrategias es diferente:

> […] (El ordenador) Puede contener una gran cantidad de conocimiento almacenado, mientras que un jugador humano también puede hacer esto. El ordenador puede efectuar repetidamente aplicaciones extraordinariamente rápidas y precisas de las comprensiones de los programadores de una forma totalmente carente de finalidad, pero lo hace en una medida que supera con mucho la capacidad de cualquier ser humano. El jugador humano necesita continuar haciendo juicios una y otra vez y elaborando planes con contenido, con una comprensión global de lo que es el juego. Estas son cualidades que no están en absoluto disponibles para el ordenador, aunque, en una buena medida, este puede utilizar su potencia computacional para compensar su falta de comprensión real (Penrose, 2012: 417-418).

Si el ordenador puede conseguir resultados muy parejos a los del ser humano en determinadas actividades (que requieren cierto ejercicio mental) es porque es extremadamente bueno en algunos aspectos, mientras que en otros llega a carecer de ellos[208]. Es decir, es ese *equilibrio* entre la virtud y el déficit lo que le hace tener la habilidad, pero lo que también lo distingue del ser humano (el cual, por su parte, tiene sus aptitudes más compensadas). Por tanto, los algoritmos tienen una capacidad de acción concreta que está limitada por su propia naturaleza.

Esta declaración, empero, no convence a quienes defienden el quehacer de los procesos algorítmicos. Siguiendo con el caso del ajedrez, Bostrom, por ejemplo, cree que quien piensa como Penrose[209] tiene una concepción equivocada de cómo se desarrollan los algoritmos y el alcance *real* de estos. Para Bostrom, el proceso de desarrollo para ver el alcance de los algoritmos se asemeja al que es experimentado en las teorías científicas, las cuales pueden explicar muchos fenómenos, pero si se siguen desarrollando acaban por poder describir de un modo más preciso el orden de aquello que nos rodea:

> La experiencia en el juego de ajedrez se logró mediante un algoritmo sorprendentemente simple. Es tentador especular que otras capacidades, como la capacidad de razonamiento general o alguna habilidad clave involucrada en la programación, también se pueden lograr a través de un algoritmo sorprendentemente simple. El hecho de que se obtenga el mejor rendimiento a la vez a través de un mecanismo complicado no significa que ningún mecanismo simple pueda hacer el trabajo tan bien o mejor. Podría ser simplemente que nadie ha encontrado la alternativa más simple. El sistema ptolemaico (con la Tierra en el centro, orbitada por el Sol, la Luna, los planetas y las estrellas) representó el estado del arte en astronomía durante más de mil años, y su precisión predictiva se mejoró a lo largo de los siglos al complicarse progresivamente el modelo: agregar epiciclos sobre epiciclos a los movimientos celestes postulados. Luego, todo el sistema fue derrocado por la teoría heliocéntrica de Copérnico, que era más simple y, aunque solo después de una mayor elaboración por parte de Kepler, más predictivamente precisa (Bostrom, 2014: 14).

[208] En el caso concreto que hemos visto, aquello de lo que carecen los ordenadores, según Penrose, es la «comprensión» del juego.
[209] No se refiere a él directamente.

La defensa de quienes comparten esta perspectiva de Bostrom es que los algoritmos aún se encuentran en fase de maduración. No podemos decir todavía hasta dónde puede llegar el margen de acción de los procesos computacionales, sobre todo si tenemos en cuenta los diferentes avances que casi de manera diaria surgen en el campo de la IA. Es injusto hacer una valoración de lo que podrán hacer las máquinas sin haber comprobado todo su potencial. Si miramos a nuestro alrededor, no es descabellado pensar que las máquinas puedan acabar por superarnos [con creces] en habilidades que *parecen* que sólo nos pertenece a los humanos. Entonces, ¿por qué arrebatarles una posibilidad que parece inevitable e incluso inminente?

Para Penrose, la cuestión no gira entorno a quitarles el mérito a lo que hacen o lo que harán los ordenadores a través de los algoritmos. Nuestro autor piensa que el desarrollo de las máquinas irá a más y que podrá alcanzar un nivel de sutileza en sus acciones cada vez mayor, tanto que será muy complicado diferenciarlo del de los seres humanos. No obstante, ello no significará que su modo de alcanzar tales metas sea el mismo por el que los seres humanos las alcanzamos.

No se trata, por tanto, de declarar una incapacidad plena de los algoritmos, en el sentido amplio. De forma concreta, esta es, alcanzar las capacidades del ser humano, dicha incapacidad sí que sería plena. Todo ello tiene que ver, de nuevo, con la diferencia entre el comprender y el pensar de las máquinas y los seres humanos. Por esta misma razón es preciso que tengamos claro [dentro de nuestras posibilidades] el modo en el que comprendemos y pensamos los humanos para, de tal manera, poder dar un veredicto más certero. Pero la gran cuestión es: ¿cómo pensamos los seres humanos?

1.4. El pensamiento y su relación con el lenguaje

Al inicio de este capítulo hemos visto que en relación a los conceptos esenciales del debate en el que Penrose está inmerso, este no ofrece definiciones o descripciones que pretendan ser precisas. En lugar de ello, encuentra más adecuado basar tales definiciones y descripciones en cómo los entiende él personalmente (que suele ser desde el plano más general), e intentando añadir algún tipo de respaldo científico.

A pesar de que Penrose lleva a cabo tal estrategia como medida de precaución para entrar en debates, lo cierto es que los temas que saca a la luz son inevitablemente objeto de polémica.

Para seguir definiendo el pensamiento humano, nuestro autor considera oportuno traer a colación [nada más y nada menos que] el papel que juega el lenguaje dentro de este.

La postura de Penrose es clara con respecto a este tema: podemos prescindir del lenguaje para pensar. Nuestro autor reconoce que para muchos sectores (obviamente los relacionados con las áreas de las humanidades) su postura es inaceptable. Pero eso es porque no atienden a lo que sucede *realmente*.

Penrose considera que las facultades de la inspiración, la intuición y la originalidad dan muy buenas pistas de cómo el verdadero pensamiento no precisa del lenguaje. Para hacer manifiesta esta idea expone una serie de

anécdotas (incluida una propia), las cuales tienen en común que en todas ellas se contribuía a problemas científicos (acción que obviamente requiere de pensamiento). Veamos cómo relata su experiencia a un problema concreto[210] en NME:

> […] Un colega (Ivor Robinson) vino a visitarme desde los Estados Unidos y habíamos iniciado una animada conversación sobre un tema muy diferente mientras paseábamos por la calle hacia mi despacho en el Birbeck College de Londres. La conversación se interrumpió al cruzar una calle lateral y se reanudó al otro lado. ¡Por lo visto, durante esos breves instantes, se me ocurrió una idea, pero luego la reanudación de la conversación la borró de mi mente!
> Ese mismo día, después de que mi colega hubiera partido, volví a mi despacho. Recuerdo que tenía una extraña sensación de júbilo que no podía explicar. Empecé a repasar en mi mente todas las cosas que me habían sucedido durante el día, intentando encontrar qué era lo que había causado este júbilo. Después de eliminar muchas posibilidades inadecuadas me vino finalmente a la mente la idea que había tenido al cruzar la calle – ¡una idea que me había proporcionado una momentánea alegría dándome la solución al problema que había estado dando vueltas en el fondo de mi cabeza! Aparentemente ese era el criterio necesario —que posteriormente denominé una «superficie atrapada»— y entonces no me llevó mucho tiempo construir las líneas generales de una demostración del teorema que había estado buscando (Penrose, 1965). Aun así, pasó algún tiempo antes de que la demostración estuviese formulada de un modo completamente riguroso, pero la idea que había tenido mientras cruzaba la calle había sido la clave. (A veces me pregunto qué habría pasado si me hubiera sucedido alguna otra experiencia gozosa durante ese día. ¡Quizá no hubiera recordado nunca la idea de la superficie-atrapada!) (Penrose, 1991: 521).

Quienes sean contrarios a Penrose pueden argumentar que, precisamente, hasta que él no verbalizó aquello que le llevó a la respuesta del problema esta no se hizo efectiva. Pero nuestro autor no estaría conforme, ya que si quiere contar algo con la anécdota es que ¡no necesitó de las palabras para llegar al pensamiento que le permitió obtener la respuesta!

El modo en el que responde Penrose a esta posible crítica es acudiendo al plano estético. Nuestro autor defiende que en la inspiración y la intuición el papel de la estética es verdaderamente importante (Penrose, 1991: 521).

¿En qué sentido puede la estética constituir una base firme para aspectos fundamentales del pensamiento como lo son la inspiración y la intuición? A esta pregunta Penrose responde de un modo muy cauteloso, ya que entiende que sus razones son avaladas más por convicciones internas que por razones que pueda demostrar científicamente.

Penrose mantiene que en su experiencia propia los juicios estéticos no sólo intervienen exclusivamente en el instante que describe en la anécdota (ese *clic* que le permitió dar con la respuesta). Nuestro autor insiste en que estos también lo hacen en alto grado en el proceso de hacer conjeturas una vez se confeccionan los argumentos rigurosos que pretende desarrollar, siempre en el marco lógico y de los hechos conocidos (Penrose, 1991: 523). En lo que respecta a este segundo aspecto parece más evidente que los juicios estéticos entren en escena, ya que es innegable que en estos se da el pensamiento. Sin embargo, ¿sucede lo propio con las intuiciones repentinas como las de la anécdota de Penrose? Nuestro autor opina que efectivamente así es, ya que

[210] Dicho problema está relacionado con los agujeros negros y la postura con respecto a ello por parte de Oppenheimer y Snyder. Para los detalles del problema véase NME, (1991: 520).

una idea estéticamente inaceptable sería rechazada y olvidada por la consciencia, que es en última instancia la que hace posible los juicios estéticos[211]. Pese a que los juicios estéticos de este tipo dependan de la consciencia, Penrose reconoce que también ellos obedecen significativamente al inconsciente. Ahora bien, la última palabra siempre la tiene la consciencia. Como vimos más arriba, esto responde más bien a una convicción interna de la que Penrose no puede dar cuenta al modo científico, algo que verdaderamente le pesa pero de lo que no duda en seguir defendiendo:

> […] Debo estar de acuerdo, asimismo, en que no puede tratarse simplemente de que la mente inconsciente esté lanzando ideas aleatoriamente. Debe haber un proceso de selección enormemente poderoso que permite que la mente consciente sea perturbada sólo por ideas que tienen alguna posibilidad. Sugeriría que estos criterios de selección –principalmente los estéticos, de alguna especie- han sido ya fuertemente influenciados por los desiderata conscientes (como la sensación de fealdad que acompaña a las ideas matemáticas que son inconsistentes con los principios generales ya establecidos) (Penrose, 1991: 523).

La consciencia permite juicios estéticos (tanto repentinos -en un modo difícil de explicar- como los someramente sopesados) en el plano de la búsqueda de respuestas científicas[212]. Nuestro autor sostiene que esto también está relacionado con la originalidad de las ideas, que él explica a través de los factores de *propuesta* y *rechazo*, siendo el primero perteneciente al inconsciente y el segundo a la consciencia. La *propuesta* de una idea original puede perfectamente ser inconsciente (de hecho, puede darse en los sueños), de tal modo que es donde por norma general se da. Pero tal *propuesta* puede quedar en nada si esta no pasa el examen de la consciencia, siendo, por tanto, el factor *rechazo* aquello que garantiza que una idea original se concretice o no. Pero, ¿de qué depende esta concretización (es decir, el modo en el que

[211] A pesar de que Penrose ande con pies de plomo a la hora de abordar temas «estrictamente» filosóficos, el tema de la importancia de la estética está presente también en SM, donde la idea subyacente queda más explícitamente expuesta. Dicha idea es que los juicios estéticos son también no-computacionales, por lo que es imposible que los ordenadores puedan llevar a cabo este tipo de juicios. Penrose encuentra en este argumento un rasgo distintivo del pensamiento humano con respecto al de los ordenadores que podría tener un peso importante. Ante la pregunta de si un ordenador inteligentemente programado podría tener la capacidad de crear obras de arte Penrose responde: Esta es una cuestión delicada, me parece a mí. La respuesta inmediata, creo yo, es simplemente «no» –aunque sólo sea porque el ordenador no puede poseer las cualidades sensitivas que son necesarias para juzgar lo bueno y lo malo, o lo soberbio y lo meramente competente. Pero, podemos preguntar: ¿por qué es necesario que el ordenador «sienta» realmente para que pueda desarrollar sus propios «criterios estéticos» y formar sus propios juicios? Uno podría imaginar que tales juicios podrían «emerger» simplemente después de un largo período de aprendizaje (de-abajo-arriba). Sin embargo, como sucede con la cualidad de comprensión, tengo la impresión de que es mucho más probable que los criterios tuvieran que formar parte de input deliberado del ordenador, habiendo sido estos criterios destilados con cuidado a partir de un análisis detallado de-arriba-abajo (muy posiblemente asistido por ordenador) que ha sido llevado a cabo por seres humanos estéticamente sensibles (Penrose, 2012: 421).

[212] Esta no es una perspectiva en la que Penrose se encuentre solo. Son muchos(as) los(as) que mantienen que al aspecto estético es fundamental en el camino que lleva a la verdad. En un tono muy parejo al de Penrose, Hardy decía: Los modelos de un matemático, al igual que los de un pintor o un poeta deben ser *hermosos*; las ideas, como los colores o las palabras, deben ensamblarse de una forma armoniosa. La belleza es la primera señal, pues en el mundo no hay un lugar permanente para las matemáticas feas (Hardy, 2014: 14).

opera la consciencia o pensamiento)? Para Penrose esta respuesta no puede, paradójicamente, concretizarse. Como sabemos, nuestro autor apuesta por una acción no-computable de la que no podemos dar cuenta hoy en día. A pesar de las dificultades que entraña esta cuestión, Penrose tiene claro que el lenguaje, por ejemplo (y con ello volvemos al principio), no es la respuesta a este problema.

En la descripción del pensamiento y sus características que hemos visto, podemos comprobar que el lenguaje es prescindible. Todo parece depender de un *algo* sutil que se nos escapa de las manos y de las palabras.

Siendo un asunto del que tan poco se puede decir certeramente, ¿en qué basa Penrose su postura? Pues una vez más volvemos a toparnos en el camino con las matemáticas. Cuando hacemos matemáticas ciertamente *pensamos*, y en este tipo de pensamiento (que contiene en sí mismo grados de intuición, inspiración y originalidad) no necesitamos el lenguaje. Penrose no quiere decir con esto, ni mucho menos, que el pensamiento pueda reducirse a matemáticas o algo por el estilo. De hecho reconoce que para quienes se dedican de un modo más directo con el lenguaje defiendan que este sea imprescindible para el pensamiento. Lo que Penrose quiere decir con su exposición es que las matemáticas siendo fuente de pensamiento, que no del pensamiento en general, y no necesitando del lenguaje, este último no puede ser tampoco la base sobre la que se sostenga el pensamiento en general. Así lo expresa nuestro autor:

> Esto no quiere decir que no piense a veces con palabras, sino que simplemente encuentro las palabras casi inútiles para el pensamiento matemático. Otros tipos de pensamiento, quizá tales como el filosofar, parecen mucho más adecuados para la expresión verbal. ¡Quizá sea esta la razón del por qué muchos filósofos parecen ser de la opinión de que el lenguaje es esencial para el pensamiento inteligente o consciente! Sin duda, personas diferentes piensan de modos muy diferentes –como ha sido ciertamente mi propia experiencia, incluso entre matemáticos (Penrose, 1991: 523)[213].

Aunque Penrose se meta en un jardín, como lo es intentar desentrañar aquello que constituye el pensamiento, este no resulta un problema de mucha envergadura para su proyecto intelectual, ya que dicho debate sigue muy abierto y su postura no es excesivamente polémica.

2. Vuelta al debate entre determinismo e indeterminismo

2.1. El determinismo: características y la visión penroseana

> *[…] Esta misma regularidad que la astronomía nos señala con respecto al movimiento de los planetas, aparece en todos los fenómenos. La curva trazada por una simple molécula de aire o de vapor responde a la misma precisión de las*

[213] En este aspecto Hardy también supone un apoyo a la postura de Penrose. El matemático inglés decía al respecto: Por otra parte, un matemático no tiene otro material para trabajar más que ideas, y, por tanto, sus modelos es probable que duren más tiempo, ya que las ideas envejecen más lentamente que las palabras (Hardy, 2014: 14).

> *órbitas planetarias. Toda diferencia en ellas, es producto de nuestra ignorancia (Laplace, 2009: 6-7).*
>
> *A pesar de la ignorancia en que nos encontramos respecto a los caminos de la naturaleza o de la esencia de los seres, de sus propiedades, elementos, proporciones y combinaciones, conocemos, sin embargo, las leyes simples y generales según las cuales se mueven los cuerpos, y vemos que algunas de esas leyes comunes a todos los seres no se desmienten jamás (Holbach, 2008: 57).*

El determinismo es, en mi opinión, un tema en el esquema de Penrose muy similar al platonismo, en el sentido de que es transversal dentro de su pensamiento, pero, sin embargo, no es explicado de forma extensa por nuestro autor. Esto tiene como inevitable consecuencia que se entienda de un modo inadecuado aquello que pretende exponer.

Antes de pasar a ver cómo nuestro autor trata este tema encuentro conveniente precisar qué se entiende por determinismo en un sentido más amplio.

Generalmente, para hablar de determinismo es necesario tener en cuenta que este concepto suele distinguirse en dos tipos: el *científico* y el *filosófico*.

El denominado determinismo *científico* se define como la «teoría según la cual el estado del mundo en un instante cualquiera determina un único futuro posible, y el conocimiento de la posición de todos los objetos, así como de las fuerzas naturales dominantes en cada momento, permitirían que un ser inteligente pudiese predecir el estado futuro del mundo con absoluta precisión»[214]. Esta definición se ajusta fielmente a la elaborada por el matemático francés Pierre-Simon de Laplace, de ahí que este tipo de determinismo también sea conocido como *laplaciano*.

Por su parte, el determinismo *filosófico* se entiende como aquella teoría en el que «todo elemento depende de algunos otros fenómenos, de manera tal que puede ser previsto, producido o impedido con seguridad, con sólo conocer, producir o impedir aquellos fenómenos[215]» (Février, 1957: 16).

Veamos cuáles son las principales semejanzas y diferencias entre ambos tipos de determinismo.

En primer lugar, tenemos la importancia de la causalidad. En el determinismo, tanto *científico* como *filosófico*, una vez conocemos de forma certera la causa de un fenómeno, podemos asegurar [de manera igualmente certera] el conocimiento del efecto que producirá. La noción de determinismo, por tanto, lleva en su seno la condición de la necesidad[216].

---

[214] La definición está tomada del Diccionario de Filosofía Akal (2004: 250).

[215] Aunque manejemos esta concepción, cabe decir que hay quienes piensan que no basta con remitir la causa de todo a otros fenómenos. Es necesario también tener en cuenta que esos fenómenos tienen a su vez que ser inteligibles porque se sustentan en postulados filosóficos indudables.

[216] Esto puede detectarse en otras expresiones del determinismo, como lo son el determinismo teológico y el fatalismo. Carlos Moya los expone del siguiente modo: […] Otra es el llamado *determinismo teológico*, basado en los supuestos de la existencia y omnisciencia de Dios. Dios, siendo omnisciente, conoce ahora el valor de verdad de cualquier proposición, incluyendo las proposiciones sobre acciones y decisiones futuras de cualquier agente. Así, si

Por otro lado, existen rasgos que distinguen a un determinismo del otro. La principal divergencia reside en que el determinismo *científico* es más concreto que el *filosófico*. El primero, como su propio nombre indica, está limitado al conocimiento de los fenómenos, es decir, a aquello que es susceptible de ser estudiado por la ciencia. Mientras, el segundo abarca también conocimientos en los que la ciencia tiene poco o nada que decir (Février, 1957: 17).

Sin embargo, para algunos(as) pensadores(as), como la gran estudiosa sobre este tema Paulette Février, ven que esta diferencia acaba siendo superficial. Una vez el determinismo *científico* se topa con ciertos límites debe abandonar a la ciencia como principio sustentador, haciéndose de esta manera casi indistinguible del determinismo *filosófico*:

> Así, se advierte cómo, incluso cuando se trata del determinismo estrictamente científico, es decir, restringido a los fenómenos observables, en las definiciones que han dado los sabios mismos se trata todavía de una creencia de un principio racional, mucho más que de una comprobación a posteriori asimilable a un dato experimental puro. Y esto no facilita la distinción precisa que nosotros nos proponemos establecer entre determinismo filosófico y científico (Février, 1957: 18).

Pienso que es importante que nos detengamos en este aspecto, porque ello ayudará a comprender el concepto de determinismo *científico* tal y como lo entiendo yo. El análisis que hace Février acerca de esta distinción tiene, en mi opinión, dos aspectos que no están tratados de forma correcta.

El primero de ellos es dar a entender que el pensamiento científico debe carecer de fundamentos metafísicos para poder seguir considerándose científico. Si bien la ciencia no sustenta la garantía de su éxito sobre una base plenamente (y ni tan siquiera de forma mayoritaria) metafísica, no menos cierto es que la ciencia desde sus inicios hasta el día de hoy no prescinde de supuestos metafísicos. ¿Hasta qué punto podríamos decir que seguimos hablando de un determinismo *científico* si entendemos que los supuestos metafísicos no alteran su naturaleza? He aquí la dificultad de la que habla Février a la hora de querer distinguir el determinismo *científico* del *filosófico* y en donde, pienso yo, se encuentra un error de exposición. Este segundo problema surge de entender el determinismo *científico* como independiente del *filosófico*. Lo correcto (y sigo matizando que esto es una opinión personal) es concebir al primero como una categoría (o si se quiere una subcategoría) del segundo. Así, el determinismo *científico* no deja de ser en ningún momento *filosófico*, porque pertenece a él, y justo en el momento que da con los límites de la ciencia debe volver a su naturaleza filosófica, no dejando se ser, a su vez, científico.

Dios sabe que la proposición «Carlos leerá mañana el diario» es verdadera, Carlos leerá el diario mañana y si sabe que la proposición es falsa, Carlos no leerá el diario mañana. Aunque Carlos cresa que depende de él leer o no el diario mañana, su creencia es falsa. Lo que hará está ya establecido, pues de otro modo Dios no sería omnisciente. Finalmente, merece también el nombre de determinismo la tesis del *fatalismo*, estrechamente conectada con la idea de destino. La tragedia de Edipo ejemplifica esta tesis fatalista, según la cual las decisiones humanas no pueden cambiar el curso inexorable de los acontecimientos (Moya, 2017: 64).

De esta forma podemos hablar del determinismo *científico* como un determinismo concreto, mientras que cuando nos referimos al determinismo *filosófico* lo hacemos acerca de un determinismo general.

Hemos visto en esta primera distinción entre determinismo *filosófico* y *científico* que sin las aclaraciones pertinentes el asunto puede volverse problemático. No obstante, no todos los tipos de determinismo dan lugar a este tipo de inconvenientes. Algunos están claramente diferenciados y apenas resultan complejos en su identificación. Ejemplos de estos tipos de determinismos son el *lógico*, el *ontológico* y el *causal*.

El determinismo *lógico* es aquel que sostiene la tesis de que sólo hay un mundo posible (Arana, 2013: 3). Suponer cualquier cambio que quiera introducírsele, por mínimo que fuera, provocaría una transformación drástica en el conjunto. Por tanto, el determinismo *lógico* implica la aceptación de un determinismo total, ya que admitir uno parcial contradiría a este con su propio término. La cadena causal en el determinismo *lógico* es única y necesaria, haciendo del universo un todo inamovible. A pesar de la inflexibilidad a la que somete este tipo de determinismo al universo, ello no impide que dentro de este determinismo pueda reconocerse uno *regional*. El determinismo *regional* permite que la realidad pueda ser parcelada en regiones enteramente independientes (Arana, 2013: 4), pero ello nunca podría afectar a la unidad inamovible del universo, ya que esta sigue siendo necesaria. En el aspecto formal, el determinismo *lógico* también tiene su rasgo particular. Después de haber visto las anteriores características es fácil ver que este tipo de determinismo está basado en el principio del tercio excluso o, si se quiere, en el de bivalencia (Moya, 2017: 64). Algunos piensan que el plano formal puede no ser tenido en cuenta perfectamente, ya que las discusiones sobre el determinismo están centradas en lo concreto. Esto, sin embargo, no se corresponde con la realidad del asunto[217].

Con respecto al determinismo *ontológico*, al igual que sucede con el *lógico*, se persiste en la idea de unidad intentando buscar la necesidad de esta, aunque con la característica de que lo hace fuera de los entes analizados. Apartando la atención de los entes el marco del determinismo *ontológico* se presenta más abierto que el del *lógico*. Para el determinismo *ontológico*, por ejemplo, la tesis del único mundo posible puede ser una mera posibilidad. Con el determinismo *lógico* teníamos un universo abocado a seguir un único camino, en el que los entes nos daban las pistas para conocer la naturaleza determinista del universo. Sin embargo, el determinismo *ontológico* contempla la posibilidad de abrir nuevos caminos, que aunque también tengan la condición de ser necesarios, la unidad propia del *lógico* desaparece. Pero si no podemos admitir que el universo es determinista a partir de los entes involucrados, ¿sobre qué cae el peso en este tipo de determinismo? Pues en aquello que permite la descripción del comportamiento del universo, esto es, aquello que lo determina: las leyes de la naturaleza. Si el universo tiene un orden no se debe al comportamiento necesario de los entes que lo componen, sino al orden que establecen las leyes de la naturaleza. Por tanto, el orden es intrínseco en la naturaleza pero extrínseco a los entes que la componen.

[217] En las grandes discusiones sobre el determinismo en la física de principios del siglo XX, se discutió de manera extendida acerca del apartado formal. Para un estudio amplio sobre la importancia del plano formal en esta discusión véase en concreto (Jammer, 1974: 2-20).

Por último tenemos el determinismo *causal*. Este tipo de determinismo normalmente ha sido objeto de discusión, ya que el apellido causal entraña no pocas contrariedades. Precisamente para evitar entrar en los terrenos empantanados que traen consigo la causalidad he decidido tomar la definición de Carlos Moya (2017), quien ofrece una perspectiva muy panorámica de dicho concepto. Su definición del determinismo *causal* es la siguiente: todo lo que sucede, sucede necesariamente (no puede no suceder) dado el estado anterior del universo y las leyes de la naturaleza (Moya, 2017: 64). Este determinismo *causal* concreto comprende a los dos anteriores tipos de determinismo, porque pone el foco tanto en los entes del universo como en las leyes de la naturaleza. ¿Hasta qué punto es conflictivo entender de esta forma el determinismo *causal*? Puede llegar a ser un asunto peliagudo, por ejemplo, si tenemos en cuenta el determinismo *legal*, que es –a grandes rasgos- aquel que se sostiene única y exclusivamente en las leyes. La dificultad residiría en que la definición de Moya es muy abierta y no lograría aclarar hasta qué punto estaríamos hablando de un tipo de determinismo (*causal*) u otro (*legal*). Pero como no nos toparemos con el determinismo *legal* a lo largo de este trabajo no existirá el problema mencionado.

Queda manifiesto que identificar los distintos tipos de determinismos es fundamental para no caer en debates infructuosos. Y no es que persiga dar con una definición precisa de determinismo, ya que ello sólo conseguiría complicar aún más el asunto. Aquello que encuentro conveniente es tener claro cuándo se está hablando de un determinismo concreto y cuándo del general.

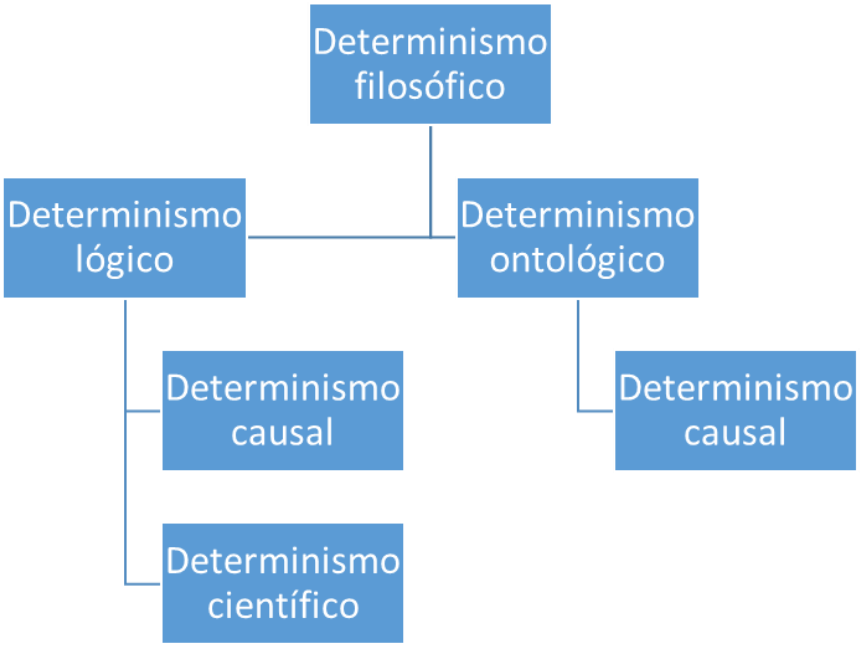

Figura 12. Esquema aclaratorio de los distintos tipos de determinismo, siendo el *filosófico* el general y los demás (diferenciados, a su vez, entre ellos) los concretos.

Una vez hecho este repaso, a continuación pasamos a ver cómo Penrose trata el determinismo en su obra. Con ello llegaremos al fin de este apartado,

este es, saber de un modo más concreto a qué tipo de determinismo pertenece el expuesto por nuestro autor. No se trata de una cuestión de encasillar el determinismo de Penrose, sino más bien de aclarar un asunto que parece no estar suficientemente explicitado en su exposición. Este aspecto no deja de ser llamativo, ya que Penrose no vacila a la hora de declararse asimismo como determinista. Sin embargo, su forma de abordar dicho concepto lo acaba arrastrando hacia una postura indiscutiblemente polémica.

En primer lugar, cabe destacar que en una parte de NME (1991: 199-203), Penrose hace un repaso de diferentes teorías científicas y las clasifica en tres categorías según estas se ajustan de forma más fiel a la realidad. La clasificación diferencia entre teorías SOBERBIAS, ÚTILES y TENTATIVAS, siendo clara la jerarquía entre ellas. El motivo de tener en cuenta tal clasificación es porque Penrose sitúa como base argumentativa (eso sí, de manera muy tímida) que las teorías SOBERBIAS tienen en común que todas ellas guardan en su seno un determinismo estricto:

> […] Normalmente, el tema del libre albedrío se discute en relación con el determinismo en física. Recordemos que en la mayoría de nuestras teorías SOBERBIAS existe un determinismo estricto, en el sentido de que si se conoce el estado del sistema en un instante cualquiera, entonces está completamente fijado para todos los instantes posteriores (o también anteriores) por las ecuaciones de la teoría (Penrose 1991: 534).

No obstante, Penrose no comete la imprudencia de hacer recaer todo el peso de su postura en este aspecto. En realidad, no parece tan siquiera que lo esté utilizando como argumento a favor del determinismo[218]. Aquello que sí constituye de un modo más palmario un punto clave en su exposición del determinismo es la relación que implanta entre este y el término «computacional» (o algorítmico).

Por norma general se suele entender que todo aquello que sea susceptible de ser computable tiene necesariamente que ser determinista. Esto tiene perfecto sentido porque la computabilidad, al igual que los procesos deterministas, tiene como una de sus características principales que aquello que acontece lo hace de un modo necesario. Por tanto, la equivalencia entre computabilidad y determinismo parece indiscutible. Pero tan sólo lo parece, porque Penrose sostiene que esto no tiene por qué ser -valga la expresión- necesariamente de tal manera.

Nuestro autor defiende que los procesos naturales, en consonancia con aquello que mantiene acerca de la consciencia (lo cual intenta no dejar de lado), contienen procedimientos que no son computables. ¿Cómo es posible que entienda que la naturaleza contiene procedimientos no-computables siendo, a la vez, determinista? Una forma cruda de decirlo es que ni el mismo Penrose lo sabe ciertamente:

> Mi propio punto de vista, aunque no esté muy bien formulado a este respecto, sería el de que algún nuevo procedimiento (GQC) toma el mando a partir de la línea divisoria cuántico-clásico e interpola entre **U** y **R** (que son ahora ambos considerados como aproximaciones), y que este nuevo procedimiento contendría un elemento *esencialmente no algorítmico*. Esto implicaría que el futuro *no sería*

[218] Como digo más arriba entre paréntesis, es que creo que sí que es un modo tímido de defender el determinismo con tal argumento, pero, sospecho, que la poca fuerza que tiene le impide seguir adelante con ello.

> *computable* a partir del presente, incluso aunque *pudiera estar determinado* por él […] Me parece bastante probable que la GQC sea una teoría determinista pero no-computable*.
> * Puede señalarse que hay al menos una aproximación a una teoría de gravitación cuántica que parece implicar un elemento de no-computabilidad (Geroch & Hartle, 1987) (Penrose, 1991: 535).

En realidad, aquello que quiere defender Penrose no encierra una dificultad inabarcable. De hecho, tal y como vimos más arriba (P.133), lo explica con el ejemplo aclarador de la meteorología. En su concepción existe el conflicto entre aquello de lo que puede dar cuenta a través de la ciencia y su convicción interna.

Para Penrose la naturaleza está determinada, pero ello no es posible demostrarse porque dicha determinación depende de un proceso no algorítmico, el cual, según veo, si se descubriese sí que podría dar la última respuesta[219]. Rubén Herce (2014), por ejemplo, si bien coincide en que

[219] Este tipo de ideas pueden resultar conflictivas, ya que existe la posibilidad de concebir que los acontecimientos puedan estar determinados, pero que, a su vez, ello no implique necesidad absoluta. Todo puede estar determinado pero también puede existir un hueco para hechos que no sean predictibles. Este es un punto de vista defendido por varios pensadores, entre ellos Carnap. Carnap cree que esto no se suele entender de tal forma porque no se tiene en cuenta que los sucesos están determinados por las leyes y la compulsión (Carnap, 1985: 186). Para el pensador alemán la clave está en que normalmente se entiende de la misma forma compulsión y predictibilidad, y esto es una equivocación que impide concebir de modo correcto el determinismo laplaciano. Los sucesos que ponen en aprietos a la defensa de un determinismo que no implica necesidad son los relacionados con los seres humanos. Defender que los sucesos se dan *necesariamente* pone en riesgo al concepto de libertad. Carnap es contrario a esta idea y piensa que, precisamente, los sucesos que conciernen a los seres humanos explican de una forma clara la diferencia entre compulsión y predictibilidad. El pensador alemán lo explica con el siguiente ejemplo:

Ahora bien, comparemos la compulsión en sus diversas formas con la determinación en el sentido de la existencia de regularidades en la naturaleza. Se sabe que los seres humanos poseen ciertos rasgos de carácter que dan regularidad a su conducta. Tengo un amigo que es sumamente afecto a ciertas composiciones musicales de Bach que raramente se ejecutan. Me entero que un grupo de excelentes músicos ofrecerá una audición privada de obras de Bach en la casa de un amigo y que en el programa figuran algunas de esas composiciones. Se me invita y se me dice que puedo llevas a alguien. Llamo a mi amigo, pero ya antes de hacerlo estoy casi seguro que él quería ir. ¿Cuál es la base sobre la que hago esta predicción? La hago por supuesto, porque conozco su carácter y ciertas leyes de la psicología. Supongamos que él viene conmigo, como yo había esperado. ¿Se vio obligado a ir? No, fue por su propia voluntad. En realidad, nunca es más libre que cuando hace una elección de esta suerte (Carnap, 1985: 187).

Tenemos entonces que el determinismo no incluye dentro de sí necesariamente la predictibilidad. El punto de vista de Carnap ofrece una explicación válida de esta condición. Sin embargo no está haciendo, en mi opinión, un análisis completo de aquello que está exponiendo. Si Carnap estuviese hablando del determinismo detectado en la naturaleza por las ciencias, o del determinismo laplaciano en el ámbito práctico (que consiste, como hemos visto más arriba, en resolver problemas de probabilidades), se le puede [e incluso se le debe] dar la razón. Pero el pensador alemán no sólo está hablando del determinismo laplaciano llevado a cabo por el matemático francés, sino también de su fondo filosófico. Y en este aspecto es donde encuentro el error de Carnap. Es cierto que Laplace es consciente de la imposibilidad del ser humano de poder alcanzar el conocimiento de su hipotético demonio. Por ello su estudio está dedicado al conocimiento probable de los sucesos del universo, que es el modo de conocer de nuestra especie. Pero esto no lo exime de las implicaciones que tiene la hipótesis del demonio. En la célebre cita dice lo que dice. Desde el punto de vista del demonio la compulsión sigue ahí aun cuando desde la perspectiva humana parece que no está. Según la física que desarrolla Laplace no podemos afirmar esto último. Pero según el

nuestro autor concibe la naturaleza como determinista y no-computable, también entiende que el sistema de la naturaleza penroseano admite cierto tipo de apertura que no se daría en un determinismo extremo (Herce, 2014: 176). Acerca de que el determinismo que expone Penrose ciertamente no es extremo[220], estoy completamente de acuerdo. No obstante, considero que la idea subyacente del determinismo que pretende defender nuestro autor puede interpretarse de un modo más cerrado del que Herce deduce.

Puestos a categorizar, personalmente me inclino a pensar que el tipo de determinismo llevado a cabo por nuestro autor es el *científico*. Realmente no veo cómo Penrose se aleja del determinismo *científico* de Laplace o de Einstein (los cuales tampoco se alejan tanto entre sí). Lo que ocurre «simplemente» es que Penrose acaba encontrándose con el problema que destacaba Février con respecto al determinismo: no poder distinguirlo del *filosófico*. Pero, tal y como defendí más arriba, esto no tiene por qué ser un problema si entendemos el determinismo *científico* como parte del *filosófico*, por lo que el primero debe volver a acudir al segundo cuando las explicaciones científicas no dan más de sí (¡que es la situación que vive la ciencia hoy en día, según denuncia el mismo Penrose!).

Penrose no niega la libertad (al menos la apariencia de esta) en lo que respecta a la dimensión humana. Sin embargo, se le hace más complicado entender la libertad (ni tan siquiera su apariencia) en el ámbito de la naturaleza del universo. Esta idea queda manifiestamente expuesta en su concepción de la selección natural, la cual, considera, no ha podido darse de forma arbitraria o «libre».

2.2. Determinismo, selección natural y principio antrópico

Según la anteriormente citada clasificación penroseana de las teorías científicas, la de la selección natural está entre las denominadas SOBERBIAS (es decir, que es de las que garantizan una descripción fiel de la naturaleza). Y no sólo eso. Tal y como apunta Herce (2014: 67), esta es la única de las teorías SOBERBIAS que no es física, lo cual hace ver su potencial y la importancia que le concede Penrose[221].

Sin lugar a dudas, tal y como sucede con la mayor parte del pensamiento de Penrose, la concepción de la selección natural que defiende nuestro autor es peculiar. Para Penrose, la selección natural de algún modo está estrechamente relacionada con el principio antrópico. Penrose se centra en los dos principios

planteamiento del demonio, que es al fin y al cabo aquel en el que cree, no existe otro modo de entenderlo. Las consecuencias filosóficas del demonio son inapelables: todo es predictible para ese ente y no hay lugar para la libertad.

[220] Para un ejemplo actual y claro de determinismo extremo véase el breve capítulo de premio Nobel en Física, Gerard t´Hooft (2019), quien incluso apuesta por una postura súper-determinista, ya que considera que esta aún no ha sido objeto de ninguna objeción (t´Hooft, 2019: 29).

[221] En primer lugar, la teoría de la selección natural en concreto de la que habla nuestro autor es la desarrollada por Darwin y Wallace. Y en segundo lugar, es de rigor aclarar que aunque Penrose reconozca un gran alcance a esta teoría, tampoco duda en recalcar que de todos modos dicho alcance está muy lejos de cualquier teoría física (Penrose, 1991: 199).

antrópicos destacados por Brandon Carter, estos son, el *débil* y el *fuerte*[222]. El *débil* nos dice que «nuestra ubicación en el universo es necesariamente privilegiada hasta el punto de ser compatible con nuestra existencia como observadores» (Arana, 2012: 332). Y por su parte, el *fuerte* «exige que sea posible que aparezcan observadores en el universo en algún estadio de su evolución» (Arana, 2012: 332). Ahora bien, ¿en qué se diferencian ambos principios? Juan Arana, siguiendo a José Manuel Alonso, cita al respecto:

> […] la diferencia entre el PAD y el PAF[223], en la formulación de Carter, consiste en que el primero alude solo a nuestra ubicación en este universo, indicando que está sesgada […] En cambio, el PAF habla no ya de nuestra ubicación en el universo, sino del universo mismo […], lo que lleva a Carter a considerar un conjunto de universos, mientras que el PAD no necesitaba considerar más que un universo (Alonso, cit. por Arana, 2012: 332).

Por su parte, Penrose entiende estos principios exactamente de la misma forma. Nuestro autor ofrece una característica que deja entrever que el principio antrópico *débil*, a pesar de su apellido no deja de ser también un planteamiento extremo. Dicho aspecto es que la exclusividad de las condiciones del mundo tal y como lo conocemos y que hace posible que las cosas acontezcan del modo en el que lo hacen:

> […] El problema se refería a varias relaciones numéricas sorprendentes que se ha observado que se dan entre las constantes físicas (la constante gravitatoria, la masa del protón, la edad del Universo, etc.). Un aspecto enigmático de esto era que algunas de estas relaciones son válidas solamente en la época actual de la historia de la Tierra, de modo que casualmente parece que estamos viviendo en un tiempo muy especial (¡con un margen de unos pocos millones de años más o menos!). Esto fue finalmente explicado por Carter y Dicke, por el hecho de que esta época coincide con la vida media de las llamadas estrellas de la secuencia principal, como es el Sol. En cualquier otra época, sigue diciendo el argumento, no existiría vida inteligente para medir las constantes físicas en cuestión -de modo que la coincidencia *tenía que* darse simplemente por el hecho de que ¡sólo existiría vida inteligente en el momento particular en que *se diera* la coincidencia! (Penrose, 1991: 537).

Con respecto al principio antrópico *fuerte*, Penrose no añade nada nuevo a lo visto más arriba. Aquello que sí cabe destacar es que nuestro autor considera que este tipo de principio antrópico no encaja en sus planteamientos.

Una vez vistos los rasgos principales del principio antrópico lo conveniente es preguntar en qué medida Penrose lo relaciona con la selección natural.

Penrose cree que algo tan sofisticado como lo es la consciencia no puede ser producto del puro azar. ¿Está reconociendo de este modo algo así como un plan oculto, el cual tenía reservado la aparición del ser humano y, con ello, de la consciencia? Como hemos visto, esto sería admitir un principio antrópico *débil* (llevado al extremo incluso) y esto no aparece explicitado en ninguno de los escritos de Penrose. Aquello que sí que se detecta en los

[222] Para las definiciones me he servido, como podrá verse citado, de la obra de Juan Arana *Los sótanos del universo*, donde trata de un modo conciso y claro tales conceptos.

[223] Son las fácilmente deducibles siglas de Principio Antrópico Débil y Principio Antrópico Fuerte.

argumentos de nuestro autor es que la idea de que una naturaleza dirigida tiene más sentido que una naturaleza arbitraria[224].

A pesar de que en ocasiones Penrose parece que vaya a dar un argumento más contundente sobre aquello que quiere defender (a mi modo de ver, un determinismo *laplaciano* sin paliativos), lo cierto es que en última instancia siempre parece recular (como todo determinismo *científico* se ve forzado a hacer). Ya no sólo acaba renunciando a la posibilidad de un principio antrópico *fuerte* (algo que en realidad no le convence en ningún momento), sino también se muestra escéptico con respecto a concebir un principio antrópico *débil* como un argumento metafísico firme que pueda explicar la aparición de la consciencia:

> Mediante la utilización del principio antrópico —ya sea en la forma fuerte o en la débil— podríamos tratar de probar que la consciencia era *inevitable* en virtud del hecho de que tenía que haber seres sensibles, es decir «nosotros», para observar el mundo, ¡así que *no* necesitamos suponer, como he hecho, que la sensibilidad tenga alguna ventaja selectiva! En mi opinión, este razonamiento es técnicamente correcto y el argumentó 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 conciencia *tiene* una poderosa ventaja selectiva, y no creo que se necesite el principio antrópico (Penrose, 1991: 538).

Es cierto que Penrose se muestra contundente en la cita a la hora de renunciar al principio antrópico [*débil*] como sustentador de la selección natural que dio lugar a la consciencia humana. Esta rotundidad, sin embargo, carece de importancia, ya que aquello que está haciendo nuestro autor es abandonar un supuesto que él mismo trajo a colación sin ninguna obligación de hacerlo. Su error es, en mi opinión, haber acudido en primer lugar al principio antrópico (ya fuera en la forma que fuese).

Por otro lado, tampoco se trata de un error de bulto por parte de Penrose. Si pone encima de la mesa el principio antrópico para darle algún tipo de crédito para más tarde quitárselo es para mostrar cómo este principio tiene sus limitaciones. Para nuestro autor, una de las limitaciones más evidentes del principio antrópico está relacionada con la incapacidad de poder dar cuenta a los misterios relacionados con la entropía, concretamente al comportamiento de esta según la segunda ley de la termodinámica.

2.3. ¿Explica el principio antrópico la entropía?

Para Penrose es importante tener en cuenta el concepto «entropía». Esta idea no sólo es relevante con respecto al estudio del universo y su origen, sino también en el plano concreto de la vida de los seres vivos. Aunque aquí no vayamos a ver en profundidad la perspectiva cosmológica de Penrose, sí que

[224] […] Parece que hubiera algo en el modo de trabajar de las leyes de la física que permitiría que la selección natural sea un proceso mucho más eficaz de lo que sería con leyes arbitrarias (Penrose, 1991: 516).

será necesario entender algunos aspectos de esta para encajarlos con la concepción que nuestro autor tiene de la entropía y el principio antrópico.

En primer lugar hay que responder a una pregunta fundamental: ¿qué es la entropía? Como hemos visto bosquejado más arriba, este concepto debe su desarrollo al campo de la termodinámica. Tal y como hace Penrose en EcalR, empezaremos viendo lo que dice la primera ley de la termodinámica para, así, tener una perspectiva más clara de lo que el concepto entropía significa.

La primera ley de la termodinámica determina que la energía total de un sistema aislado se conserva (Penrose, 2006: 928). Esta ley nos ayuda a entender la naturaleza de la energía y también a conocer su valor, el cual «permanece constante pese al hecho de que pueden tener lugar todo tipo de procesos complicados, [por lo que] la energía total después del proceso es igual a la energía antes del proceso» (Penrose, 2006: 928-929).

Penrose expone la primera ley para hacer manifiesta la diferencia que existe entre esta y la segunda ley de la termodinámica. Si bien la primera habla de una igualdad, la segunda lo hace de una desigualdad. Es aquí donde entra el concepto de entropía.

De modo general, la entropía se conoce como el «desorden manifiesto» dentro de un sistema. Esta definición, sin embargo, no hace justicia al alcance del concepto. No obstante, esto no debe resultar un problema para lo que veremos acerca de ello, ya que veremos la aplicabilidad de este concepto, sobre todo, a un rasgo muy concreto de la naturaleza: la posibilidad de vida de los seres vivos de nuestro planeta. Ahora bien, ¿qué nos dice en concreto la segunda ley de la termodinámica? Básicamente, que en la naturaleza la entropía tiende a aumentar, es decir, que el universo está cada vez más y más desordenado.

A pesar de que esta definición de la entropía no sea ni ortodoxa ni precisa, ella no nos alejará demasiado de aquello que pretende exponer Penrose acerca de dicha concepción. En realidad, las implicaciones *manifiestas* de la segunda ley son llamativamente intuitivas y no guardan misterios. De ahí que la mayoría de los físicos sientan cierta simpatía por esta, hasta el punto de creer que las teorías físicas más razonables deben satisfacerla (Penrose, 2006: 937).

Para hacer visible las implicaciones de la segunda ley, Penrose desarrolla un supuesto concreto (tomado de Boltzmann[225], aunque con añadidos propios). La imagen supone un espacio de fases[226] ($\mathcal{P}$) en el que un sistema físico (*x*) representado en un punto de dicho espacio se comporta de una manera determinada. El espacio de fases $\mathcal{P}$ está compuesto por subregiones, que se denominan como «granulado grueso» de $\mathcal{P}$ (Penrose las denomina también «cajas»). En el supuesto tenemos que el sistema *x* (en el momento AHORA) está situado en uno los granulados gruesos ($\mathcal{V}$), el cual tiene un volumen menor que aquellos que le rodean. Una vez *x* se pone en movimiento y en un espacio de fases en el que no existe ninguna inclinación o tendencia especial (como lo es este) su destino es acabar en «granulados gruesos» con

[225] Como bien se sabe, Ludwig Boltzmann (1844-1906) fue quien introdujo el concepto entropía tal y como se conoce hoy en día en termodinámica.

[226] Tomando la definición penroseana, el espacio de fases comprende que «para un sistema clásico de *n* partículas (sin características distintivas), en un espacio $\mathcal{P}$ de 6*n* dimensiones, cada uno de cuyos puntos representa la familia completa de posiciones y momentos de todas las *n* partículas» (Penrose, 2006: 929).

más volumen (¡en una abrumadora mayoría de los casos!). El sistema *x* está abocado al desorden y de forma (muy probablemente) irreversible:

> […] Una vez que *x* entra en una caja con cierta medida de entropía, se hace muy poco probable que pueda volver en un período de tiempo razonable a una caja de entropía significativamente menos que aquella de la que procedía. Alcanzar una entropía significativamente menor supondría encontrar un volumen absurdamente más minúsculo, y las probabilidades en contra de ello son inmensas (Penrose, 2006: 937).

Todo parece encajar a la perfección. Pero en la expresión «irreversible» de más arriba se encuentra una de las claves para entender que las implicaciones de la segunda ley no son tan definitivas. Con la conclusión del supuesto expuesto por Penrose *debemos* concluir una *asimetría* temporal. Esto es tremendamente conflictivo, ya que la física en la que se desarrolla el supuesto entiende la temporalidad de modo *simétrico* (Penrose, 2006: 938). ¿En qué modo se traduce la conflictividad entre la *simetría* y la *asimetría* temporal en relación a la segunda ley y el supuesto tratado?

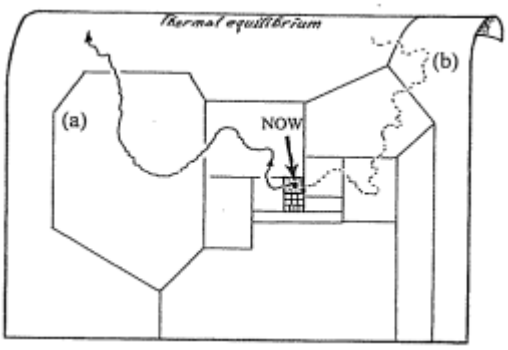

Figura 13. Representación del supuesto concreto de Penrose del espacio de fases $\mathcal{P}$. Este espacio de fases se compone de subregiones (que son de «granulado grueso» - o «cajas») y donde podemos ver que el sistema *x* (situado en una «caja» $\mathcal{V}$ y en el instante definido como punto NOW –AHORA- de la ilustración) tenderá a moverse a subregiones más grandes, de acuerdo a lo que nos dice la segunda ley.

Situándonos en el espacio de fases supuesto, en el que la entropía tiende a aumentar, tratemos de ver el comportamiento del sistema atrás en el tiempo (justo antes del punto AHORA). Si bien teóricamente la entropía debería aumentar, lo que sucede, sin embargo, es que su entropía es cada vez menor (a consecuencia de tener que provenir de «granulados gruesos» con más volumen), ¡lo cual viola la segunda ley si tenemos en cuenta la *simetría* temporal! ¿Significa esto que la segunda ley lleva inevitablemente a conclusiones absurdas en el plano físico? Penrose no lo entiende de ese modo, a pesar de que en el caso concreto del espacio de fases así lo determine. Aquello que sí expone nuestro autor es que la segunda ley tiene diferentes dimensiones con características especiales[227] y estas constituyen un misterio para la ciencia hasta donde la conocemos hoy en día. Es por ello por lo que cree conveniente examinar dichos casos en los que la entropía muestra características especiales.

[227] Estas son las que se dan en el origen del universo, los agujeros negros y los seres vivos.

Nuestra atención estará centrada, como vimos más arriba, en el caso especial de los seres vivos en nuestro planeta. La entropía en los seres vivos es especial en la medida en que para que sea posible que la vida se dé es necesario que la entropía permanezca en un nivel bajo. El papel del Sol en este escenario es imprescindible pero, matiza Penrose, no en el sentido en el que generalmente se cree:

> […] Hay extendida la falsa idea de que nuestra supervivencia depende de la *energía* que el Sol suministra. Esto es equívoco, pues para que dicha energía sea de alguna utilidad para nosotros debe proporcionarse en una forma de baja entropía. Si el cielo entero hubiera estado uniformemente iluminado, por ejemplo, con alguna temperatura uniforme –ya fuera la del Sol o cualquier otra cosa-, entonces no habría forma de utilizar esta energía […]. Un suministro de energía en equilibrio térmico es inútil. Sin embargo, nosotros tenemos la fortuna de que el Sol es un punto caliente en un *fondo frío*. Durante el día, la energía llega a la Tierra desde el Sol, pero durante el transcurso del día y la noche vuelve de nuevo al espacio. El balance neto de energía se reduce simplemente (en promedio) a que devolvemos toda la energía que recibimos (Penrose, 2006: 948).

Es decir, que la vida en la Tierra no depende simplemente de la energía procedente del Sol, sino del equilibrio que existe entre esta y la que envía la Tierra de vuelta al espacio. El equilibrio consiste en que el Sol envía a nuestro planeta fotones amarillos de alta frecuencia que contienen más energía; mientras que en el sentido contrario (esto es, de la Tierra al Sol) se emiten fotones infrarrojos de baja frecuencia que contienen menos energía. Este intercambio *equilibrado* evita, por ejemplo, que la Tierra sufra un calentamiento extremo. Pero no sólo eso, sino que también permite que seres vivos como las plantas surjan y que estas (a través de la fotosíntesis) mantengan su entropía baja y la de los seres (entre ellos, por supuesto, los humanos) que se benefician de ellas, (ya sea comiéndolas, comiendo algo que se las comen o respirando el oxígeno que liberan) (Penrose, 2006: 948). ¿Significa, entonces, que los seres vivos contradicen la segunda ley? En un principio puede parecer que sí, sobre todo teniendo en cuenta que esta se cumple en prácticamente todo el universo conocido. Sin embargo, si concebimos a los seres vivos como sistemas abiertos, estos pueden evitar contradecir lo que la segunda ley postula (Herce, 2014: 112).

Tenemos entonces que la vigencia de la segunda ley no corre peligro y el aumento de la entropía es un hecho que se cumple. Ahora bien, atendiendo a este aspecto y volviendo al hilo conductor de este punto (el determinismo), surge una pregunta: ¿está Penrose reconociendo que con el *innegable* aumento de la entropía se da lugar a una libertad, la cual daría de frente con el determinismo que él mismo defiende? En realidad no tiene por qué entenderse de ese modo. Penrose puntualiza en MFF que entender la entropía o la segunda ley en términos de organización es un error común que es conveniente evitar:

> Debo aclarar una idea que suele generar confusión. En términos coloquiales, podríamos decir que los estados de baja entropía, al ser «menos aleatorios», son por tanto «más organizados», y la segunda ley nos dice por lo tanto que la organización en el sistema se está reduciendo continuamente. Sin embargo, desde otro punto de vista, se podría decir que la organización en el estado de alta entropía en el que acaba el sistema es exactamente igual que la del estado inicial de baja entropía. La razón para esta afirmación es que (con ecuaciones dinámicas deterministas) la organización nunca se pierde, porque el estado final de alta entropía contiene

> cantidades enormes de correlaciones detalladas en los movimientos de las partículas, que son de una naturaleza tal que, si invirtiésemos cada movimiento exactamente, el sistema entero desharía el camino hasta llegar al estado inicial «organizado» de baja entropía. Esta no es más que una característica del determinismo dinámico, y nos dice que hablar tan solo de «organización» no nos aporta nada en cuanto a la comprensión de la entropía y de la segunda ley. Lo fundamental es que baja entropía corresponde a orden *manifiesto* o *macroscópico*, y que las sutiles correlaciones entre posiciones o los movimientos de los componentes submicroscópicos (partículas o átomos) *no* son cosas que contribuyan a la entropía del sistema. De hecho, esta es una cuestión central en la definición de la entropía, y sin esos adjetivos, «manifiesto» o «macroscópico», en las descripciones anteriores de entropía, no habríamos sido capaces de avanzar hacia una comprensión de la entropía y del contenido físico de la segunda ley (Penrose, 2017: 313-314).

Por tanto, según Penrose, el determinismo no se ve afectado por la entropía, sino que, en cierto sentido, resulta de algún modo beneficiado por las implicaciones de esta y la segunda ley[228].

Es fácil ver por qué un concepto como el de la entropía sirve para defender la postura determinista penroseana. Al fin y al cabo un concepto que incluya algún tipo de *tendencia* o *finalidad* generalmente se erige como un apoyo para el determinismo. Y si esto es así, ¿por qué Penrose es tan reacio a aceptar el principio antrópico? El motivo de que nuestro autor vacile en este aspecto concreto se debe, como vimos bosquejado más arriba, a que en términos cosmológicos el principio antrópico no responde de un modo adecuado a los misterios que entraña la entropía.

Penrose llega a este veredicto a través del análisis de una postura cosmológica concreta, esta es, la inflacionaria[229]. Penrose define [a grandes rasgos] esta teoría[230] del siguiente modo:

> [Según los inflacionistas] nuestro universo, casi inmediatamente tras originarse en el Big Bang, en un período extraordinariamente breve de entre 10^{-36} y 10^{-32} segundos tras *ese* suceso trascendental, se vio sujeto a una *expansión exponencial* –denominada *inflación*- similar a la presencia de una descomunal constante cosmológica Λ_{infl} (Penrose, 2017: 375).

Penrose reconoce que la teoría inflacionaria cuenta con un grado de aceptación por la comunidad científica ciertamente alto (Penrose, 2017: 375). A pesar de ello, nuestro autor cree que esta teoría entra en conflicto con principios y leyes demasiado fundamentales como para que sea aceptada sin ambages. De hecho, la rechaza por el modo en el que esta contradice, precisamente, la segunda ley de la termodinámica.

La contradicción es fácil de localizar. Si suponemos que el universo en su inicio sufrió una *expansión exponencial*, quiere decir (muy resumidamente) que alcanzó el grado de máxima entropía en *aquel* entonces (¡contradiciendo al aumento de la entropía de la segunda ley!). ¿Cómo una teoría que

[228] En cuanto que el determinismo no se ve envuelto en contradicción alguna, llegando incluso a ser casi indiferente para dicha ley, como puede verse en la cita anterior.

[229] Aquí no veremos la corriente inflacionaria, ya que ello nos alejaría innecesariamente del asunto que nos ocupa. De todos modos, si se quiere profundizar en el análisis que Penrose hace de esta véase (Penrose, 2006: 1002-1016) y, sobre todo, (Penrose, 2017: 374-394).

[230] Para la definición, nuestro autor se sirve de los trabajos de Alan Guth y Alexéi Starobinski.

contradice de un modo tan flagrante la segunda ley cuenta con el crédito de la comunidad científica? En primer lugar, la segunda ley, como hemos visto, no resulta tan clara para explicar ciertos fenómenos de la naturaleza (algunos tan fundamentales como la vida de los seres vivos en nuestro planeta). En segundo lugar, la teoría inflacionaria responde satisfactoriamente a tres problemas cosmológicos muy desconcertantes (estos son denominados *del horizonte*, *de la suavidad* y *de la planitud*).

Para Penrose, en cambio, los inflacionistas en el fondo no resuelven debidamente ninguno de los tres problemas. Aunque les concede cierto crédito en lo que respecta a los problemas *del horizonte* y *de la planitud*, en referencia al *problema de la suavidad* entiende que este queda sin resolverse de ninguna de las maneras. Como vimos más arriba, la segunda ley en principio tampoco *encaja* según postula la teoría inflacionaria. Esto, sin embargo, no es obstáculo para los inflacionistas, ya que para responder a la contradicción con respecto a la segunda ley estos frecuentemente apelan al principio antrópico. Y he aquí el motivo por el que Penrose siente cierto recelo para con la concepción de este principio. Su postura contraria se basa en que considera que los inflacionistas no manejan una concepción adecuada del principio antrópico, en el sentido de que estos otorgan a dicho principio un alcance mayor del que le pertenece:

> En mi opinión, resulta perturbador constatar con qué frecuencia los físicos teóricos acaban recurriendo a tales argumentos [principio antrópico *fuerte*] para compensar la falta de capacidad predictiva de que adolecen sus diversas teorías […] Mi opinión es que esta es una situación muy triste e inútil para una teoría (Penrose, 2017: 409).

Como vimos en el apartado anterior, la idea de un principio antrópico *débil* no desagrada a Penrose (aspecto que permanece incluso en sus trabajos más actuales, como MFF). Otro asunto es que se conciba al principio antrópico (sea en la forma que sea) como la última respuesta.

Sin lugar a dudas, el principio antrópico [*débil*] y la segunda ley podrían servir de apoyo al determinismo de Penrose, pero él mismo sabe que el modo en el que están desarrolladas estas ideas hoy en día, dicho soporte se da *parcialmente*.

3. Neurociencia y especulación

3.1. Algunos de los rasgos del cerebro y su relación con la consciencia

> *[…] Y, finalmente, cuando el* alma razonable *se halle en esta máquina tendrá su sede principal en el cerebro y allí desempeñará la misma función que el fontanero que tiene que estar en los respiraderos adónde van a parar todos los tubos de esas máquinas, cuando quiere estimular o impedir o cambiar de alguna manera sus movimientos (Descartes, 2011: 684).*

Hasta ahora hemos visto que Penrose trata el debate de la computabilidad o no-computabilidad de la mente y la consciencia humana desde la ciencia, concretamente desde la física y la matemática. También es digno de mención que otorga a la filosofía un papel importante en este debate. Pero tratándose de la consciencia, ¿no es conveniente tener en cuenta, al menos de forma panorámica, la perspectiva de la neurociencia? El mismo Penrose defiende que sí, a pesar de que reconoce que ello puede acarrearle algunos problemas (ya que no es un campo de su dominio[231]).

Una vez decide entrar en materia de neurociencia, Penrose se propone ofrecer un esquema cerebral no muy amplio, pero sí lo suficientemente completo como para que podamos hacernos una idea del funcionamiento general del cerebro.

La primera de las características a destacar[232] del cerebro es la división que generalmente se hace de este. El cerebro observado desde arriba se puede ver muy claramente dividido en dos mitades, cada una de las cuales se conoce como *hemisferios* (izquierdo y derecho). De un modo menos claro, también se compone de cuatro lóbulos, uno en la zona delantera (lóbulo *frontal*) y los otros tres (lóbulos *parietal*, *temporal* y *occipital*) en la zona trasera. Otra parte importante la encontramos también en la zona trasera, el *cerebelo*, el cual posee un tamaño pequeño y una forma tímidamente esférica. Y luego otras partes aún menos evidentes, pero que son generalmente muy tenidas en cuenta, tenemos la médula espinal, el puente, el acueducto de Silvio (que está situado posteriormente al puente), el tálamo, el hipotálamo, hipocampo, el cuerpo calloso y el mesencéfalo (Penrose, 1991: 463).

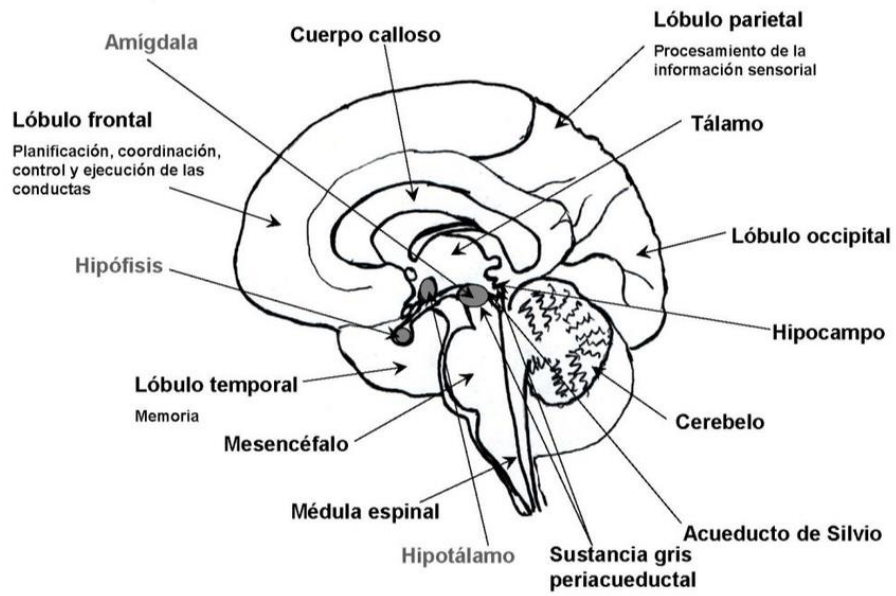

Figura 14. Mapa general del cerebro humano

El cerebro y el cerebelo tienen capas externas (*sustancia gris*) e internas (*sustancia blanca*). La *sustancia gris* que compone la capa externa en ambas

[231] Si lo comparamos, claro está, con campos como la matemática o la física.

[232] Como podrán verse en las citas, todas aquellas características que veremos son las destacadas por Penrose (sobre todo las expuestas en NME).

partes se denomina, *corteza cerebral* y *corteza cerebelar* (Penrose, 1991: 464).

Con el esquema del cerebro suficientemente claro, Penrose pasa a analizar, también de un modo general, las acciones que corresponden a cada parte. De la corteza cerebral, por ejemplo, destaca que algunas de sus partes son fundamentales para múltiples funciones, llegando a ser en ocasiones, al menos, llamativas. Uno de esto casos es el *córtex visual*, el cual se encuentra en el interior del lóbulo occipital (recordemos que es uno de los tres que estaban situados en la zona trasera del cerebro). Tal y como puede deducirse de su nombre, se entiende que del *córtex visual* depende la recepción y la interpretación de la visión. Lo llamativo para Penrose (aunque no creo que nuestro autor sea una excepción en esta consideración) es que la visión dependa de una región que está situada en la parte trasera del cerebro, cuando los ojos están en la zona delantera (Penrose, 1991: 464). Sin embargo, esto no constituye la única *rareza* en el modo de proceder de las partes del cerebro. De hecho, en lo que se refiere a la capacidad motora, los hemisferios del cerebro son encargados de su lado opuesto (el hemisferio *izquierdo* está relacionado con la parte *derecha* del cuerpo, mientras que el *derecho* de la *izquierda*). El modo en el que el cerebro procesa el sonido (desde el lóbulo *temporal*) también sigue este mismo patrón. Por su parte, la facultad olfativa (la cual se sitúa en el lóbulo *frontal*) no cumple dicha pauta (es decir, que las acciones del lado *derecho* proceden del *derecho* y las del *izquierdo* con del *izquierdo*) (Penrose, 1991: 465). Pero con el tacto, volvemos a encontrarnos con el mismo cruce anterior. Las sensaciones táctiles están relacionadas con el *córtex somatosensorial*, situado en el lóbulo *parietal*, más concretamente detrás de la división entre los lóbulos *frontal* y *parietal* (Penrose, 1991: 466).

Todas las partes vistas hasta ahora, estas son los córtex visual, auditivo, olfativo, somatosensorial y motor, pertenecen a lo que Penrose denomina como región *primaria* (Penrose, 1991: 466). La *secundaria* es aquella en la que las distintas acciones de los córtex se relacionan entre sí, con la excepción de nuevo del córtex olfativo, el cual es marcadamente diferente y del que se tienen pocas certezas (Penrose, 1991: 467). Y por último tenemos la región *terciaria*, también denominada como córtex de *asociación*. En esta región también se relacionan los diferentes córtex, pero de un modo muy sutil y complejo. Tanto es así que es en ella donde se considera que se desarrolla la memoria, donde se construyen las imágenes del mundo, permite la evaluación de planes y donde se entiende y formula el habla (Penrose, 1991: 467). Tenemos, entonces, que la región *terciaria* es esencial con respecto a aspectos que nos caracteriza como humanos. El habla, sin duda, es uno de tales aspectos. Las áreas principales del cerebro que se relacionan con el habla son las de Broca y Wernicke[233].

[233] Estos nombres se deben, de manera obvia, a sus descubridores, el médico francés Paul Pierre Broca (1824-1880) y el neurólogo alemán Karl Wernicke (1848-1905).

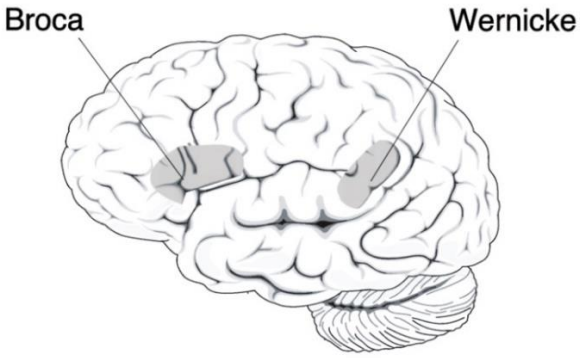

Figura 15. Puntos del cerebro en los que se sitúan las áreas de Broca y Wernicke

Al relacionarse el habla con el lado izquierdo del cerebro, tanto el área de Broca como el de Wernicke están situados en dicha parte. La función que se le asigna al área de Broca es la de formar el habla, mientras que a la de Wernicke se le otorga la función de comprender el habla (Penrose, 1991: 469).

Dejando a un lado las actividades procedentes de los lóbulos, pasemos a ver la principal del cerebelo. Dicha función es, cuanto menos, peculiar. Cuando los seres humanos estamos en fase de aprendizaje (en el plano motor, como por ejemplo al andar, conducir, bailar, etc.), la actividad depende del cerebro. Pero una vez los movimientos pueden realizarse de un modo automático, estos pasan a depender directamente del cerebelo. Es más, si intentamos añadir aspectos diferentes a lo aprendido (intentar caminar diferente, aprender nuevos pases de baile…) el cerebro vuelve a activarse, mientras que la acción del cerebelo pasa a ser menos fluida (Penrose, 1991: 470). Cabe destacar que el cerebelo no presenta la misma naturaleza que el cerebro, en el sentido de que el lado derecho controla el lado derecho y el izquierdo el izquierdo (Penrose, 1991: 471).

Con respecto a las partes anteriormente nombradas, debemos saber que el *hipocampo*, por ejemplo, tiene un papel fundamental en el plano de la memoria. El *hipotálamo*, por su parte, es donde se dan las emociones (siendo responsable de las manifestaciones, tanto físicas como mentales, de las emociones). Y como partes importantes de conducción entre las distintas partes del cerebro tenemos el *cuerpo calloso*, el *tálamo* y el *mesencéfalo* (Penrose, 1991: 471).

Suponiendo que la estructura del cerebro tiene una forma tal, no es difícil comprender por qué los(as) partidarios(as) del punto de vista A encuentren posible que la actividad del cerebro, y con ello todo de lo que de ella surge (por ejemplo la consciencia), pueda ser reproducida en términos algorítmicos. Esto no significa que quienes defienden el punto de vista A conciban la actividad cerebral del modo simplificado en el que la hemos visto descrita. Por supuesto que saben de lo complejo del asunto y también de lo necesario que es seguir profundizando en los estudios del cerebro para saber de un modo más preciso el funcionamiento de este.

Penrose está de acuerdo en casi todo lo descrito. La salvedad, como ya bien se sabe, reside en que la actividad del cerebro puede ser traducida por procedimientos algorítmicos. Nuestro autor reconoce que existen numerosas capacidades del cerebro que son susceptibles de ser computables, pero en

ninguno de los casos la consciencia puede ser una de ellas. La defensa de Penrose es que debe haber un procedimiento no-algorítmico que permita explicar el fenómeno de la consciencia. Pero, ¿en qué parte del cerebro residiría tal proceso no-algorítmico? Si bien en NME no podemos encontrar una respuesta clara a tan problemática pregunta, en SM Penrose, siguiendo los estudios de Stuart Hameroff, responde con una teoría propia (que compartiría con el mismo Hameroff). Dicha teoría sitúa al proceso no-algorítmico que anda buscando nuestro autor en una parte muy concreta del cerebro: los microtúbulos.

3.2. La relación los microtúbulos y la consciencia

Ya sabemos que el punto de vista C de Penrose comparte con el punto de vista A que la respuesta al debate de la consciencia debe ser respondida en base a criterios científicos. Su postura no se limita a este aspecto, sino también considera que el fenómeno de la consciencia se da en el *material interno* del ser humano (de ahí que le resulte imprescindible el estudio del cerebro humano). Es decir, todo aquello que suponga a la consciencia surgiendo a partir de algo que esté *fuera* de ese *material interno* humano (Penrose toma el ejemplo de la postura dualista) carece de sentido para nuestro autor:

> […] Tener una «materia mental» externa que no está sometida a leyes físicas es salirnos de algo que podría llamarse razonablemente una explicación científica, y es recurrir al punto de vista D (Penrose, 2012: 370).

Si la ciencia resulta tan decisiva para este debate, ¿por qué no consigue dar una respuesta más contundente a la postura dualista, la cual realmente no se ha visto amenazada ni por quienes defienden C o A? Ciertamente es complicado darles una contrarréplica contundente a los dualistas (de hecho, Penrose lo reconoce). Sin embargo, nuestro autor piensa que si dicha contrarréplica no surge es porque la física[234] sobre la que se asientan los principios de los estudios acerca de la mente y el cuerpo humano no *puede* responder a ello.

Penrose está en desacuerdo con la idea de que la acción del cerebro se tenga que explicar en base a la física clásica. Para nuestro autor es fundamental que se tenga en cuenta la acción cuántica de la física moderna para el estudio del cerebro. Esto no se da de dicha forma, ya que los biólogos siguen creyendo que no es necesario salirnos del marco clásico (Penrose, 2012: 368). Esto es casi un absurdo, porque se conoce muy certeramente que las fuerzas químicas tienen una importancia capital en las acciones del cerebro, y dichas fuerzas químicas tienen un origen innegablemente mecano-cuántico (Penrose, 2012: 368). No obstante, Penrose reconoce que es necesario encontrar efectos de la acción cuántica que sean muy significativos, que hagan evidente que tal acción tiene lugar en nuestra biología. Para ello es imprescindible que se dé una coherencia cuántica a gran escala. Pero, ¿qué es la coherencia cuántica? La respuesta de nuestro autor es la siguiente:

[234] Dicha física es la clásica.

> […] Este fenómeno se refiere a circunstancias en que grandes números de partículas pueden cooperar colectivamente en un simple estado cuántico que permanece esencialmente no enmarañado con su entorno. (La palabra «coherencia» se refiere, en general, al hecho de que las oscilaciones en lugares diferentes varían al unísono. Aquí con coherencia *cuántica* estamos interesados en la naturaleza oscilatoria de la función de onda, y la coherencia se refiere al hecho de que estamos tratando con un simple estado cuántico). Semejantes estados tienen lugar muy espectacularmente en los fenómenos de superconductividad (donde la resistencia eléctrica cae a cero) y superfluidez (donde la fricción del fluido, o viscosidad, cae a cero). El ingrediente característico de tales fenómenos es la existencia de un *intervalo de energía* que tiene que ser superado por el entorno para llegar a perturbar este estado cuántico. Si la temperatura en dicho entorno es demasiado alta, de modo que la energía de muchas partículas es suficientemente grande para que superen este intervalo y se enmarañen con el estado, entonces la coherencia cuántica se destruye (Penrose, 2012: 371).

La explicación deja entrever un problema fundamental y que el mismo Penrose se apresura en apostillar (Penrose, 2012: 371), este es, las condiciones *especiales* de la superconductividad y de la superfluidez. Estas condiciones especiales se traducen en que debido a las temperaturas a las que se dan estos fenómenos se hace difícil la posible localización de la coherencia cuántica dentro del cerebro. Para Penrose, esto no supone un contratiempo definitivo, ya que existen estudios en los que este problema se ha visto, al menos, medianamente resuelto (Penrose, 2012: 372). La búsqueda de la acción cuántica debe continuar.

Este intento de introducir la acción cuántica por parte de Penrose no es nueva ni exclusiva de SM, ya que en NME clamaba por la misma idea, pero él mismo no quedó convencido con sus propios argumentos[235].

Para su nueva propuesta, la cual se denomina ***RO** orquestada*[236], encuentra oportuno analizar la naturaleza biológica de las células del cerebro. A pesar de que dedica una explicación digna de la importancia y la forma de las neuronas (Penrose, 2012: 372-375), Penrose sostiene que aquello en lo que debe centrarse la atención es en el *citoesqueleto*, ya que considera que el quehacer de ello es más fundamental que el de las neuronas.

¿Qué es el citoesqueleto y qué lo hace tan importante? Según la definición de Rubén Herce, «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» (Herce, 2014: 177). Es decir, podemos entenderlo como la base fundamental de las células. Para Penrose la función del citoesqueleto que verdaderamente le interesa es aquella en la que este se comporta como «sistema nervioso» de la célula (Penrose, 2012: 377). Esto es así, porque nuestro autor sigue

[235] El motivo más destacable de este desacuerdo propio es el siguiente: La propiedad de perturbar el entorno de la neurona que se dispara es la característica que me ha parecido siempre más incómoda para el tipo tosco de propuesta que yo había defendido previamente en NME, en la que la superposición cuántica del disparo y el no disparo simultáneo de familias de neuronas parece ser realmente necesaria (Penrose, 2012: 375).

[236] Con respecto a esto hay que tener varios aspectos en cuenta. En primer lugar, esta propuesta Penrose la comparte con Hameroff. En segundo lugar, debe su nombre al procedimiento no-algorítmico que buscan para explicar el fenómeno de la consciencia, al cual llaman **RO** (que significa *reducción objetiva*). Y en tercer lugar, la propuesta ha tenido su evolución (aunque sin cambiar mucho en esencia).

[principalmente] a Hameroff, uno de los grandes defensores de que el citoesqueleto desempeña dicha función.

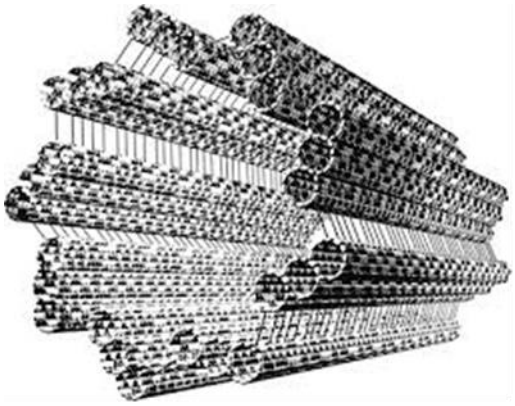

Figura 16. Representación del esquema del citoesqueleto de las neuronas

El citoesqueleto está compuesto de filamentos de actina, microtúbulos y filamentos intermedios (Penrose, 2012: 378). Si bien todos estos componentes son importantes, aquellos que le interesan en particular a Penrose son los *microtúbulos*[237]. Aquello de lo que se están compuesto los microtúbulos es de tubos cilíndricos huecos, con 25 nm[238] de diámetro exterior y 14 nm de diámetro interior de manera aproximada (Penrose, 2012: 378). Tales tubos se organizan a veces en fibras mayores de tipo tubo formados, a su vez, por nueve dobletes o tripletes (e incluso tripletes parciales) de microtúbulos, dispuestos de tal manera que se asemeja a un abanico (habiendo a veces un par de microtúbulos en el centro) (Penrose, 2012: 378). Los microtúbulos están formados por subunidades llamadas *tubulinas*, las cuales, al mismo tiempo, constan de dos partes (por lo que son «dímeros»), denominadas α-tubulina y β-tubulina. Estas dos partes pueden darse, como mínimo, en dos configuraciones geométricas diferentes, conocidas como *conformaciones* diferentes (Penrose, 2012: 379).

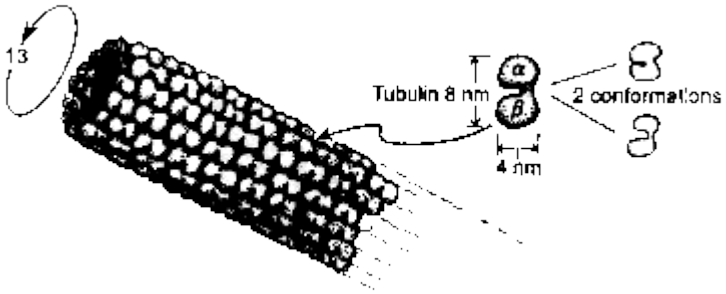

Figura 17. Representación de un microtúbulo con sus tubulinas α y β y dos de sus posibles conformaciones

[237] Aunque veamos algunas de las características de los microtúbulos, considero que es conveniente no entrar en demasiados detalles de la descripción de estos, porque ello realmente no aportaría en exceso a lo que aquí pretendo exponer. Para la descripción detallada véase (Penrose, 2012: 378-388).

[238] Equivale a la milmillonésima parte de un metro.

Una de las características más importantes de las conformaciones de las partes de la tubulina es que pueden cambiar de un estado a otro según las alternativas para sus polarizaciones eléctricas, estando las partes influidas por los estados de polarización de cada uno de sus seis vecinos, ajustándose así reglas de la conformación de las partes y sus vecinos[239] (Penrose, 2012: 384). Esta característica en particular es importante porque otorga a los microtúbulos una capacidad de interconexión fundamental[240]. Penrose defiende, incluso, que ella tiene una importancia capital en la influencia de los microtúbulos en las neuronas, aunque esto pertenece a un plano especulativo. Lo cierto es que una vez nuestro autor termina de describir los microtúbulos y aquello que se sabe de estos, su postura se asienta cada vez más en dicho plano especulativo. Por otro lado, Penrose no vacila en reconocer esta situación, aunque ello no le lleva a caer en el pesimismo con respecto a su postura, sino más bien todo lo contrario:

> […] una cosa parece clara en mi opinión. Hay pocas posibilidades de que una exposición enteramente clásica del citoesqueleto pudiera explicar adecuadamente su comportamiento. Esta es una situación diferente de la de las propias neuronas, para las que las discusiones en términos enteramente clásicos parecen ser básicamente apropiadas. De hecho, un examen de la literatura actual sobre la acción del citoesqueleto revela el hecho de que continuamente se está apelando conceptos mecano-cuánticos, y tengo pocas dudas de que esto aumentará en el futuro (Penrose, 2012: 389-390).

Aunque por ahora sólo se puedan hacer conjeturas, Penrose sostiene que existen evidencias de que los microtúbulos tienen un papel esencial para que la consciencia tenga lugar. Nuestro autor trae a colación el caso de la utilización de anestesias en organismos y cómo estos afectan directamente al citoesqueleto, provocando que la función de los microtúbulos quede interrumpida. Esto podría no tener importancia si la consciencia no se viese afectada en forma alguna, ¡pero coincide con la pérdida de esta! Si bien este argumento está muy lejos de ser definitivo, sí que es ciertamente lícito tenerlo en cuenta[241].

Sin lugar a dudas, la propuesta ***RO** orquestada* de Penrose y Hameroff tiene grietas en su esquema[242]. La primera de ellas es que el procedimiento no-algorítmico RO tiene que obedecer a las leyes cuánticas, lo cual supone un conflicto si tenemos en cuenta la decoherencia cuántica[243] (Herce, 2014: 183-

[239] Si bien la estructura de la propuesta de Penrose y Hameroff ha cambiado desde el modelo que he decidido seguir para explicar este asunto (el de SM), lo cierto es que estos cambios no repercuten de forma drástica en las implicaciones de aquello que defienden desde el principio. Estos cambios responden, más bien, al intento de una descripción más precisa de dicha estructura.

[240] Esta es una idea que Penrose toma de los estudios de Djuro Koruga y Hameroff.

[241] Penrose señala que este caso no es exclusivo en los seres humanos, sino en todo organismo que contiene microtúbulos (nuestro autor nombra los ejemplos de las amebas y los paramecios), lo cual puede hacer plantear la cuestión de si estos también tienen algún tipo de consciencia. No obstante, Penrose no decide no extenderse demasiado sobre dicho asunto, ya que cree que es un debate en el que poco se puede decir realmente (Penrose, 2012: 391).

[242] Las siguientes «grietas» que veremos son, en gran medida (como podremos ver en las citas), las señaladas por Rubén Herce (2014), porque estas constituyen, ciertamente, una observación elocuente y actual de la propuesta ***RO** orquestada*.

[243] En mecánica cuántica se conoce como decoherencia cuántica a la capacidad de transición de un estado cuántico a un estado clásico (John Gamble, 2008: vi).

184). Este aspecto es perfectamente concebible científicamente hablando, pero ello obliga, de forma casi necesaria, a seguir en el plano de la conjetura.

La segunda (que es consecuencia de lo que implica la primera) consiste en que no tendríamos la oportunidad de observar el fenómeno de la consciencia, ya que el procedimiento **RO**, al ser cuántico, haría tremendamente complicada su experimentación en un nivel macroscópico (Herce, 2014: 184). En mi opinión, este aspecto realmente no me perece tan definitivo como dificultad de la propuesta de Penrose y Hameroff. Si bien los procedimientos cuánticos en el plano macroscópico tienen una traducción –digámoslo así- compleja, no menos cierto es que sus efectos pueden acabar siendo observados.

La tercera tiene que ver con la composición biológica que requieren los cerebros de los seres humanos, la cual, en consideración de Penrose, no se diferenciaría (en algunos casos) de otras ciertas especies (Herce, 2014: 184). Ello, como vimos más arriba, plantea la difícil cuestión de si tales especies podrían tener cierto tipo de consciencia. Por lo que la propuesta más que aclarar sugiere asuntos de igual o más dificultad.

Si bien las contrariedades que presenta la propuesta de Penrose y Hameroff son para tener en cuenta, también es de rigor que Penrose matiza en numerosas ocasiones que esta propuesta es sólo el punto de partida y nunca una respuesta definitiva:

> En la visión que estoy proponiendo provisionalmente, la consciencia sería alguna manifestación de este estado citoesquelético interno cuánticamente enmarañado y de su implicación en el juego (**RO**) entre los niveles cuántico y clásico de actividad. El sistema de neuronas de tipo ordenador clásicamente interconectadas se vería continuamente influido por esta actividad citoesquelética, como manifestación de lo que conocemos como «libre albedrío», cualquier cosa que sea. El papel de las neuronas, en esta imagen, se parece más quizá a un *dispositivo amplificador* en el que la acción citoesquelética a menor escala se transfiere a algo que puede influir en otros órganos del cuerpo –tales como los músculos. En consecuencia, el nivel neuronal de descripción que proporciona la imagen actualmente vigente del cerebro y la mente es una mera *sombra* del nivel más profundo de acción citoesquelética -¡y es en este nivel más profundo donde debemos buscar las bases físicas de la *mente*! (Penrose, 2012: 398).

Otra parte importante de la propuesta[244] es el papel que juega el tiempo en la consciencia. Para comprender de un modo adecuado dicho papel, Penrose considera que es absolutamente necesario tener una concepción de la física acertada. Veamos en el punto que sigue cómo nuestro autor lo plantea.

3.3. La consciencia y el tiempo

> *[…] No pretendo saber qué cosa es el tiempo (ni siquiera si es una «cosa») pero adivino que el curso del tiempo y el tiempo son un solo misterio y no dos (Borges, 1974: 648)*

[244] Si bien en la parte neurocientífica Penrose se apoya en Hameroff, en esta es nuestro autor quien hace la mayor contribución filosófica.

Imagínese que se tumba en un prado un día soleado con nubes pasajeras que interrumpen el azul del cielo. Simplemente mire cómo las formas de esas nubes cambian y parecen adoptar figuras que le resultan familiares (montañas, animales, las olas del mar…). Esta es una actividad que, sin duda, hemos podido experimentar a lo largo de nuestra vida (podríamos cambiar el prado por la playa o cualquier otro lugar). A pesar de que el acto descrito en sí mismo pueda parecer carente de importancia, por otro lado no podremos negar que dicho ejercicio se realiza (y he aquí el *quid* del asunto) *conscientemente*. El incesante paso de las nubes y [si nos quedamos en la misma posición un largo rato] el cambio de claridad del cielo nos advierte de que el tiempo pasa. La del tiempo, sin duda, es una de las pistas más significativas que tenemos a la hora de detectar la presencia de la consciencia, sobre todo si este tiempo tiene coherencia (es decir, que no existen saltos).

Pero, ¿de veras ese fluir del tiempo es tan definitivo? Penrose no está tan seguro. Por supuesto que admite que el papel del tiempo es innegablemente importante, pero que, por norma general, entendamos de un modo inadecuado tal papel también es evidente para él. La razón por la que tenemos una concepción errónea del tiempo es que en este asunto concreto pasamos por alto que este no puede ser no puede ser separado del espacio (¡tal y como nos dice la relatividad!). Cuando observamos el cielo y las nubes, el tiempo parece «fluir», pero, sin embargo, no podemos decir lo mismo del espacio. ¿Nos engaña, entonces, la consciencia? El mismo Penrose reconoce que la relación entre la consciencia y el tiempo es, cuanto menos, extraña:

> En realidad, es *únicamente* el fenómeno de la consciencia el que nos exige pensar en términos de un tiempo que «fluye». Según la relatividad, uno tiene simplemente un espacio-tiempo tetradimensional «estático», sin nada que «fluya» en él. El espacio-tiempo está precisamente *allí* y el tiempo no fluye más de lo que lo hace el espacio. Es sólo la consciencia la que parece necesitar que el tiempo fluya, de modo que no deberíamos sorprendernos si la relación entre consciencia y tiempo es también extraña en otros aspectos (Penrose, 2012: 406).

Tanto en NME como en SM, Penrose cita dos experimentos en los que se intentaba dar respuesta al tiempo que tarda el ser humano en realizar un acto de libre arbitrio (que corresponde a la parte activa de la consciencia). El resultado de tales experimentos[245] fue que el acto libre [consciente] tarda un segundo (e incluso algo más) en concretarse. Las conclusiones a las que pueden llegarse son, al menos, tres. La primera es que la consciencia, en tanto que acto voluntario en sí, no existe, ya que dependería de un mandato inconsciente previamente establecido. La segunda es que puede existir una capacidad de elegir en el último instante (dándose en ese segundo -y algo más- que necesita la consciencia), cambiando tal mandato previo, aunque esto sería posible sólo en ocasiones y no sería de manera constante. Y la tercera es que el sujeto percibe de forma errónea su propio acto consciente, que se daría antes de ese segundo –y algo más (Penrose, 2012: 407-408). Sea cual sea la conclusión que se escoja, lo cierto es que el papel de la consciencia queda en una posición perjudicada. Es por ello mismo por lo que a Penrose no le convence ninguna de ellas. Según nuestro autor, tal y como vimos más

[245] Uno es el llevado a cabo por Kornhuber y sus colaboradores, y el otro el realizado por Libet y sus colaboradores. Para los detalles de estos experimentos véase (Penrose, 1991: 545-548) y (Penrose, 2012: 407-408).

arriba, el error reside en la concepción manejada del tiempo en estos experimentos, que es al fin y al cabo la que generalmente se concibe:

> […] ¿*Verdaderamente* existe un «tiempo real» en el que tiene lugar una experiencia consciente, y tal que el «instante de experiencia» particular debe preceder al instante de cualquier efecto de una «respuesta voluntaria» a dicha experiencia? En vista de la relación anómala que mantiene la consciencia con la propia noción de tiempo […], me parece que es cuando menos posible que *no* exista semejante «tiempo» claro y preciso en el que deba ocurrir un suceso consciente (Penrose, 2012: 409).

¿Cuál es la solución que plantea Penrose (si es que ofrece alguna)? Nuestro autor acude al planteamiento de su propuesta. Recordemos que la ***RO orquestada*** clama por la búsqueda de un procedimiento no-algorítmico (que respondería a los principios de la física cuántica), el cual daría con una explicación más certera acerca de la consciencia. Si la consciencia explicada en términos de física clásica no otorga una respuesta satisfactoria en su relación con el tiempo, es casi forzoso concluir que la acción cuántica necesita entrar en la escena (Penrose, 2012: 410). Penrose admite que todo esto responde de un modo mayoritario a sus convicciones propias que a evidencias científicas. Sin embargo, también se encarga de aclarar que tales convicciones no son ciegas, sino que se sustentan en algo más concreto que la simple fe en esas convicciones.

4. La «realidad» de Penrose

4.1. Realidad, ciencia y metafísica (los tres mundos)

> *[…] El realista elige así* su *realidad en la realidad (Bachelard, 1978: 33)*

Un tema fundamental para Penrose, el cual se hace extensivo tanto a su faceta científica como filosófica, es la realidad. Con respecto a cómo trata Penrose la realidad, considero, le sucede lo mismo que con el platonismo (tema con el que también está estrechamente relacionado), en el sentido de que nuestro autor no aborda la realidad desde el *tecnicismo* filosófico, lo cual generalmente es considerado un error. Esta, sin embargo, no es mi opinión. Sin duda, Penrose aporta una perspectiva, cuanto menos, interesante y particular acerca de la realidad.

En primer lugar, Penrose suele ser claro a la hora de considerarse a sí mismo como realista (Penrose, 1991: 571[246]), (Penrose, 1991: 377). Nuestro autor considera que existe una realidad *ahí fuera* y que es susceptible de ser conocida. El modo de conocer dicha realidad sería a través de juicios. Pero entendamos esto correctamente. Cuando Penrose habla de «juicios» no se refiere a ningún tipo de «creación», sino de *valorar* hasta qué punto lo que se nos *presenta* se *corresponde* con la realidad. Es decir, que la última palabra siempre dependería de esa realidad en sí y en ninguno de los casos de tales juicios. Por otra parte, el papel activo del ser humano es imprescindible para

[246] La página corresponde a las notas de la edición citada.

que la realidad sea explicada[247]. Por este motivo es conveniente encontrar los juicios que *más* nos garanticen ese *acercamiento* a la realidad.

Ahora bien, ¿cuál (o cuáles) campo(s) de conocimiento pueden avalar este tipo de juicios? La respuesta de Penrose es la esperada: las matemáticas y la física. No obstante, esto no significa que nuestro autor restrinja tal capacidad a estos dos campos de manera exclusiva. Su respuesta obviamente está motivada porque él es especialista en estos terrenos. A pesar de que entienda que tanto las matemáticas como la física garantizan un grado de certeza incuestionable, ello no le lleva al error de negárselo a cualquier otro campo.

Habiendo dejado claro este último aspecto, pasemos a ver el modo en el que Penrose entiende que las matemáticas y la física deben aportar los juicios que nos acerquen a la realidad. Para nuestro autor el punto de partida es doble: las teorías aceptadas por la comunidad científica y los resultados de los experimentos. Esto puede parecer muy esquemático y hermético, pero en realidad lo plantea en unos términos más laxos:

> […] 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 (Herce, 2014: 42).

La realidad, defiende Penrose, si bien *está ahí* no por ello debemos concebirla de manera estática. Por tanto, el modo de llegar a ella tampoco puede serlo. El rasgo dinámico de la realidad penroseana queda explicado en la concepción metafísica que elabora nuestro autor, esta es, la de los tres mundos[248].

Penrose toma la idea de los tres mundos a partir de la teoría popperiana, aunque nuestro autor insiste en que los enfoques son diferentes[249] (Penrose, 2012: 433). La teoría consiste, como su propio nombre indica, que existen tres mundos independientes con diferentes contenidos, pero que se interrelacionan entre sí. El esquema de los mundos se entiende de la siguiente manera (Penrose, 2012: 434):

- *Mundo mental*: este mundo es el que conocemos más directamente, ya que está constituido por nuestras percepciones conscientes. No obstante, el mundo mental es el menos accesible a la ciencia (al menos, por ahora). El contenido de este mundo está formado por ideas (ya sean sentimientos como el dolor, memorias, como ideas de objetos del mundo físico).

[247] En este punto volvemos al –digámoslo así- *especial* rasgo común que Penrose mantiene con los intuicionistas en el plano de las matemáticas, con los cuales compartía la concepción de la importancia creadora del ser humano, pero difiriendo en que esa creación surgiera plenamente de nuestro intelecto. Véase (Cap. 2, § 2.2).

[248] En este apartado nos centraremos en ver las características principales que Penrose describe en SM.

[249] Aunque acude a ideas de Platón y Berkeley (también con sus pertinentes diferencias), el esquema es claramente más afín al de Popper.

- *Mundo físico*: tal y como su nombre indica, este mundo contiene todo aquello que es físico (sillas, mesas, cerebros, átomos…). Para Penrose, el mundo mental es más directo que el mundo físico, aunque admite que este último cada vez nos es menos ajeno, gracias a lo que la ciencia nos dice de él.

- *Mundo matemático platónico*: en este mundo encontraríamos las ideas matemáticas, en el sentido más amplio. Es decir, en él se encuentran tanto las matemáticas que conocemos (los números naturales y las operaciones que con ellos podemos realizar); como las matemáticas a las que no tenemos acceso directo (el número pi íntegramente, las soluciones a los problemas matemáticos sin resolver e incluso aquellos que aún no han sido planteados). Penrose reconoce que este mundo resulta difícil de aceptar por muchos (Penrose, 2012: 434), pero ello no impide que nuestro autor le dé una gran importancia. Es precisamente la creencia en este mundo por la que Penrose se considera a sí mismo platónico. Pero antes de profundizar en este asunto veamos cómo se da la interrelación entre los tres mundos.

La manera en la que los tres mundos se interrelacionan es *emergiendo* los unos de los otros, aunque no habiendo un mundo originario, ya que esta interrelación sería cíclica (por lo que ver un posible principio o fin resultaría una tarea inútil). En detalle, el mundo físico emergería de una parte del mundo matemático platónico, mientras que este último lo haría del mental que, a su vez, lo haría del físico. Lejos de pensar que el esquema que ofrece es claro, Penrose reconoce la dificultad de concebirlo y es por ello por lo que considera que la interrelación es *misteriosa*[250] (Penrose, 2012: 435).

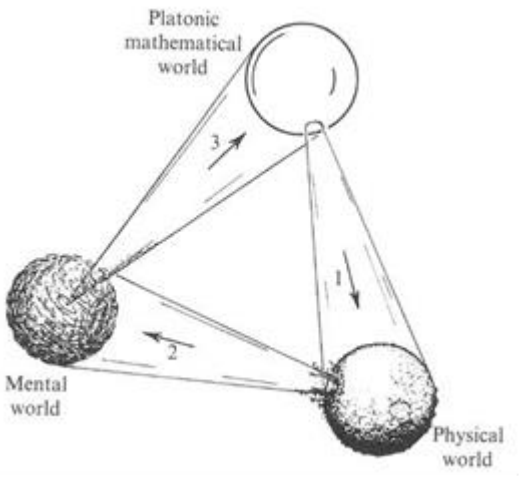

Figura 18. Representación de los tres mundos de Penrose

Penrose no entra en demasiados detalles para explicar todos los misterios. De hecho, sólo lo hace con aquellos que involucran al mundo matemático platónico que es, como vimos más arriba, el de más difícil aceptación.

[250] Este aspecto ha sido duramente criticado, porque Penrose no ofrece muchas más explicaciones que las mencionadas, lo cual hace que el esquema se tambalee inevitablemente. Para una crítica de este corte véase (Badía, 2008).

Teniendo en cuenta este último aspecto y que Penrose es matemático, hace más comprensible que añada dicha aclaración.

No cabe duda que entender la realidad como compuesta por más de un mundo es difícilmente explicable, y más si dicha idea es presentada en forma de bosquejo. Por otro lado, pienso que es inevitable caer en explicaciones *insuficientes* cuando intentamos abordar la realidad, por lo que señalar a Penrose por ello me parece injusto. Es cierto que se le puede exigir una exposición algo más extensa, pero también hay que entender que, al fin y al cabo, está hablando en términos metafísicos, terreno con el que no está familiarizado. Pero dejemos a un lado las diferentes valoraciones que puedan hacerse de las ideas de Penrose y pasemos a ver la explicación de los misterios que involucran al mundo matemático platónico.

Nuestro autor piensa que si es tan complicado para muchos la concepción del mundo matemático platónico es, precisamente, porque no conocen de un modo adecuado el alcance de las matemáticas y la influencia que estas tienen en el mundo físico. Para que dicho alcance e influencia queden manifiestos, Penrose recurre a las teorías físicas, en concreto a la de la relatividad de Einstein y del esquema newtoniano. Según nuestro autor, las teorías físicas son una prueba satisfactoria tanto para ver la relación de las matemáticas con el mundo físico (ya que las teorías físicas tienen un marcado fundamento matemático) como para atisbar su alcance (el grado explicativo que tienen):

> […] La teoría gravitatoria de Newton había permanecido durante 250 años, y había alcanzado una precisión extraordinaria, de algo así como una parte en diez millones […] Se había observado una anomalía en el movimiento de Mercurio, pero esto no era ciertamente motivo para abandonar el esquema de Newton. No obstante, Einstein percibió, a partir de bases físicas profundas, que uno podría mejorarlo si cambiaba el propio marco de la teoría gravitatoria. En los años inmediatamente posteriores a la propuesta de Einstein, hubo sólo unos pocos efectos que la apoyaran, y el incremente en precisión sobre el esquema de Newton era irrelevante. Sin embargo, ahora, casi 80 años después de que se propusiera por primera vez la teoría, su precisión global ha aumentado hasta algo del orden de diez millones de veces mayor. Einstein no estaba simplemente «advirtiendo pautas» en el comportamiento de los objetos físicos. Estaba desvelando una subestructura matemática profunda que estaba ya oculta en la propia marcha del mundo (Penrose, 2012: 437).

El mismo Penrose reconoce que este no es un argumento definitivo para defender la existencia del mundo matemático platónico, pero sí que tiene la convicción de que es necesario tenerlo en cuenta para que dicho mundo no sea rechazado *a priori* (Penrose, 2012: 438).

El motivo por el que nuestro autor insiste en esta idea es porque dicha idea está apoyada en otra subyacente y que dentro del pensamiento penroseano es una parte importante. Tal idea ya la vimos anteriormente: el platonismo.

Atendiendo al platonismo que vimos más arriba, podemos ver que en él estábamos inmersos en el otro misterio que involucra al mundo matemático platónico (es decir, el de su relación con el mundo mental). Como hemos visto, la teoría de los tres mundos nos dice que el mundo matemático platónico *emerge* del mundo mental. Y bajo esta declaración cabe la pregunta: ¿cómo puede ser que un mundo cuyos contenidos son perfectos *emerja* de uno con contenidos imperfectos? Es Penrose quien se percata de dicho conflicto conceptual y lo cierto es que no logra resolverlo de un modo del todo satisfactorio. Nuestro autor reconoce que la flecha en el esquema de los tres mundos que representa el emerger del mundo matemático platónico a

partir del mundo mental lleva necesariamente a la idea de que el último es más originario que el primero. No obstante, Penrose, como platónico, es totalmente contrario a dicha concepción. ¿Cómo salva, entonces, el entuerto? Pues restándole importancia al esquema para, de esa forma, intentar no perjudicar al platonismo:

> […] El punto esencial sobre las flechas […] no es tanto su dirección sino el hecho de que en cada caso representan una correspondencia en el que una *pequeña* región de un mundo engloba todo el mundo siguiente (Penrose, 2012: 439).

Con este tipo de afirmaciones es fácil ver que Penrose elaboró este esquema teniendo en mente que no sería definitivo y que habría matices que necesitarían ser perfilados, sobre todo los relativos a dejar clara su postura platonista. Para nuestro autor, si bien el esquema y la teoría de los tres mundos pertenecen a un plano conjetural[251] y metafísico, la existencia platónica, en cambio, tiene una naturaleza objetiva:

> […] En mi opinión, la existencia platónica es simplemente una cuestión de objetividad y, en consecuencia, no debería verse como algo «místico» o «acientífico», pese a que así la consideran algunos (Penrose, 2006: 58).

Este tipo de cuestiones llevan necesariamente a una reelaboración, al menos, del esquema de los tres mundos. Ciertamente así lo hace Penrose, que si bien no cambia en exceso los principios de la teoría, la estructura del esquema sí que se ve sutilmente cambiada.

4.2. Evolución de los tres mundos

Como hemos visto al final del punto anterior, Penrose en su reelaboración del esquema de los tres mundos no pretende hacer cambios drásticos con respecto a las ideas básicas de la teoría. De hecho, esencialmente no existen cambios en absoluto. El esquema sí que experimenta una diferencia sutil a la vez que significativa.

En el anterior esquema (Fig. 11), podíamos ver que las flechas se proyectaban de un mundo a otro, abarcándolo en su totalidad. La implicación de dicha característica es que la relación entre los mundos se da de una manera completa, y esto es algo que Penrose no dice en ningún momento. Por ello entiende que lo más conveniente es cambiar la forma de la flecha, de tal manera que estas accedan a *una parte* del mundo y no a su totalidad, que es aquello que nuestro autor defiende en todo momento. En la siguiente figura se puede observar este cambio.

---

[251] A pesar de que la teoría de los tres mundos es una idea metafísica que defiende, Penrose insiste en aclarar que simplemente es eso, una teoría y que es susceptible de cambios:
[…] estoy tratando de no juzgar de antemano la cuestión de cuál de los mundos, *si los hay*, debe considerarse como primario, secundario o terciario (Penrose, 2012: 440). [La cursiva es mía].
[…] Para englobar cualquiera de estas posibilidades alternativas, (el esquema) tendría que ser dibujado de nuevo, de modo que permitiera que alguno o todos estos mundos se extendiera más allá del ámbito de su flecha precedente (Penrose, 2012: 440).

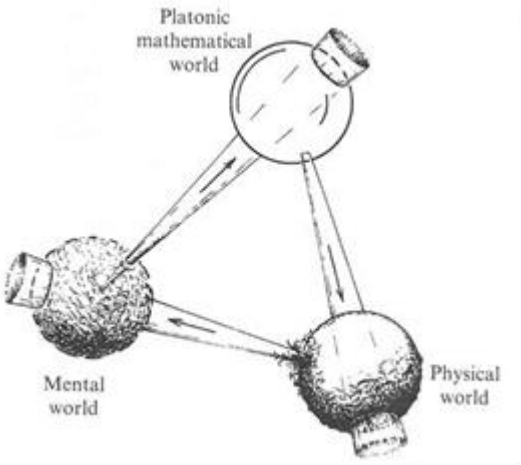

Figura 19. Estructura de los tres mundo de Penrose reelaborada

El aspecto mencionado queda resuelto de manera satisfactoria, pero, como podemos ver, la cuestión que atañe a la dirección de las flechas no cambia en absoluto. Esto lleva a preguntarnos por qué Penrose se toma la molestia de reestructurar el esquema si con él va a seguir sin poder explicar una parte tremendamente conflictiva con respecto al platonismo que defiende.

En realidad, esta pregunta se responde con la actitud que nuestro autor muestra con respecto a este tema. Si bien Penrose desarrolla sus ideas de forma seria e intenta darles visibilidad a través de este tipo de esquemas, lo cierto es que al fin y al cabo se trata de una cuestión puramente metafísica. Es decir, a pesar de que ciertamente crea en aquello que está exponiendo, existen unos límites claros en sus explicaciones y estos no pueden ser sorteados a través de la ciencia o de un método científico. Así que nuestro autor más que cometer un error, considero que lo evita (no mezclando cuestiones puramente metafísicas con la ciencia).

Sin duda, la realidad es un tema de una complejidad innegable e intentar formar y defender cualquier tipo de teoría conduce de forma necesaria hacia el debate, ya sea en el plano de la filosofía o de la ciencia. Es por ello por lo que las ideas de Penrose no son una excepción. Nuestro autor se limita a intentar expresar aquello en lo que cree y ofrece los argumentos más científicos y honestos que puede permitirse.

5. Una última pregunta

El pensamiento de Penrose constituye un gran reto ya no sólo a nivel científico sino también en el filosófico. El modo que tiene de abordar los temas es verdaderamente peculiar, llegando a crear confusión en aquello que quiere tratar en algunas ocasiones. A pesar de que intente no dejar títere con cabeza a través de sus constantes explicaciones explícitas, no puede impedir que sigan surgiendo dudas en torno al tema principal de su pensamiento. ¿Es la respuesta a la Inteligencia Artificial? ¿Es la respuesta a los fundamentos de

las matemáticas? ¿Es la respuesta al fenómeno de la consciencia? Realmente la respuesta a todas estas cuestiones es sí, porque, como hemos visto, Penrose defiende que todas ellas están relacionadas de algún modo. Ahora bien, el asunto que subyace a todas ellas es el determinismo. La verdadera motivación del pensamiento de Penrose es poder responder a la pregunta que se hacían los científicos de la primera mitad del siglo XX, y esta es si los procesos naturales son deterministas o indeterministas. Es por ello por lo que considero que la primera y última pregunta que plantea Penrose, a propósito del debate entre Einstein y Bohr, es: ¿podemos realmente hablar de un triunfo del indeterminismo? El hecho de que lo pregunte lleva consigo la respuesta que tiene reservada Penrose.

Epílogo

…

- No intentemos saber qué pasará y dejemos a un lado nuestros deseos de cara al futuro. Centrémonos mejor en lo que tenemos aquí y ahora. Haciendo tal ejercicio no te queda otro camino que el de ceder. Pero conociéndote, querido Einstein, sé que eso no sucederá.

- Está claro que en algún momento tendremos que dejarlo, pero antes de ello me gustaría plantearte otra cuestión. ¿Aceptas el fatalismo como una forma de determinismo?

- Por supuesto que lo es.

- Bien. Si te hago esta pregunta en concreto es para hacerte ver con un ejemplo fatalista aquello que nos distancia y que, según creo, está el origen de tu error. ¿Te apetece acabar con tal ejemplo?

- No reprimas tus ganas, pues son las mismas que las mías.

- Un caso bastante aclarador nos lo da una historia que sé que conoces a la perfección, esta es, la tragedia *Edipo rey* de Sófocles. En dicha obra, como bien sabes, el protagonista conoce su destino pero no tiene el poder de cambiarlo, a pesar, incluso, de tener la voluntad de querer hacerlo. A lo largo de la trama podemos ver cómo Edipo, sin saberlo, se va acercando cada vez más al final que el oráculo le había revelado. En esta historia puedo verte, mi querido Bohr, y esa imagen es la que me aclara por qué estás sumergido en un error que nos distancia de forma definitiva. Te sitúas en el papel del oráculo, quien puede saber todo, mientras que esta tragedia nos enseña, entre otras cosas, que en este mundo sólo podemos aspirar a ser Edipo. Mi muy estimado Bohr, hay que ser Edipo.

- Ahora entiendo por qué quieres sacarte los ojos y no ver la realidad que se nos está presentando delante de nosotros. También entiendo por qué te sitúas en el papel de Edipo, y es porque aun conociendo lo que sucederá no eres capaz de acertar en qué modo lo hará. Esa es, precisamente, la tragedia del determinismo. Por otro lado, el mundo en el que vivimos no tiene cabida para un oráculo, pero estoy seguro de que si la tuviese este podría contarnos qué grado de libertad nos pertenece en lugar de dictarnos un futuro inamovible.

- De veras que disfruto mucho de tu elocuencia y de tus contraargumentos. Pero sabes, al igual que yo, que no podemos convencer al otro si nos tornamos hacia nuestras convicciones propias, y más si estas requieren una explicación definitiva. Este hecho es algo que nos tiene irremediablemente alejados y se antoja imposible que se resuelva como a nosotros nos gustaría que se resolviese.

Ambos se sonríen y se despiden disponiéndose cada uno a regresar a su respectiva habitación del hotel. Saben que la discusión acabó por hoy, pero muy probablemente no lo hará jamás.

Bibliografía

La siguiente bibliografía consta de dos partes. La primera concierne a las obras del autor principal de este trabajo, mientras que la segunda contiene las demás obras, tanto citadas, consultadas o recomendadas (libros, artículos y capítulos de libros).

Es conveniente aclarar que las traducciones de las citas de las obras en idiomas extranjeros son propias, aunque también es de rigor decir que he intentado en la medida de lo posible contrastar los conceptos con traducciones al castellano.

Como último apunte decir que las distintas obras del (la) mismo(a) autor(a) están ordenados, como podrán verse, por orden alfabético.

PRIMARIA

- Penrose, Roger, Beyond the doubting of a shadow: A reply to commentaries on *Shadows of the mind*, Psyche 2(23) (http://psyche.cs.monash.edu.au/v2/psyche-2-23-penrose.html), 1996.

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

-, *Fashion, Faith and Fantasy in the New Physics of the Universe*, Nueva Jersey, Princeton University Press, 2016.

-, *La nueva mente del emperador*, trad. por Javier García Sanz, Barcelona, Grijalbo Mondadori, 1991.

-, *Las sombras de la mente*, trad. por José Javier García Sanz, Barcelona, Crítica, 2012.

-, *Moda, Fe y Fantasía en la nueva física del universo,* trad. por Marcos Pérez Sánchez, Barcelona, Penguin Random House Grupo Editorial (Debate), 2017.

-, *Shadows of the Mind A Search for the Missing Science of Consciousness*, Nueva York, Oxford University Press, 1994.

❖ Penrose, Roger, Shimony, Abner, Cartwright, Nancy, Hawking, Stephen, *Lo grande, lo pequeño y la mente humana,* trad. por Javier García Sanz, Madrid, Cambridge University Press, 1999.

SECUNDARIA

❖ Arana, Juan, "Aristóteles y la filosofía de las matemáticas", Philosophia, 78, 2018, pp. 9-30.

-, "Determinismo y libertad en Karl Popper", *Anuario filosófico,* 34, 2001, pp. 119-138.

-, "Erwin Schrödinger, filósofo de la Biología", revista Thémata, 20, 1998 159-174.

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

-, *Los sótanos del universo: La determinación natural y sus mecanismos ocultos,* Madrid, Biblioteca Nueva, 2012.

-, *Materia, Universo, Vida*, Madrid, Tecnos, 2001.

-, "La visione cósmica di Erwin Schrödinger, trad. por Rita Bettaglio, perteneciente a: Conferenze di Scienza e Fede, Roma, Ateneo Pontificio Regina Apostolorum, 2008.

-, "Los múltiples rostros del determinismo", perteneciente a: Conferencia inaugural de la Primera Semana de Investigación Interdisciplinar: Determinismo e Indeterminismo. De la Física a la Filosofía, Buenos Aires, 2013.

-, "Panteísmo y ética en la vida y obra de Albert Einstein", Thémata 14, 1995, 181-196.

❖ Aristóteles, *Obras completas*, trad. por Tomás Calvo Martínez, Guillermo R. De Echandía, Madrid, Gredos, Tomo I, 2011.
❖ Audi, Robert [ed.], *Diccionario Akal de Filosofía*, trad. por Huberto Marraud y Enrique Alonso, Madrid, Ediciones Akal, 2004.
❖ Avigad, Jeremy, "Methodology and metaphysics in the development of Dedekind´s theory of ideals" en José Ferreirós & Jeremy J. Gray, *The Architecture of Modern Mathematics: Essays in History and Philosophy*, Oxford, Oxford University Press, 2006, pp. 159-186.

- Bachelard, Gastón, *El agua y los sueños: Ensayo sobre la imaginación de la materia*, trad. por Ida Vítale, México D.F., Fondo de Cultura Económica, 1978.
- Badía Serra, Eduardo, "Roger Penrose: Una aproximación elemental a su filosofía de la ciencia", Teoría y Praxis, 13, 2008, pp. 53-80.
- Batey, Mavis, "Breaking machines with a pencil" en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017, pp. 97-108.
- Belot, Gordon, "The Representation of Time and Change in Mechanics" en J. Butterfield & J. Earman, *Handbook of the Philosophy of Science*, Ámsterdam, Elsevier, 2007, pp. 133-228.
- Benacerraf, Paul, Putnam, Hilary, *Philosophy of mathematics: Selected readings*, Nueva York, Cambridge University Press, 1983.
- Bernays, Paul, "On Platonism in mathematics", en Paul Benacerraf & Hilary Putnam, *Philosophy of mathematics: Selected readings*, Nueva York, Cambridge University Press, 1983, pp. 258-271.
- Berto, Francesco, *There´s something about Gödel: The complete guide to the Incompleteness Theorem*, Oxford, Wiley-Blackwell, 2009.
- Bohr, Niels, *La teoría atómica y la descripción de la Naturaleza*, trad. por Miguel Ferrero Melgar, Madrid, Alianza, 1988.

-, *Nuevos ensayos sobre física atómica y conocimiento humano*, trad. por Carlos Rodríguez, Madrid, Aguilar, 1970.

-, Reply to "Can Quantum-Mechanical Description of Physical Reality Be Considered Completed?", Physical Review, 48, 1935, 696-702.

- Borges, Jorge Luis, *Obras completas*, Buenos Aires, Emecé Editores, 1974.
- Born, Max, *Natural Philosophy of cause and chance*, Oxford, Clarendon Press, 1949.
- Born, Max, Born, Hedwig, Einstein, Albert, *Correspondencia (1916-1955)*, trad. por Félix Blanco, México D.F., Siglo XXI editores, 1971.
- Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014.
- Brouwer, L.E.J., "Intuitionism and formalism" en Paul Benacerraf & Hilary Putnam, *Philosophy of mathematics: Selected readings*, Nueva York, Cambridge University Press, 1983, pp. 77-89.
- Brown, James Robert, *Platonism, Naturalism, and Mathematical Knowledge*, Londres, Routledge, 2012.
- Burtt, Edwin Arthur, *Los fundamentos metafísicos de la ciencia moderna*, trad. por Roberto Rojo, Buenos Aires, Editorial Sudamericana, 1960.
- Butterfield, J., Earman, J., *Handbook of the Philosophy of Science*, Ámsterdam, Elsevier, 2007.

- Capek, Milic, *El impacto filosófico de la física contemporánea*, trad. por Eduardo Gallardo Ruíz, Madrid, Tecnos, 1973.
- Carnap, Rudolf, *Fundamentación lógica de la física*, trad. por Néstor Miguens, Barcelona, Orbis, 1985.

-, "The logicist foundations of mathematics" en Paul Benacerraf & Hilary Putnam, *Philosophy of mathematics: Selected readings*, Nueva York, Cambridge University Press, 1983, pp. 41-51.

- Chopra, Deepak, "Reality and consciousness: A view from the East"; Comment on "Conciousness in the universe: A review of the «Orch OR» theory" by Stuart Hameroff and Roger Penrose, Physics of Life Reviews, 11, 2014, 81-82.
- Churchland, Paul, *Neurophilosophy at work*, Nueva York, Cambridge University Press, 2007.
- Copeland, Jack, *Inteligencia artificial: Una introducción filosófica*, trad. por Julio César Armero San José, Madrid, Alianza, 1996.
- Copeland, Jack, "Intelligent machinery", en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017a, pp. 265-276.
- Copeland, Jack, "Turing´s great invention: the universal computing machine" en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017b, pp. 49- 56.
- Copeland, Jack, Bowen, Jonathan, Sprevak, Mark, Wilson, Robin, *The Turing Guide*, Oxford, Oxford University Press, 2017.
- Corry, Leo, "La teoría de las proporciones de Eudoxio interpretada por Dedekind", Mathesis, 10, 1994, 1-24.
- Detlefsen, Michael, "Formalism", en Stewart Shapiro [ed.] *The Oxford Handbook of Philosophy of Mathematics and Logic*, Nueva York, Oxford University Press, 2005, pp. 236-217.

-, "Philosophy of Mathematics in the twentieth century" en Stuart Shanker (ed.), *Philosophy of Science, Logic and Mathematics in the twentieth century*, Routledge, Londres, 1996, pp. 50-123.

- Dennett, Daniel, *La conciencia explicada: Una teoría interdisciplinar*, trad. por Sergio Balari Ravera, Barcelona, Paidós, 1995, pp. 33-78.

-, *Dulces sueños: Obstáculos filosóficos para una ciencia de la conciencia*, trad. por Julieta Barba y Silvia Jawerbaum, Buenos Aires, Katz, 2006.

- Descartes, René, *Los principios de la filosofía*, trad. por Guillermo Quintás, Madrid, Alianza, 1995.

-, *Obras completas*, trad. por Ana Gómez Rabal, Madrid, Gredos, 2011.

- Diéguez Lucena, Antonio, *Filosofía de la ciencia*, Madrid, Biblioteca Nueva, 2005.
- Drake, Stillman, *Galileo*, trad. por Alberto Elena, Madrid, Alianza, 1980.
- Einstein, Albert, Born, Max y Hedwig, *Correspondencia (1916-1955)*, trad. por Félix Blanco, México D.F., Siglo veintiuno editores, 1973.
- Einstein, Albert, *El significado de la relatividad*, trad. por Carlos E. Prélat, Madrid, Editorial Espasa Calpe, 2005.

-, *La física*, trad. por Rafael Grinfeld, Buenos Aires, Editorial Losada, 1974.

-, *Mi visión del mundo*, trad. por Sara Gallardo y Marianne Bübeck, Barcelona, Tusquets Editores, 2013.

-, *Sobre la teoría de la relatividad especial y general*, trad. por Miguel Paredes Larrucea, Madrid, Alianza Editorial, 2002.

- Einstein, A., Podolsky, B., Rosen, N., "Can Quantum-Mechanical Description of Physical Reality Be Considered Completed?", Physical Review, 47 (1935), 777-780.
- D´Espagnat, Bernard, *En busca de lo real. La visión de un físico*, trad. por Tomás Fernández Rodríguez, Miguel Ferrero Melgar y José A. López Brugos, Madrid, Alianza Editorial, 1983, pp. 88-101, 112-141.

-, *On Physics and Philosophy*, New Jersey, Princeton University Press, 2006, pp. 89-101, 113-127, 319-323.

- Espinoza, Miguel, Torretti, Roberto, *Pensar la ciencia*, Madrid, Editorial Tecnos, 2004.
- Feferman, Solomon, "Penrose's Gödelian Argument: A Review of Shadows of the Mind by Roger Penrose," *Psyche*, 2 (7), 1995, http://psyche.cs.monash.edu.au/ v2/psyche-2-07-feferman.html.
- Ferreirós, José, "Hilbert, logicism, and mathematical existence", (Springer) Synthese, 170, 2009, 33-70.
- Ferreirós, José, "Matemáticas y platonismo(s)", La Gaceta de la Real Sociedad Española de Matemáticas, 2, 1999, 446-473.
- Ferreirós, José, *Mathematical Knowledge and the Interplay of Practices*, New Jersey, Princeton University Press, 2016.
- Ferreirós, José, Gray, Jeremy J., *The Architecture of Modern Mathematics: Essays in History and Philosophy*, Oxford, Oxford University Press, 2006.
- Ferrero Blanco, Juan José, *Galileo Galilei: el filósofo*, Bilbao, Universidad de Deusto, 1986.
- Février, Paulette, *Determinismo e Indeterminismo*, trad. por Raquel de Gortari, México D.F., Universidad Nacional Autónoma de México, 1957, pp. 15-26, 131-171.

- Floyd, Juliet, Bokulich, Alisa [eds.], *Philosophical Explorations of the Legacy of Alan Turing: Turing 100*, Boston, Springer, 2017.
- Floyd, Juliet, "Turing on "Common sense": Cambridge Resonances" en Juliet Floyd & Alisa Bokulich [eds.], *Philosophical Explorations of the Legacy of Alan Turing: Turing 100*, Boston, Springer, 2017, pp. 103-152.
- Forouzan, B.A., *Introducción a la ciencia de la computación: De la manipulación de datos a la teoría de la computación*, trad. por Lorena Peralta, México D.F., International Thompson Editores, 2003, pp. 321-325.
- Fuentes Fernández, Jorge, "¿Ondas o partículas?: La teoría de la doble solución de Louis de Broglie", Ciencias, 117, 2015, pp. 14-25.
- Galilei, Galileo, *Opere di Galileo Galilei*, Torino, Unione Tipografico-Editrice Torinese, 1964, Vol. 1.
- Galilei, Galileo, Kepler, Johannes, *El mensaje y el mensajero sideral*, trad. por Carlos Solís Santos, Madrid, Alianza, 1990.
- Gamble, John, *Foundations of Quantum Decoherence*, Wooster, The college of Wooster, 2008, (Tesis doctoral).
- Gardner, Howard, *Multiple Intelligences: New Horizons*, Nueva York, Basic Books, 2006.
- Ghosh, S., Sahu, S., Bandyopadhyay, A., "Evidence of massive global synchronization and the conciousness"; Comment on "Conciousness in the universe: A review of the «Orch OR» theory" by Stuart Hameroff and Roger Penrose, Physics of Life Reviews, 11, 2014, pp. 83-84.
- Gödel, Kurt, *Obras completas*, trad. por Jesús Mosterín, Madrid, Alianza Editorial, 2006, pp. 53-89.
- Goldstein, Catherine, Norbert Schappacher, Joachim Schwermer [eds.] *The Shaping of Arithmetic after C. F. Gauss's Disquisitiones Arithmeticae*, Nueva York, Springer, 2007.
- Greenberg, Joel, "The Enigma machine" en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017, pp. 85-96.
- Grosholz, Emily, Breger, Herbert [eds.], *The Growth of Mathematical Knowledge*, Boston, Springer-Science+Business Media, B.Y, Volume 289, 2000.
- Hameroff, S., Penrose, R., "Conciousness in the universe: A review of the «Orch OR» theory", Physics of Life Reviews, 11, 2014, 39-78.

-, Reply to seven commentaries on "Conciousness in the universe: A review of the «Orch OR» theory", Physics of Life Reviews, 11, 2014, 94-100.

- Hale, Bob, Wright, Crispin, "Logicism in the Twenty-first Century" en Shapiro, Stewart [ed.], *The Oxford Handbook of Philosophy of Mathematics and Logic*, Nueva York, Oxford University Press, 2005, pp. 166-202.

- Hardy, Godfrey Harold, *A mathematician apology*, Londres, Baaltis publishing, 2014.
- Heisenberg, Werner, *Física y filosofía*, trad. por Fausto de Tezanos Pinto, Buenos Aires, Ediciones La isla, 1959.
- Herce, Rubén, *De la Física a la Mente: El proyecto filosófico de Roger Penrose*, Madrid, Biblioteca Nueva, 2014.

-, "Penrose on what scientists know", Found Sci, (2016), 21: 679.

- Hofstadter, Douglas, *Gödel, Escher, Bach: Una eterna trenza dorada*, trad. por Mario Arnaldo Usabiaga Brandizzi, México D.F., Consejo Nacional de Ciencia y Tecnología, 1982.
- Holbach, Paul Henri Thiry, *Sistema de la naturaleza*, trad. por Nerina Bacín, José Manuel Bermudo, Miguel Estapé y Alín Salom, Pamplona, Laetoli, 2008.
- Holton, Gerald, *Victory and Vexation in Science: Einstein, Bohr, Heisenberg and others*, Londres, Harvard University Press, 2005.
- Honner, John, *The Description of Nature: Niels Bohr and the Philosophy of Quantum Physics*, Nueva York, Oxford University Press, 1987.
- Howard, Don, "Was Einstein Really a Realist?" Perspectives on Science: Historical, Philosophical, Social, 1 (1993), 204-251.
- Jumper, C., Scholes, G.D., "Life- Warm, wet and noisy?"; Comment on "Conciousness in the universe: A review of the «Orch OR» theory" by Stuart Hameroff and Roger Penrose, Physics of Life Reviews, 11 (2014), 85-86.
- Kennedy, Juliette, "Turing, Gödel and the "Bright Abyss" en Juliet Floyd & Alisa Bokulich [eds.], *Philosophical Explorations of the Legacy of Alan Turing: Turing 100*, Boston, Springer, 2017, pp. 63-92.
- Kirk, G.S., Raven, J.E., Schofield, M., *Los filósofos presocráticos: Historia crítica con selección de textos*, trad. por Jesús García Fernández, Madrid, Gredos, 2008.
- Kline, Morris, *Mathematical Thought from Ancient to Modern Times*, Volumen I, Nueva York, Oxford University Press, 1972.
- Kreisel, Georg, "Hilbert´s programme" en Paul Benacerraf & Hilary Putnam, *Philosophy of mathematics: Selected readings*, Nueva York, Cambridge University Press, 1996, pp. 207-238.
- Kuhn, Thomas, *La estructura de las revoluciones científicas*, trad. por Agustín Contín, Buenos Aires, Fondo de Cultura Económica, 2004.
- Ladrière, Jean, *Limitaciones Internas de los Formalismos: Estudio sobre la significación del Teorema de Gödel y teoremas conexos en la teoría de los fundamentos de las matemáticas*, trad. por José Blaso, Madrid, Tecnos, 1969.
- Laplace, Pierre-Simon, *Essai philosophique sur les probabilités*, Cambridge, Cambridge University Press, 2009.

-, *Exposición del sistema del mundo*, trad. por José Luis Arántegui Tamayo, Barcelona, Crítica, 2006.

- Lindley, David, *Incertidumbre: Einstein, Heisenberg, Bohr y la lucha por la esencia de la ciencia*, trad. por Joan Soler, Barcelona, Editorial Ariel, 2008.
- Lindström, Per, "Penrose's New Argument," Journal of Philosophical Logic, 30, 2001, pp. 241–50.
- Linnebo, Øystein, *Philosophy of Mathematics*, Nueva Jersey, Princeton University Press, 2017.
- Lucas, John, "The face of freedom"; Comment on "Conciousness in the universe: A review of the «Orch OR» theory" by Stuart Hameroff and Roger Penrose, Physics of Life Reviews, 11, 2014, 87-88.
- Maddy, Penelope, *Realism in Mathematics*, Nueva York, Oxford University Press, 2003.
- McDermott, Drew, "Penrose is wrong: A Review of Shadows of the mind by Roger Penrose", *Psyche*, 2(17), 1995, http://psyche.cs.monash.edu.au/v2/psyche-2-17-mcdermott.html
- Minsky, Marvin Lee, *La sociedad de la mente: la inteligencia humana a la luz de la inteligencia artificial*, trad. por Lidia Espinosa de Matheu, Buenos Aires, Ediciones Galápago, 1986.
- Muñoz, Jorge, Moya, Paco, "Números reales y complejos" en Libros Marea Verde, 2014, www.apuntesmareaverde.org.es.
- Moya, Carlos, *El libre albedrío: un estudio filosófico*, Madrid, Cátedra, 2017.
- Ordoñez, Javier, Navarro, Víctor, Sánchez Ron, José Manuel, *Historia de la ciencia*, Madrid, Editorial Espasa Calpe, 2007, pp. 559-618.
- Pino, S., Di Mauro, E., "How to conciliate Popper with Cartesius"; Comment on "Conciousness in the universe: A review of the «Orch OR» theory" by Stuart Hameroff and Roger Penrose, Physics of Life Reviews, 11, 2014, 91-93.
- Platón, *República*, trad. por Conrado Eggers Lan, Madrid, Gredos, Tomo IV, 1992.
- Ponte Azcárate, María, *Realismo y entidades abstractas: Los problemas del conocimiento en matemáticas*, Servicio de publicaciones de la Universidad de La Laguna, 2006 (Tesis doctoral).
- Poole, David, Mackworth, Alan, Goebel, Randy, *Computational Intelligence: A logicial approach*, Oxford, Oxford University Press, 1998.
- Popper, Karl R., "Indeterminism in Quantum Physics and in Classical Physics", British Journal for the Philosophy of Science 1(2), 1950, pp. 117-133.
- Posy, Carl, "Intuitionism and Philosophy", en Stewart Shapiro [ed.] *The Oxford Handbook of Philosophy of Mathematics and Logic*, Nueva York, Oxford University Press, 2005, pp. 318-355.

- Prigogine, Ilya, *Las leyes del caos*, trad. por Juan Vivanco, Barcelona, Grijalbo Mondadori, 1997.
- Prigogine, Ilya, Stengers, Isabelle, *La nueva alianza: Metamorfosis de la ciencia*, trad. por María Cristina Martín Sanz, Madrid, 2002.
- Proudfoot, Diane, "Child machines" en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017a, pp. 315-326.

-, "Turing´s concept of intelligence" en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017b, pp. 301- 308.

- Rabossi, Eduardo, *Filosofía de la mente y ciencia cognitiva*, Paidós, Barcelona, 1995.
- Randell, Brian, "Turing and the origins of digital computers" en Jack Copeland, Jonathan Bowen, Mark Sprevak & Robin Wilson, *The Turing Guide*, Oxford, Oxford University Press, 2017, pp. 67-76.
- Rioja, Ana, "La dualidad onda-corpúsculo en la filosofía de Max Born", Thémata 14, 1995, 251-284.
- Rioja, Ana, Ordoñez, Javier, *Teorías del universo (Volumen 3): De Newton a Hubble*, Madrid, Síntesis, 2006, pp. 211-253.
- Rodríguez Valls, Francisco [ed.], *La inteligencia en la naturaleza: Del relojero ciego al ajuste fino del universo*, Madrid, Biblioteca Nueva, 2012, pp. 101-118.
- Russell, Bertrand, Whitehead, Alfred North, *Principia Mathematica*, Cambridge, Cambridge University Press, 1997.
- Russell, Bertrand, "Selection from *Introduction to Mathematical Philosophy*, en Paul Benacerraf & Hilary Putnam, *Philosophy of mathematics: Selected readings*, Nueva York, Cambridge University Press, 1983, pp. 160-182.
- Russell, Stuart J., Norvig, Peter, *Inteligencia Artificial: Un enfoque moderno*, trad. por Juan Manuel Corchado Rodríguez, Fernando Martín Rubio, José Manuel Cadenas Figueredo, Luis Daniel Hernández Molinero, Enrique Paniagua Arís, Raquel Fuentetaja Pinzán, Mónica Robledo de los Santos y Ramón Rizo Aldeguer, Madrid, Pearson Educación S.A., 2004.
- Sánchez Ron, José Manuel, "Einstein y la filosofía del siglo XX", Pensamiento y Cultura, 728, 2007, 833-853.

-, "Las filosofías de los creadores de la mecánica cuántica", Thémata 14, 1995, 197-222.

- Saunders, Simon, "Complementarity and Scientific Rationality", Foundations of Physics 35(3), 2005, pp. 417-447.
- Searle, John, "Mentes y cerebros sin programas" en Eduardo Rabossi, *Filosofía de la mente y ciencia cognitiva*, Paidós, Barcelona, 1995, pp. 413-444.

-, “Minds, brains and programmes”, The behavioral and brain sciences, 3, 1980, pp. 417-457.

- Schrödinger, Erwin, *Mi concepción del mundo* seguido de *Mi vida*; *Mi concepción del mundo* trad. por Jaime Fingerhut; y *Mi vida* trad. por Arthur Klein, Barcelona, Tusquets, 2011.

-, *La naturaleza y los griegos*, trad. por Víctor Gómez Pin, Barcelona, Tusquets, 1997.

-, *La nueva mecánica ondulatoria y otros escritos*, trad. por Xavier Zubiri y Juan Arana, Madrid, Biblioteca Nueva, 2001.

-, *¿Qué es la vida?*, Salamanca, Textos de biofísica, 2005.

-, *¿Qué es una ley de la Naturaleza?*, trad. por Juan José Utrilla, México D.F., Fondo de Cultura Económica, 1975

-, *What is life?*, Cambridge, Cambridge University Press, 1967.

- Scardigli, Fabio, t´Hooft, Gerard, Severino, Emanuele, Coda, Piero, *Determinism and Free Will: New Insights from Physics, Philosophy, and Theology*, Cham, Springer, 2019.
- Shanker, Stuart G. [ed.], *Philosophy of Science, Logic and Mathematics in the Twentieth Century*, Londres, Routledge, 1996.
- Shapiro, Stewart “Mechanism, Truth, and Penrose’s New Argument,” Journal of Philosophical Logic, 32, 2003, pp. 19–42.
- Shapiro, Stewart [ed.], *The Oxford Handbook of Philosophy of Mathematics and Logic*, Nueva York, Oxford University Press, 2005.
- Soler Gil, Francisco, “¿Es la conciencia un estado de la materia?”, Naturaleza y Libertad. Revista de estudios interdisciplinares, 10, 2017, pp. 299-310.
- Solís, Carlos, Sellés, Manuel, *Historia de la ciencia*, Madrid, Espasa, 2009.
- Steiner, Mark, “Penrose and Platonism”, en Emily Grosholz, Herbert Breger, *The Growth of Mathematical Knowledge*, Springer-Science+Business Media, B.Y, Volume 289, 2000, pp. 133-142.
- Tandy, Charles, “Are you (almost) a zombie? Conscious thoughts about Conciousness in the universe by Hameroff and Penrose”, Physics of Life Reviews, 11, 2014, 89-90.
- Tegmark, Max, *Life 3.0: Being Human in the Age of Artificial Intelligence*, Nueva York, Alfred A. Knopf, 2017.
- t´Hooft, Gerard, “Free Will in the Theory of Everything”, en Fabio Scardigli, Gerard t´Hooft, Emanuele Severino & Piero Coda, *Determinism and Free Will: New Insights from Physics, Philosophy, and Theology*, Cham, Springer, 2019, pp. 21-48.
- Turing, Alan, “Maquinaria computacional e Inteligencia”, trad. por Cristóbal Fuentes Barassi, Santiago de Chile, Universidad de Chile, 2010.

- Tuszynski, Jack A., “The need for a physical basis of cognitive process”; Comment on “Conciousness in the universe: A review of the «Orch OR» theory” by Stuart Hameroff and Roger Penrose, Physics of Life Reviews, 11, 2014, 79-80.
- Van Atten, Mark, *Brouwer meets Husserl on the phenomenology of choice sequences*, París, Springer, 2007.
- Yates, Frances, *Giordano Bruno y la tradición hermética*, trad. por Domènec Bergadà, Barcelona, Ariel, 1983.