

Roger Penrose

Una aproximación elemental a su filosofía de la ciencia.

Eduardo Badía Serra¹.

Este pequeño trabajo no es un profundo análisis de la ciencia y de la filosofía de Penrose. No tiene tal pretensión, ni cercanamente. Es simplemente una exposición breve y concisa de la obra del científico-filósofo, una especie de recensión si se quiere, cuyo único propósito es colocar dicha obra en el tapete de la discusión y el conocimiento que, seguramente, otros más conocedores que yo de la misma, podrán hacer.

This short article is not a deep analysis of the science and the philosophy of Penrose. It does not intend to do that, not even close to that. It is simply a brief and concise presentation of the work done by such scientist and philosopher, a sort of review whose only purpose is to bring his work up for discussion and knowledge that others more knowledgeable than me will certainly make.

NOTA ACLARATORIA.

Intentar una introducción a la filosofía de Roger Penrose es, debo decirlo, una acción aventurada e incluso imprudente. Penrose es un científico muy difícil de entender y más aún de comprender, y, aunque no dudo en afirmar que dentro de su obra científica hay un marcado, elegante y profundo fondo filosófico, calificarlo como filósofo sin duda despertará la crítica. Si ya sus posiciones dentro de la física y de la matemática son fuertemente controversiales y discutidas, más aun podrían serlo sus opiniones filosóficas, campo en el cual se manifiesta como un decidido y convencido platónico.

1. Ingeniero químico, colaborador de Teoría y Praxis.

Pero el intento puede validarse por varias razones: En primer lugar, la necesidad reconocida de que la filosofía debe entrar en el conocimiento de la ciencia para poder proponer un acceso a la realidad más, vale decir, real; en segundo lugar, porque las opiniones de Penrose denotan, no sólo un adecuado y correcto conocimiento de la historia de la filosofía y de sus problemas fundamentales y eternos, sino un conveniente acomodamiento a las condiciones actuales del enfoque de dichos problemas; y finalmente, porque los temas que toca son de una importancia capital para el conocimiento y la cultura, y cito entre estos, el del problema mente-cuerpo y su relación con la Inteligencia Artificial Fuerte, el del origen y desarrollo de la conciencia, su teoría de los tres mundos y el acceso a la realidad, sus posiciones ante el método científico, su clasificación de las teorías, su "*explicación*" de la realidad universal", etc.

Roger Penrose nació en el Reino Unido, en Essex, en 1931. De formación matemática, su relación con los físicos teóricos más famosos del momento, entre los que puede contarse el mismo Stephen Hawking, le hizo incursionar en este importante campo del conocimiento, haciéndolo de una manera tan elegante como profunda. Indudablemente, los problemas filosóficos que la ciencia sabe manifestar en sus investigaciones le llevaron también al estudio de esta ya vieja *ciencia de las ciencias*. Sin que haya habido en él una formación filosófica académica y sistemática, su profunda formación científica le ha permitido entrar en los problemas filosóficos con una admirable propiedad. Es actualmente profesor emérito de matemáticas en la Universidad de Oxford, y entre sus principales obras se encuentran: La nueva mente del Emperador; Las sombras de la mente; Lo grande, lo pequeño y la mente humana; El camino a la realidad; Cuestiones cuánticas y cosmológicas; La naturaleza del espacio y del tiempo; Técnicas de topología diferencial en relatividad.

De nuevo, mis disculpas por el atrevimiento, en la esperanza de que hablar de la obra filosófica de Penrose provoque el efecto esperado: Que los más preparados entren en su estudio y la expongan y aclaren mejor.

I. - TRES MUNDOS DE PENROSE Y LA REALIDAD DE LOS CONCEPTOS MATEMÁTICOS

Antes de exponer la teoría de los tres mundos de Penrose, considero necesario anticipar su posición sobre la realidad y la verdad de los conceptos matemáticos, muy ligados con ella.

I.1 La realidad de los conceptos matemáticos.

Penrose plantea dos posiciones alrededor de la realidad de los objetos del mundo matemático: La primera, que sostiene que no hay nada real en ellos, que los objetos matemáticos sólo son conceptos, idealizaciones mentales que hacen los matemáticos. La segunda, que sostiene que hay una realidad profunda en ellos que va más allá de las elucubraciones mentales de un matemático particular, y que el pensamiento matemático va siendo guiado hacia alguna verdad exterior, que tiene realidad por sí misma y que sólo se nos revela parcialmente a alguno de nosotros. (1-123). Siempre que la mente percibe una idea matemática toma contacto con el mundo platónico de las ideas matemáticas. Según el punto de vista platónico, las ideas matemáticas tienen una existencia propia y habitan en un mundo ideal que sólo es accesible por la vía del intelecto. Penrose sostiene la existencia real del mundo matemático platónico, en oposición a muchas posiciones, sobre todo de la filosofía, que lo consideran producto de una mera ficción o de una imaginación desbordante. (3-53)

Penrose participa de la segunda posición. Los objetos matemáticos no son invenciones, elaboradas *construcciones mentales* que no tienen auténtica realidad pero que engañan incluso a sus mismos inventores haciéndoles creer que son reales, sino *descubrimientos* de verdades que "*ya están ahí*", verdades cuya existencia es independiente de los matemáticos. Cita como ejemplos, los números complejos y el conjunto de Mandelbrot. Siente que en el caso de las matemáticas, la creencia en algún tipo de existencia etérea y eterna, al menos para los conceptos más profundamente matemáticos, es muy fuerte. Hay en los conceptos matemáticos una compulsiva unicidad y universalidad que parecen ser de un orden diferente del que se pudiera esperar de las artes o de la ingeniería. Y cita que esto ya había sido planteado de alguna manera por Platón, por lo que a veces, a esta posición se le suele llamar *Platonismo Matemático*. (1-124,125)

Sin embargo, algunas verdades matemáticas parecen tener una realidad platónica más fuerte, más profunda, más interesante, que otras. Tales verdades son las que se identificarían más estrechamente con el funcionamiento de la realidad física, como el sistema de los números complejos. Esta identificación puede hacer comprensible el cómo la mente establece la conexión entre el mundo físico y el mundo platónico de las matemáticas, de la cual se hablará posteriormente. Hay que recordar que muchas partes del mundo matemático, precisamente las más profundas y las más interesantes, no tienen carácter algorítmico. Por ello, la acción no algorítmica debería tener un papel de importancia muy considerable dentro del mundo físico, y este papel estará muy ligado, íntimamente ligado, con el propio concepto "*mente*". (1-507,508)

Ya Platón decía que los objetos de la geometría pura, (líneas, rectas, círculos, triángulos, planos, etc.), sólo se realizaban aproximadamente en el mundo de las cosas físicas reales. Los objetos matemáticamente precisos de la geometría pura no poblaban este mundo físico sino un mundo diferente, "*el mundo ideal de Platón de los objetos matemáticos*", que no nos es accesible del modo físico ordinario sino por la vía del intelecto. Nuestra mente entra en contacto con el mundo de Platón cada vez que contempla una verdad matemática, percibiéndola mediante el ejercicio del razonamiento y la intuición matemática. El mundo ideal se consideraba diferente y más perfecto que el mundo material de nuestra experiencia externa, pero tan real como este. Por ejemplo, los objetos de la geometría euclidiana pura pueden ser estudiados por el pensamiento, aunque no necesariamente el mundo físico imperfecto de la experiencia externa se ajuste a este ideal, y el funcionamiento de este mundo externo real puede ser entendido finalmente sólo por virtud de las matemáticas exactas, lo que, en términos del mundo ideal de Platón, quiere decir "*accesible por la vía del intelecto*". (1-193,194)

I.2 La verdad matemática.

Penrose enfoca este problema desde tres puntos de vista: El del *formalismo*, el del *intuicionismo*, y el del *platonismo*. Como podría fácilmente anticiparse, rechaza los dos primeros.

Para el *formalismo*, los enunciados matemáticos carecen de significado, y no son otra cosa que cadenas de símbolos en algún sistema matemático formal. Son, así, una especie de juegos sin significado. Penrose rechaza esto, pues, para él, es el significado y no la ciega computación algorítmica lo que constituye la sustancia de las matemáticas. Según él, la noción formalista de la verdad debe ser necesariamente incompleta. La falta de interés que profesan los formalistas por la verdad matemática le parece a él un muy extraño punto de vista desde la filosofía de las matemáticas. (1-133,134,137,138). Desprende del teorema de Godel, que es un *golpe devastador al formalismo*, que el concepto de verdad matemática no puede ser encapsulado dentro de ningún esquema formalista, pues es algo que trasciende el mero formalismo, dado que hay en él *algo absoluto e infuso* que trasciende las meras *construcciones humanas*. (1-141,142).

En cuanto al *intuicionismo*, también llamado *finitismo*, que, según Penrose, es mantenido por una minoría *bastante pequeña* de matemáticos, estos rechazan la existencia consumada de cualquier conjunto infinito. Son una respuesta a Russel y su denodado afán por el uso de conceptos infinitos en el razonamiento matemático, (Paradojas de Russel). Se dice que las raíces del *intuicionismo* o *finitismo* se encuentran en Aristóteles. Para el *intuicionismo*, los conjuntos, infinitos o no, no deben pensarse como si tuvieran existencia por sí mismos, sino que deben pensarse en términos de las reglas por las que

se puede determinar su pertenencia a ellos. (1-144). El *intuicionismo* rechaza la ley del *tercio excluido*, que afirma que la negación de la negación de un enunciado es equivalente a su afirmación, lo cual, en términos del sentido común, es una verdad evidente, y que es la base del método matemático conocido como *reducción al absurdo*. Los intuicionistas niegan esta ley. Su objeción consiste en que dicho principio puede ser *no constructivo* en cuanto que a veces lleva a una afirmación de la existencia de cierta entidad matemática, sin que se ofrezca ninguna construcción real de la misma. (3-93, 102). Para ellos, el hecho de que un concepto matemático exista, exige previamente una construcción mental definida, esto es, es una *existencia constructiva*. En la lógica intuicionista, de la falsedad de la no existencia de un objeto no se puede deducir que este exista realmente. Esta posición, para Penrose, es demasiado extrema.

Penrose se adhiere, como ya se ha anticipado, al *platonismo matemático*, para quienes tanto la existencia de los conceptos matemáticos como el carácter absoluto de la verdad matemática son esencialmente la misma cosa. *La independencia del sujeto* les confiere su existencia platónica, y sus *detalles más finos* quedan fuera del alcance de las computadoras. Gödel era un gran platónico, (1-142, 143); sin embargo, hace ver que a medida que los conjuntos matemáticos, por ejemplo, se vuelven desmesuradamente enormes y son contruidos de maneras muy vagas, se puede incluso dudar de que su existencia o su inexistencia sea realmente algo absoluto, e incluso comiencen a adquirir algo de la cualidad de *cuestión de opinión* en lugar de *idea infusa*. Otorgar la calidad de existencia absoluta a un conjunto desmesuradamente enorme y expresado de manera muy vaga es, dice Penrose, llevar el platonismo matemático a un nivel extremo, llevarlo *hasta sus últimas consecuencias*. (1-143, 144). Penrose no oculta sus simpatías por el *platonismo matemático*, en el sentido que la verdad matemática es absoluta, externa y eterna, y no se basa en criterios hechos por el hombre, y de que los objetos matemáticos tienen una existencia intemporal por sí mismos, independiente de la sociedad humana o de los objetos físicos particulares. (1-147)

Las verdades matemáticas son verdades necesarias, por lo que ninguna información real, en el sentido técnico, pasa al descubridor. Toda la información ha estado allí todo el tiempo, y es sólo cuestión de atar los cabos y ver la respuesta. Por ello, siguiendo a Platón, *el descubrimiento matemático es sólo una forma de recuerdo*. (1-505)

I.3 Los tres mundos de Penrose. Prejuicios y misterios.

Penrose inicialmente plantea la existencia de dos mundos, *el mundo físico y el mundo platónico de las ideas matemáticas*. (2-13, 14, 15). Realmente, este mundo platónico contiene otras ideas tales como la de la belleza y la

de la bondad, pero él prefiere, aun reconociendo lo anterior, circunscribir dicho mundo al de las ideas matemáticas. (2-14; 3-66-67-68). Refutando a quienes consideran que este *mundo platónico de las ideas matemáticas* no puede tener una existencia independiente, y que parten de que los conceptos matemáticos son meras idealizaciones del mundo físico, del cual emergen, él sostiene que es el mundo físico el que emerge del *intemporal mundo de las matemáticas*. (2-15).

Hay aquí que recordar que, si bien ha ahondado en la física y en la filosofía, él es primariamente un matemático, cuestión que sabe afirmar repetidamente en su obra. Para él, la clave para entender la naturaleza reside en unas matemáticas incuestionables, y a ellas se debe el primer avance trascendental de la ciencia. (3-50). Por ejemplo, los números reales y los números complejos constituyen un ingrediente esencial en la comprensión del mundo físico. (3-120 a 127). No debe extrañarnos entonces su posición. Penrose defiende la existencia real del mundo platónico asumiendo un concepto un tanto diferente de cómo debe interpretarse la categoría de *existencia*. Él entiende por *existencia* tan sólo la objetividad de la verdad matemática.

La existencia platónica para él se refiere a la existencia de un canon externo objetivo que no depende de nuestras opiniones individuales ni de nuestra cultura concreta. Este es un concepto que incluso piensa puede aplicarse a otros objetos diferentes de las matemáticas, tales como la moralidad y la estética. Bajo tal concepto coloca ejemplos específicos, entre ellos, el famoso *último teorema de Fermat*, (ninguna potencia enésima positiva de un número entero puede ser la suma de otras dos potencias enésimas positivas si n es un número entero mayor que 2), y el igualmente famoso *Conjunto de Mandelbrot*, ya citado, cuya solidez, para él, está fuera de duda pues su estructura se revela independientemente del matemático o del computador que lo examine. La existencia del *último teorema de Fermat* y del *Conjunto de Mandelbrot* sólo puede estar dentro del mundo platónico de las matemáticas. (3-59, 60).

Como puede verse, para Penrose, las formas matemáticas del mundo de Platón no tienen evidentemente el mismo tipo de existencia que los objetos físicos ordinarios, vale decir, una silla o una mesa; no tienen localización en el tiempo y en el espacio. Las nociones matemáticas objetivas son entidades intemporales, no debe considerarse que nacieron en el instante en que fueron humanamente percibidas por primera vez. Las espirales concretas del *conjunto de Mandelbrot* no alcanzaron su existencia en el instante en que fueron vistas por primera vez en la pantalla o en la impresora de un computador; ni surgieron cuando la idea general que hay tras él fue propuesta por primera vez por un ser humano; dichos diseños ya *existían* desde el principio de los tiempos, en el sentido potencial e intemporal con que necesariamente

se iban a revelar en la forma exacta en que hoy los percibimos, con independencia de qué momento o qué lugar eligiera cualquier ser perceptivo para examinarlos. (3-60). Aún más, para Penrose, la existencia matemática es diferente no sólo de la existencia física sino también de una existencia que es atribuida por nuestras percepciones mentales; pero con todo, hay una conexión misteriosa y profunda con cada una de estas otras dos formas de existencia: La física y la mental. (3-61). A estos dos mundos, el de las ideas platónicas de las matemáticas, y el físico, Penrose agrega, como se ve, un tercero, modificando la idea de Popper de su *mundo de la cultura*, al que prefiere llamar *mundo mental*. (2-80,81,82). Estos tres mundos están conectados y hay una dependencia profunda del mundo físico con respecto a leyes matemáticas exactas.

Esta estructura de *relación* y dependencia entre los tres mundos plantea un problema fundamental, el de aclarar la relación que se establece entre ellos, y tres grandes misterios: El primero es el de la relación que existe entre las matemáticas y la física. El segundo es el de la relación entre el mundo físico y el mundo de la mentalidad. Y el tercer misterio: El de la relación entre el mundo mental y el de las matemáticas, esto es, ¿qué subyace a nuestra capacidad de acceder a la verdad matemática? (2-81-82).

Pero Penrose reconoce en su planteamiento, a la par del problema de base y de los tres misterios citados, tres creencias de él mismo, a las que también llama prejuicios: El primero, (relación del mundo matemático-platónico con el mundo físico), es que el mundo físico entero no puede, en principio, describirse en términos matemáticos, aunque si las matemáticas se saben elegir correctamente, se puede con ellas describir adecuadamente y de forma muy precisa el mundo físico, de lo que deduce que este se comporta de acuerdo con las matemáticas. Una pequeña parte del mundo platónico engloba entonces a nuestro mundo físico. (2-111; 3-61).

El segundo prejuicio, (relación del mundo mental con el mundo físico), es que no se trata de sostener que todos los elementos del mundo físico, (y particularmente ciertas estructuras físicas como el cerebro humano vivo, sano y despierto), tienen su correspondencia en el mundo mental, sino más bien que no existen objetos mentales *"flotando ahí fuera que no estén basados en la física"*. *"Aunque el cerebro de un gato puede evocar realmente cualidades mentales, no estoy exigiendo lo mismo de una piedra"*, dice. (3-62).

Y el tercer prejuicio, (relación del mundo mental con el mundo matemático), es que cualquier objeto individual en el mundo platónico es accesible, en cierto sentido, a nuestra mentalidad, aunque considera evidente que sólo una pequeña fracción de nuestra actividad mental tiene que estar interesada

en la verdad matemática absoluta. (3-62). Así, pues, todo el mundo físico se encuentra gobernado de acuerdo con leyes matemáticas, por ecuaciones o por algunas nociones matemáticas futuras fundamentalmente diferentes de aquellas que hoy se etiquetan con el nombre de ecuaciones. (3-62).

A los tres prejuicios anteriores agrega tres más: El cuarto reside en que piensa que está adoptando una actitud científica demasiado fría al dibujar el diagrama de los tres mundos de una forma que implica que toda mentalidad tiene sus raíces en la fisicidad, pues, si bien no se tiene evidencia científica razonable de la existencia de mentes que no tengan una base física, no se puede estar seguro de ello, sobre todo, si se considera el pensamiento religioso, que sabe admitir vehementemente la posibilidad de mentes independientes de lo físico.

El quinto prejuicio reside en representar el mundo platónico dentro del ámbito de la mentalidad, con lo que se pretende indicar que, al menos en principio, no hay verdades matemáticas que estén más allá del alcance de la razón. (3-63,64).

Y el sexto y final, que engloba una paradoja: ¿Cómo es posible que cada mundo parezca englobar al siguiente en su totalidad? Este es, para Penrose, un misterio aún más profundo que trasciende a los otros, y plantea al mismo una posibilidad, el que los tres mundos no sean en absoluto independientes sino que meramente reflejen, individualmente, aspectos de una verdad más profunda sobre el mundo como un todo de la que se tiene muy poca idea en el momento presente. (3-6).

Penrose entonces plantea una variación al diagrama de los tres mundos en el sentido de dibujarlo admitiendo *tres posibles violaciones a sus prejuicios*. Tener en cuenta la posibilidad de acción física más allá del alcance del control matemático, la posibilidad de que pudiera haber mentalidad que no estuviera enraizada en estructuras físicas, y la posibilidad de que pudieran existir enunciados matemáticos verdaderos cuya verdad fuera en principio inaccesible mediante la razón y la intuición. (3-64)

Hay, pues, *"algo característicamente misterioso en la forma en que estos tres mundos se interrelacionan -en donde cada uno de ellos parece casi emerger de una pequeña parte de su predecesor"*. (2-111). Esta forma misteriosa de interrelación no reside sólo en su precisión sino también en su sofisticación sutil y en su belleza matemática. *"Hay asimismo un profundo e indudable misterio en cómo puede llegar a suceder que la materia física adecuadamente organizada -y aquí me refiero en concreto a cerebros humanos (o animales) vivos- pueda evocar de algún modo la cualidad mental del conocimiento consciente. Por último, hay también un misterio en cómo percibimos la verdad matemática. No se trata solamente de que nuestros*

cerebros estén programados para <calcular> de manera fiable. Hay algo mucho más profundo que eso en las intuiciones que incluso los más humildes de entre nosotros tenemos cuando apreciamos, por ejemplo, los significados reales de los términos <ceros>, <uno>, <dos>, <tres>, <cuatro>, etc.", dice. (3-64, 65).

Pero es posible, en opinión de Penrose, que dichos misterios puedan ser entendidos e incluso resueltos, al menos en parte, utilizando la noción matemática, -un tipo de matemática "no-local"- de cohomología, esto es, una noción que proporciona un medio de calcular el grado de imposibilidad de una figura. "Es importante reconocer los rompecabezas y los misterios cuando se presentan. Pero el solo hecho de que suceda algo muy enigmático no significa que nunca seremos capaces de comprenderlo", dice. (2-111)

II. ALGUNAS POSICIONES GNOSEOLÓGICAS.

II.1 El método científico y las herramientas de la ciencia.

Para Penrose, los extraordinarios avances en el conocimiento han sido adquiridos gracias a:

- a) Cuidadosa observación física;
- b) Soberbia experimentación;
- c) Razonamientos físicos de gran profundidad;
- d) Intuición; y
- e) Razonamientos matemáticos.

Tales avances, que han conducido al enorme progreso de la ciencia, han sido posibles al encontrar el equilibrio correcto entre las restricciones, tentaciones y revelaciones de la teoría matemática, por un lado, y la acción precisa de las acciones del mundo físico, por otra, normalmente mediante experimentos cuidadosamente controlados. Cuando la guía experimental está ausente, como en el caso de la investigación fundamental más actual, este equilibrio se rompe. (3-1351, 1356, 1357). Efectivamente, Penrose reconoce el equilibrio y la mutua complementariedad que debe existir entre la teoría y la experimentación, entre lo deductivo y lo inductivo. Sin embargo, se advierte en él que, como dice en (1-183), *"es la teoría física la que fundamentalmente sustenta nuestra tecnología"*.

II.2 La *belleza y los milagros* como parte del método científico.

Penrose es un físico teórico y un matemático, aunque también un filósofo, y de primera línea, "que no teme abordar problemas que sus contemporáneos

despachan considerándolos sin sentido", como dice Gardner en el prólogo a *La mente nueva del Emperador*, (1-12). Sus intereses científicos más importantes corren alrededor de los problemas actuales de la astrofísica y la cosmología. Malcom Longair, quien prologa su libro *Lo grande, lo pequeño y la mente humana*, (2), dice en el mismo que los trabajos más famosos de Penrose son los que se refieren a dos teoremas relacionados con las teorías relativistas de la gravedad, en los que trabajó junto con Stephen Hawking. Uno de ellos que afirma que dentro de un agujero negro debe existir inevitablemente una singularidad física, es decir, una región en la que la curvatura del espacio, o, en forma equivalente, la densidad de la materia, se hace infinitamente grande; el otro en el que se afirma que, de acuerdo con las teorías relativistas clásicas de la gravedad, existe inevitablemente una singularidad física similar en el origen de los modelos cosmológicos de *gran explosión*, resultados, ambos, que indican que en cierto sentido, estas teorías no están acabadas, puesto que las singularidades físicas deberían ser evitadas en cualquier teoría físicamente significativa. (2-8,9). Gardner, en el prólogo citado, coincide con Longair en lo anterior. (1-14).

Sin embargo, al margen de su formación eminentemente teórica, Penrose reconoce el importantísimo papel que las experiencias empíricas han jugado dentro del desarrollo de la ciencia y la tecnología. Esas, en buena medida, dice, se han derivado de la experimentación y la observación empírica, pero ello no quita que él reconozca que el papel principal en tal desarrollo corresponde a la teoría, como ya hemos apuntado.

Penrose reconoce, pues, el papel central que dentro del mismo han tenido y tienen la inducción y la deducción, la teoría y la práctica empírica, pero agrega que deben incluirse como parte del método científico otras formas de conocimiento: La intuición, cuestión ya aceptada por muchos otros científicos, y la belleza y los milagros. Resulta un tanto extraño admitir a la belleza y a los milagros como partes del método científico, pero viniendo tal posición de una persona como Penrose, debe darse crédito, al menos, para debatir y discutir tal asunto. *"Dos poderosas fuerzas impulsoras internas han tenido una gran influencia en la dirección de la investigación teórica, pese a que normalmente pasan desapercibidas en escritos teóricos serios, por miedo, sin duda, a que pudiera parecer que dichas influencias se hayan desviado demasiado de las reglas estrictas del método científico correcto. La primera de estas es la belleza, o elegancia,..... En cuanto a la segunda, a saber, el irresistible atractivo de lo que frecuentemente se califica de 'milagros', tan sólo la he sugerido.....; pero, como puedo dar fe por mi experiencia personal, estos pueden ejercer realmente una poderosa influencia en la dirección de la propia investigación de cada uno".* (3-1389).

En su opinión, los criterios estéticos son importantes en la formación de los juicios, y pone como ejemplo el cómo el *agudo sentido de la belleza* de

Dirac le hizo capaz de descubrir su famosa ecuación para el electrón, considerada una de las diez ecuaciones más importantes de la ciencia actual, y que en vano habían tratado de encontrar muchos otros científicos. A Penrose le parece evidente que la importancia de los criterios estéticos no sólo se aplica a los juicios de inspiración instantáneos sino a los juicios mucho más frecuentes que se hacen en el trabajo matemático y científico. (1-496, 497).

"Muchas de las ideas que se cree que han impulsado un avance importante en la teoría física también se consideran de una belleza cautivadora". (3-1389). Y coloca como ejemplos, en este caso, a la geometría euclídea, la dinámica de Newton, el electromagnetismo de Maxwell, la relatividad general de Einstein, la estructura de la mecánica cuántica, el espín mecanicocuántico, la ecuación de ondas relativista de Dirac, el formalismo de integrales de camino para la QFT, la teoría matemática de Cantor, el teorema de incompletud de Godel, la teoría matemática de categorías, el último teorema de Fermat, la exquisita teoría de Riemann de la función.

En cuanto a los milagros, cita Penrose, como producto de los mismos, la teoría de la supergravedad, y las simetrías espejo. (3-1389 a 1395).

II.3 Refutación del Falsacionismo de Popper.

Penrose ofrece una clara refutación del *Principio de Falsabilidad* de Popper, que, en mi opinión, lo anula ya definitivamente, y lo hace tomando como ejemplos los casos de la *supersimetría en la moderna física de partículas*, *la teoría del monopolio magnético de Dirac*, y *la teoría de la homogeneidad e isotropía del universo*.

Karl Popper ofreció un criterio que parecía razonable para la admisibilidad científica de una teoría propuesta, a saber, que sea *refutable observacionalmente*. Como decimos, Penrose no admite que refutar *observacionalmente* lleve a admitir científicamente una teoría. La teoría de la *supersimetría* en la moderna física de partículas, que es para él una idea teórica con una evidente elegancia matemática, ingrediente fundamental además en la construcción de QFT y en la teoría de cuerdas, y considerada parte del modelo estándar en la física de partículas actual, no tiene, sin embargo, ningún apoyo experimental serio.

Veamos lo que expone Penrose. (3-1366): *"La teoría predice supercompañeras para todas las partículas observadas en la naturaleza, pero ninguna de aquellas ha sido observada hasta ahora, debido al efecto de un mecanismo de ruptura de simetría, de naturaleza desconocida actualmente, que hace que las supercompañeras sean tan masivas que las energías necesarias para crearlas están más allá de las posibilidades de los aceleradores actuales. Con posibilidades de energía aumentadas, las supercompañeras podrían*

encontrarse, con ello se habría conseguido un hito en la teoría física con implicaciones importante para el futuro. Pero supongamos que las supercompañeras siguen sin encontrarse. ¿Refutaría esto la idea de la supersimetría? En absoluto. Podría argumentarse (y probablemente se haría) que había habido demasiado optimismo sobre el grado de la ruptura de la simetría, y que se necesitarían energías aún mayores para encontrar las supercompañeras perdidas". Aunque hay un carácter popperiano en dicho modelo, ello no la hace inaceptable como teoría científica.

En cuanto a la teoría del monopolio magnético del físico inglés Dirac, de nuevo el argumento de Penrose es el siguiente: *"El argumento de Dirac es que la mera existencia de un único monopolio magnético en alguna parte del cosmos proporcionaría una explicación al hecho de que cada partícula del universo tiene una carga eléctrica que es un múltiplo entero de un valor dado (como se observa en la realidad). La teoría que afirma que tal monopolio existe en 'alguna parte' es típicamente apopperiana. Dicha teoría podría ser establecida mediante el descubrimiento de tal partícula, pero parece que no es 'refutable' como requeriría el criterio de Popper; pues si la teoría es falsa, ¡por mucho que los experimentadores busquen en vano, su incapacidad para encontrar un monopolio no refutaría la teoría! Pese a todo, se trata de una teoría científica, perfectamente digna de una consideración seria". (3-1366, 1367).*

Finalmente, Penrose coloca el ejemplo de la cosmología. De nuevo, citamos: *"La región del universo que se encuentra fuera de nuestro horizonte de partículas está más allá de la observación directa. Pese a todo parece una propuesta científica razonable la de que dicha región debería asemejarse, en una amplia escala, a la región que es accesible a la observación directa. La teoría de que la región inobservable se parece a la región observacionalmente accesible, que, de hecho, es parte del modelo estándar de la cosmología, aunque no de la mayoría de los esquemas inflacionarios, no es refutable observacionalmente". (3-1367)*

Penrose recurre aquí, como efecto definitivo de su refutación a Popper, al caso de las tres posibilidades de la geometría espacial que corresponden al universo observable: Curvatura positiva, negativa o nula, que saben expresarse como $K > 0$, $K < 0$ o $K = 0$, respectivamente. *"Cuando $K = 0$, la teoría afirma su carácter de refutable observacionalmente. Pero si según la teoría, $K \neq 0$ pero la realidad fuera $K = 0$, entonces dicha teoría no podría ser refutada porque siempre habría algún margen de incertidumbre en las observaciones que no descartaría una muy ligera curvatura espacial positiva o negativa. Notemos que el caso $K > 0$ podría en principio ser refutado si en realidad $K < 0$, y $K < 0$ podría ser refutado si en realidad $K > 0$. Por el contrario, $K = 0$ no puede ser confirmado directamente, mientras que $K \neq 0$ podría ser confirmado observacionalmente, (si el universo resultara ser de esa forma). Por ello,*

las dos afirmaciones $K>0$ y $K<0$ son popperianas en el sentido restringido de que son refutables en ciertas circunstancias -aunque no son susceptibles de ser refutadas sí, en realidad, $K=0$ - y también son confirmables individualmente. ¡Notemos que $K=0$ es completamente popperiana, en principio, pero no confirmable!'. (3-1367,1368).

II.4 Clasificación de las Teorías.

Penrose hace una interesante clasificación de las teorías en cuatro clases: Extraordinarias, o Supremas; Útiles; Provisionales; y Erróneas, o Descaminadas. (1-13).

Las teorías Supremas, o Extraordinarias, son aquellas que, sin exigir su aplicación sin refutación a los fenómenos del mundo, sí exigen que el alcance y la exactitud con que se aplican sea excepcional, en el sentido apropiado. *“Resulta notable que todas las teorías supremas de la naturaleza han resultado ser asaz fértiles como fuente de ideas matemáticas, lo que sin duda nos dice algo profundo sobre la conexión entre el mundo real de nuestra experiencia física y el mundo platónico de las matemáticas”, dice. (1-212).* Son ejemplos de ellas, la geometría euclidiana, la teoría de la selección natural de Darwin y Wallace, la estática y la dinámica newtonianas, cuyo nacimiento dio lugar al cálculo infinitesimal, y que por efecto de difusión en el espacio no puede ser verdadera en nuestro mundo, el electromagnetismo de Maxwell, la teoría de la relatividad de Einstein, la teoría cuántica, y la electrodinámica cuántica.

Es enorme la exactitud de las teorías físicas que Penrose reconoce y clasifica como Supremas o Extraordinarias. La geometría euclidiana fue desde el principio una teoría física suprema, además de ser un elemento elegante y lógico de la matemática pura. (1-185 a 190). Es exacta con un margen de error menor que el diámetro de un átomo de hidrógeno sobre un rango de un metro, y su falta de completa exactitud se debe a los efectos de la relatividad general. La mecánica newtoniana es exacta hasta exactamente una parte en 107, aunque no es totalmente exacta debido, de nuevo, a los efectos de la relatividad. La electrodinámica de Maxwell es válida en un enorme rango de escalas que va desde los tamaños de las partículas fundamentales, donde se utiliza en combinación con la mecánica cuántica, hasta las dimensiones de las galaxias distantes, lo que corresponde a un rango de escalas de 1035 o más. La teoría de la relatividad de Einstein puede decirse que es exacta hasta casi una parte en 1014, aproximadamente el doble de cifras significativas que la mecánica newtoniana, por lo que se considera que esta queda incluida dentro de aquella. La mecánica cuántica, por su parte, es una teoría extraordinariamente exacta, (2-50), a tal grado que permite explicar cosas como la estabilidad de los átomos, las líneas

espectrales, las fuerzas químicas, la radiación del cuerpo negro, la fiabilidad de la herencia, el funcionamiento de los láseres, los superconductores y superfluidos. (2-51,52).

Entre las teorías Útiles coloca el modelo de los Quarks de Gell Mann-Zweig para los hadrones, la cromodinámica cuántica, la teoría de los leptones de Glashow, Salam, Ward y Weimberg, y la teoría del *big bang o gran explosión* sobre el universo. También es una teoría Útil la del movimiento planetario de Eudoxo, que duró más de 2000 años, y que fue posteriormente desarrollada por Hiparco y Ptolomeo. (1-185 a 190)

Son teorías Provisionales las de Kaluza-Klein, la de la supersimetría o supergravedad, las de las cuerdas y de las supercuerdas, las GUT, la del escenario inflacionario, e incluso su propia teoría de los *twistor*: (1-185 a 190). Para Penrose, la teoría de cuerdas, que es uno de los soportes más utilizado por los físicos teóricos para acceder a una teoría del todo, es sólo una teoría provisional, pues no existe, al menos por el momento, ninguna razón observacional para creer que ella, en particular, sea física, aunque esté ciertamente motivada por poderosas aspiraciones físicas; el que haya sido estudiado por muchos físicos no la hace física, aunque sea probablemente por ahora la mayor de las modas en la investigación física fundamental.

Una teoría puede pasar, en el tiempo, de una a otra clasificación, o incluso desaparecer de las clasificaciones porque hayan sido olvidadas o se hayan desvanecido, o porque se hayan incorporado. Ejemplo de las primeras son el sistema ptolemaico, que fue por tantos años una teoría Útil pero que se disolvió posteriormente como teoría física; otros ejemplos de teorías Útiles del tipo finalmente acertado son la concepción del movimiento planetario elíptico de Kepler, y la tabla periódica de Mendeleiev. (1-185 a 190)

Penrose no clasifica a la Termodinámica dentro de sus teorías porque no la considera una teoría física, aunque parece que el argumento anterior lo utiliza como una excusa para evitar definir si una teoría es Suprema, o si simplemente es una teoría Útil. Y probablemente sólo clasifica al *big bang* como Útil debido a que considera que existe en ella una singularidad, por lo que no es una teoría acabada puesto que las singularidades físicas deberían ser evitadas en cualquier teoría físicamente significativa, según afirma en *Lo grande, lo pequeño y la mente humana*. (2-8).

Aunque la teoría de la gran explosión proporciona la imagen más convincente que se tiene para comprender las características a gran escala de nuestro universo, es incompleta en varios aspectos: La imagen resulta correcta sólo si se fijan las condiciones iniciales de una forma muy cuidadosa, ya que antes de una milésima de segundo, el comportamiento del universo se aparta de

la física ensayada y comprobada, y por ello se basa en extrapolaciones razonables de las leyes conocidas de la física. La pregunta no es cuáles fueron esas condiciones iniciales, sino porqué fueron así, con lo que Penrose entra a ubicar el problema en el plano especulativo estrictamente filosófico. Rechaza así la imagen inflacionaria del universo en sus etapas primitivas, y deja la respuesta futura a la *teoría cuántica de la gravedad*. (2-8,9). *En la gran explosión, dice, todo falla, exactamente, en el comienzo*. (2-38)

III. EL ACCESO DEL HOMBRE A LA REALIDAD.

III.1 Relatividad general y teoría cuántica, ¿Teorías certeras sobre la realidad?

Penrose, en oposición a físicos teóricos que creen que ya se está cerca de una teoría del todo que sea capaz de explicar adecuadamente la naturaleza fundamental del mundo físico, y que saben apoyarse en las diferentes versiones sobre las que se va actualizando la teoría de cuerdas, cree que todavía se está muy lejos de tal *teoría final*. (3-1351,1352). Ciertamente, de alguna manera la teoría de cuerdas ha permitido avances en el conocimiento de la realidad, y también otras teorías, pero piensa que "*sería imprudente predecir con mucha confianza que estas teorías estén a punto de dar los pasos adicionales necesarios que nos pudieran guiar hacia el verdadero camino a la comprensión de la realidad física*". Para él, la mayor contribución ha provenido de la relatividad general de Einstein, aunque también han contribuido grandemente la teoría cuántica y la teoría cuántica de campos, QFT, aunque sostiene que estas últimas aún no tienen la coherencia necesaria debido al problema de la *paradoja de la medida*, lo que las hace teorías incompletas. Sin embargo, mantiene la esperanza de que tal incompletud estará resuelta en algún momento del siglo XXI, con lo cual la contribución cuántica a la explicación de la realidad será aun mayor que la de la relatividad general. (3-1352,1353).

La relatividad general fue, en gran medida, la *teoría de una persona*, al margen de las aportaciones esenciales que Einstein recibió de Lorentz, Poincaré, Mach, Minkowski y otros; la teoría cuántica fue por el contrario, la teoría de muchas personas, impulsada externamente por los extraordinarios resultados de cuidadosos experimentos. (3-1373). Penrose tiene la sensación, debido al escaso impulso experimental de la mecánica cuántica, de que el progreso futuro de la ciencia será producto de algo parecido a la aproximación einsteiniana de *una persona*. (3-1373).

III.2 El mundo platónico de las ideas matemáticas, ¿una teoría del todo?

Pero, ¿Qué es la realidad física? Los científicos modernos tratan siempre de

evitar la respuesta al *qué*, y prefieren responder en términos de *cómo se comporta*. Acoto aquí la debilidad fundamental del enfoque anterior, pues la realidad *no se comporta* sino simplemente *es*. Podría aceptarse replantear la pregunta de la ciencia en términos de *cómo es la realidad*, y aquí Penrose considera tal pregunta una cuestión fundamental. Él la replantea así: *¿Cómo describimos las leyes que rigen nuestro universo y sus contenidos?*(3-1375), aun y reconociendo que así planteada se deja sin respuesta otra pregunta aun más fundamental todavía: *¿Qué es lo que se está comportando así?*, y aún más, *¿porqué se está comportando así?* Pero, dice, *sin saber qué es la realidad, es difícil ver porqué se comporta como lo hace*. (3-1376).

Penrose reconoce la limitación de la ciencia y del conocimiento en cuanto a su capacidad de dar respuesta a las preguntas por el *qué* y por el *porqué*. Como se sabe, ello corresponde más bien a la filosofía, y aquí, transitando hacia esta, el gran físico y matemático vuelve de nuevo a buscar respuesta en el mundo platónico de las ideas. Cuando los físicos modernos describen las cosas invariablemente en términos de modelos matemáticos, lo que hacen en el fondo es *"tratar de encontrar la realidad dentro del mundo platónico de las ideas matemáticas"*, que es, en el fondo, una especie de teoría del todo, dice, pues entonces la realidad física aparecería como un reflejo de leyes puramente matemáticas. (3-1376). Esta posición, como se ha visto, constituye el primero de los profundos misterios en su teoría de los tres mundos.

Penrose resume y concreta su posición diciendo que ciertamente el mundo de Platón proporciona un tipo de realidad para las nociones matemáticas, aunque sin embargo no puede identificarse la realidad física dentro de la realidad abstracta del mundo de Platón. *"Cada uno de los tres mundos - matemático-platónico, físico y mental- tiene su propio tipo de realidad, y donde cada uno está basado (profundamente y misteriosamente) en el que le precede (tomando los mundos cíclicamente) "*. (3-137). Ciertamente él quisiera pensar que el mundo platónico es el más fundamental de los tres, pero existe el misterio, o la paradoja, adicional del aspecto cíclico de estos mundos, donde cada uno de ellos parece capaz de englobar en su totalidad al que le sigue, aunque en sí mismo parezca depender solo de una pequeña parte de su predecesor. (3-1378).

IV. EL UNIVERSO, EL PROBLEMA DE LA CONCIENCIA, EL PRINCIPIO ANTRÓPICO Y EL DETERMINISMO UNIVERSAL.

IV.1 ¿El universo y la conciencia como actos de la Creación?

El universo se origina, dice Penrose, a partir de un gas difuso del cual se condensaron las estrellas, y que fue ´escupido´, según la ´teoría estándar´,

como consecuencia de la *Gran Explosión*, el "*big bang*". Sin embargo, no fue esa una explosión ordinaria en la que el material se expele a partir de un punto central hacia el espacio preexistente: Aquí, el propio espacio es '*creado*' en la explosión, y no existe, ni existió, punto central. (1-388). No hay, entonces, '*espacio vacío preexistente*', en el cual se '*vacíe*' la materia producida por la explosión; el mismo espacio, es decir, la superficie del globo, nace de la explosión. (1-389).

Penrose enlaza así el problema de la creación del universo con el origen y la razón de ser de la conciencia y el papel del hombre dentro de él. ¿No acaso lo anterior podría servir como una nueva expresión del hecho bíblico de la Creación, en la cual, todo, incluso el mismo espacio, fue creado de la nada, e incluso la misma conciencia en el hombre aparece como producto de un solo acto único y unitario? Aquí, él expone un argumento importante cuando afirma que la conciencia no tiene porqué ser producto de la selección natural, con todo y el respeto que le tiene a esta teoría en sus expresiones de Darwin y Wallace, y que él cataloga como una de sus "*teorías supremas*" dentro de su clasificación de las teorías.

IV.2 La necesidad del Principio Antrópico para explicarse el fenómeno de la evolución de la conciencia.

¿Qué importancia tiene la conciencia para el universo en su totalidad? ¿Podría existir un universo sin habitantes conscientes? ¿Están las leyes de la física especialmente diseñadas para permitir la existencia de vida consciente? ¿Hay algo especial en nuestra localización particular en el universo, ya sea en el espacio o en el tiempo? Estas son las preguntas claves que se hace el *Principio Antrópico*, que en su forma '*débil*' se refiere a la localización espacio-temporal de la vida consciente o inteligente. El *Principio Antrópico* en su forma '*fuerte*' va más lejos, pues no sólo se interesa en nuestra localización espacio-temporal en el universo sino de una infinidad de universos posibles. En opinión de Penrose, este *Principio Antrópico Fuerte* tiene un carácter algo dudoso, y los físicos tienden a invocarlo cuando no tienen una teoría lo bastante buena para explicar los hechos observados.

El *Principio Antrópico Débil*, por el contrario, le parece incuestionable siempre que se utilice cuidadosamente. El *Principio Antrópico*, en cualquiera de sus expresiones, ayuda a probar que la conciencia es inevitable en virtud del hecho de que tendría que haber seres sensibles, es decir, '*nosotros*', para observar el mundo, lo cual implica que no es necesario suponer que la sensibilidad tenga alguna ventaja selectiva. Al menos, dice, la versión '*débil*' podría proporcionar una razón para que la conciencia exista sin que tenga que ser favorecida por la selección natural. "*Cualquier universo que pueda ser observado debe ser capaz, como necesidad lógica, de soportar la*

mentalidad consciente, puesto que la conciencia es precisamente la que desempeña el papel final como 'observador', dice. (3-1378). Sin embargo, Penrose no cree que sea el Principio Antrópico la razón auténtica, o al menos no la única razón para la evolución de la conciencia. Para él hay suficientes pruebas procedentes de otras direcciones para convencerse de que la conciencia tiene una poderosa ventaja selectiva, y no cree que el Principio Antrópico sea necesario. (1-510,511,512).

IV.3 El Principio Antrópico y el asunto de los muchos universos.

Penrose enfoca el asunto de los *muchos universos* en relación con los observadores, la objetividad y los estados de conciencia. *"El punto de vista de los 'muchos universos' es manifiestamente dependiente de tener una comprensión adecuada de lo que constituye un 'observador consciente', puesto que cada 'realidad percibida' está asociada con un 'estado de observador', de modo que no sabemos qué estados de realidad (i. e., mundos) están permitidos hasta que sepamos qué estados de observador están permitidos. Dicho de otra forma, el comportamiento del mundo aparentemente objetivo que es realmente percibido depende del camino que siga la conciencia de uno a través de las miríadas de alternativas cuánticamente superpuestas. En ausencia de una teoría adecuada de los observadores conscientes, la interpretación de los muchos universos debe necesariamente seguir estando fundamentalmente incompleta", dice. (3-1380).*

IV.4 El determinismo universal, la insuficiencia de la física y la naturaleza necesariamente no algorítmica del enlace de los mundos clásico y cuántico.

En cuanto al universo y su carácter determinista, Penrose reconoce no sólo esta naturaleza determinista del universo sino de los *'muchos universos'* posibles. Ello le lleva a confirmar su teoría de que el gobierno del universo no es, posiblemente, algorítmico, de lo cual se hablará posteriormente. El *determinismo fuerte*, dice, no deja lugar o espacio para el *"libre albedrío"*, pero sin embargo, debe existir algún nuevo procedimiento que permita establecer una conexión entre la línea divisoria cuántico-clásica que contenga un elemento esencialmente no algorítmico, y que implicaría que el futuro no sería computable a partir del presente, incluso aunque pudiera estar determinado por él. Recordemos que para él, la mecánica cuántica es determinante y precisa, y su indeterminación aparece sólo cuando se realiza la medida, lo que implica amplificar el suceso desde el nivel cuántico hasta el nivel clásico; es decir, son el observador y sus instrumentos quienes ponen la indeterminación, y no el propio mundo cuántico en sí.

En el *determinismo* fuerte no sólo el futuro está determinado por el pasado sino que la historia entera del universo está fijada, según un esquema matemático preciso, para cualquier instante, incluso llegando a considerar la idea de los muchos universos, en la que no sólo la simple historia del universo estaría matemáticamente fijada sino la totalidad de miríadas de miríadas de historias del universo posibles. Para Penrose, si se tiene determinismo fuerte pero sin muchos universos, entonces el esquema matemático que gobierna la estructura del universo tendría que ser posiblemente no algorítmico, lo cual fundamenta su posición en relación con el problema mente-cuerpo y la propia posición de la Inteligencia Artificial Fuerte, que reduce todo a lo algorítmico, y que él definitivamente rechaza, como se verá luego. (1-509,510).

Cabe hacer aquí algunas puntualizaciones en cuanto a cómo Penrose enfoca el problema de la física y su explicación de la realidad. Para Penrose, la física no ha llegado a dominar la cualidad profunda en la naturaleza misma de la materia que determina como esta debe comportarse. La física, dice, tiene aun muchos misterios que desentrañar y muchas intuiciones que obtener. La tecnología del mundo moderno se ha derivado, en buena medida, de una gran riqueza de experiencias empíricas; sin embargo, es la *teoría física* la que fundamentalmente sustenta nuestra tecnología. (1-182,183).

La física clásica admite la existencia de un *mundo objetivo ahí afuera*. Lo mismo dice el sentido común. Este es un mundo determinista, independiente de nosotros mismos y de cómo lo observamos. La filosofía recoge este pensamiento y lo admite y respalda. Pero la teoría cuántica nos obliga a cambiar tal visión del mundo y de la realidad física, pues sólo ella posibilita el obtener una visión más correcta de dicho mundo y de la realidad física, incluso de la realidad de la misma conciencia, aunque siempre insuficiente e inadecuada. Físicos como Niels Bohr no admitieron nunca la existencia de una imagen objetiva del mundo; *en el mundo cuántico no hay nada* ahí afuera, es cosa sólo de las mediciones. La teoría cuántica proporciona sólo un procedimiento de cálculo y no intenta de manera alguna describir el mundo como realmente *es*. Incluso Einstein nunca pudo aceptar que la teoría cuántica pudiera ser algo más que una mera descripción provisional del mundo físico; y de alguna forma, lo mismo pensaron Schrodinger, de Broglie y Dirac. (1-336,358).

Penrose no acepta la posición de Bohr; la llama *derrotista*. Él si otorga una *realidad física objetiva* al mundo cuántico. (1-270 a 302). La mecánica cuántica capacita para comprobar si algo podría haber sucedido pero no sucedió. Comprueba lo que los filósofos llaman *hechos contrafácticos*, y además permite que ¡efectos reales resulten de hechos contrafácticos! (2-61). El estado cuántico, pues, proporciona una imagen objetiva, aunque complicada y algo paradójica, del mundo real. Pero como los estados cuánticos

en su mayoría no se parecen en absoluto a los clásicos, puede que su teoría resulte errónea cuando se aplique a cuerpos macroscópicos. Penrose, reconociendo lo anterior, piensa que es necesario encontrar una nueva ley para poder describir adecuadamente el mundo físico y su realidad objetiva, incluidos aquí el conocimiento de las mentes y de las conciencias, ley que permita enlazar los mundos cuántico y clásico. (1-359). Es la comentada *Teoría del Todo*. En su opinión, es la *Teoría de la Gravitación Cuántica* la que podría ayudar a resolver el problema. (1-414), y, en contra de la mayoría de físicos, piensa que es la Relatividad General la que proporcionará los elementos para lograrlo.

V. INTELIGENCIA ARTIFICIAL. EL PROBLEMA MENTE-CUERPO Y EL ORIGEN DE LA CONCIENCIA.

El problema mente-cuerpo, viejo dilema filosófico que adquiere características paradigmáticas en la filosofía moderna, a partir de Descartes, es enfocado por Penrose en toda su obra, pero principalmente en *La Mente Nueva del Emperador* (1) y en *Las Sombras de la Mente*, (4). Constituye uno de los ejes centrales de su filosofía, en relación con el asunto de la conciencia, y es un aporte grande y calificado al análisis y la comprensión del asunto. En una conclusión anticipada, pareciera que Penrose considera insostenible el punto de vista, aparentemente prevaleciente en la corriente filosófica actual, de que nuestro pensamiento es básicamente lo mismo que la acción de una computadora muy sofisticada. Según Gardner, quien prologa el primero de los libros citados, este es *"el ataque más poderoso que se haya escrito contra la Inteligencia Artificial Fuerte"*, IA fuerte. (1-12). Tratando de exponer lo más ordenadamente posible la posición de Penrose en un caso tan complejo como éste, veamos previamente un poco cuál es la posición de la IA fuerte y de qué manera él la va analizando:

V.1 Penrose y la Inteligencia Artificial Fuerte en relación con el problema mente-cuerpo.

Los defensores de la Inteligencia Artificial (IA) fuerte creen que los algoritmos forman la sustancia de los pensamientos, de los sentimientos, del entendimiento y de la percepción consciente. Ello los lleva a cierta forma de dualismo que en el fondo tratan de evitar. (1-38, 39, 40). Esta posición es de alguna manera aceptada incluso por algunos opositores de la IA fuerte, como el filósofo norteamericano Searle, para quien se pueden identificar los algoritmos con los estados mentales, pero reconociendo una diferencia que pudiera considerarse fundamental: La diferencia entre la función del cerebro humano, (que puede alojar la mente), y las computadoras electrónicas, (que, según él, no pueden hacerlo), pudiendo ambos ejecutar el mismo algoritmo, radica solamente en la construcción material de cada uno: Los objetos biológicos,

(cerebros), pueden poseer *intencionalidad y semántica*, lo que él considera como las características definitorias de la actividad mental, mientras que los electrónicos no.

A Penrose, esta tesis le parece "*sospechosamente dogmática*", quizá no menos dogmática que las afirmaciones de la misma IA fuerte que sostienen que la simple ejecución de un algoritmo puede producir un estado de conciencia. Penrose piensa que Searle *"ha sido confundido por los computólogos y los informáticos, quienes a su vez han sido confundidos por los físicos, que parecieran creer que cualquier cosa es una computadora digital"*. (1- 38, 39, 40). *Stephen Hawking rechaza en alguna forma la posición de su compañero y amigo Penrose, cuando dice, (2-134): "Personalmente me siento incómodo cuando las personas, en especial los físicos teóricos, hablan sobre la conciencia. La conciencia no es una cualidad que uno pueda medir desde fuera.... Yo prefiero hablar de inteligencia, que es una cualidad que puede medirse desde fuera, y no veo ninguna razón por la que la inteligencia no pueda ser simulada en un ordenador. Nosotros, ciertamente, no podemos simular inteligencia humana por el momento, como demostró Roger con su problema de ajedrez. Pero Roger admitió también que no existía línea divisoria entre inteligencia humana e inteligencia animal"*. Sólo cabe recordar aquí la posición tan radicalmente opuesta a lo especulativo y al papel de la filosofía en el enfoque de la realidad, así como su posición reduccionista de las ciencias a la física, que ha mantenido siempre Hawking.

Esto pareciera ir en concordancia con la interpretación y el papel que se otorga al hardware y al software en la IA fuerte: Estableciendo una equivalencia entre dispositivos físicos de cómputo, la IA fuerte considera al hardware de relativamente poca importancia, (quizás incluso totalmente irrelevante), y en cambio, el software, esto es, el programa, los algoritmos, es un ingrediente vital. (1-41). El hardware, pues, no es importante en los fenómenos mentales, lo cual a Penrose le parece algo injustificado, pues se basa en la presunción de que el cerebro, o la mente, es una computadora digital, y supone que cuando pensamos no está en juego ningún fenómeno físico concreto que requiera estructuras físicas concretas, (biológicas o químicas), como las que tienen los cerebros. El punto de vista de la IA fuerte sostiene que al tratarse solamente de una cuestión de hardware, toda la física implicada en el funcionamiento del cerebro puede ser simulada mediante la introducción de un software convertidor apropiado. Concretamente, para ellos, el cerebro actúa de acuerdo con algún tipo de algoritmo, (1-46,47), y la mente es la encarnación de un algoritmo suficientemente complejo activado por algunos objetos del mundo físico, (1-506). Penrose objeta y se opone a lo anterior. Para él, las mentes conscientes no son entidades algorítmicas. (1-506).

Veamos ahora un poco más detalladamente las objeciones de Penrose a la posición de la IA fuerte, y su misma posición al respecto del problema mente-cuerpo y el origen de la conciencia.

A Penrose, la conciencia le parece un fenómeno de tal importancia que no puede creer que sea producido *"accidentalmente"* por un cómputo complicado. Pero aún más: Según su punto de vista, tiene que haber algo de naturaleza no computable en las leyes físicas que están por venir, y dicha posición la fundamenta y respalda en el teorema de Godel, que implica que la indemostrabilidad formal de una cierta proposición matemática es señal de que de hecho es verdadera. De ello concluye Penrose que nuestro pensamiento, al menos el matemático, tiene *"componentes no computables"*. Si se admite que existen procesos físicos no computables, habrá que ver cómo el cerebro hace uso de ellos. Penrose argumenta entonces que existe una relación directa entre esta no computabilidad y el puente entre el nivel cuántico y el nivel clásico, de lo cual se hablará posteriormente, que a su vez se relaciona con el proceso de medida cuántica ya mencionado, y el lugar en que el cerebro aprovecha los efectos de coherencia cuántica para acoplarlos a la actividad neuronal que se observa a gran escala en él parece ser los *"microtúbulos de Stuart Hameroff"*, que forman parte del citoesqueleto celular.

¿Es necesario recurrir a tal aparato para explicar el origen de la conciencia?, se pregunta. Para él, la conciencia *"es el fenómeno en el que se hace conocida la misma existencia del universo"*. Un universo gobernado por leyes que no permiten la conciencia no es un universo en absoluto. Todas las descripciones matemáticas del universo que se han dado hasta ahora incumplen este criterio. Es sólo el fenómeno de la conciencia el que puede conjurar un presunto universo "teórico" a la existencia real. La mente consciente no puede trabajar como una computadora, ni siquiera pese a que mucho de lo que interviene en la actividad mental podría hacerlo. Hay que plantear y contestar las siguientes preguntas: ¿Qué sucede a cada uno de nuestros flujos de conciencia después de morir? ¿Dónde estaba antes de que naciera cada uno? ¿Podríamos convertirnos en otro, o haber sido algún otro? ¿Por qué percibimos? ¿Por qué estamos aquí? ¿Por qué hay un universo en el que podamos estar? Estos son enigmas que tienden a llegar con el despertar de la conciencia en cualquiera de nosotros -y sin duda con el despertar de la verdadera autoconciencia, dentro de cualquier criatura u otra entidad en que primero llegue. Si una máquina es un ser pensante, sentimental, sensible, comprensivo, consciente, entonces comprar esa máquina, adquirirla, disponer de ella, implicará responsabilidades morales. El simple hecho de ponerla en marcha para satisfacer nuestras necesidades sin tener en cuenta su propia sensibilidad ya sería censurable, sería como maltratar a un esclavo. Desconectarla, o venderla, cuando había ya llegado a sentirse muy unida a nosotros, nos plantearía dificultades morales. *"Tendríamos que evitar causarle a la*

computadora el dolor que se supone es capaz de sentir”, dice. (1-23). Penrose rechaza que la IA pueda sentir placer o dolor. Para él, *“hay mucho más que entender de las cualidades mentales de lo que puede obtenerse de la IA”*. (1-23).

Penrose acepta que la individualidad de una persona no tiene nada que ver con la individualidad que pudieran atribuirse a sus constituyentes biológicos, sino más bien está relacionada con la configuración espacio-temporal de dichos constituyentes. Los defensores de la IA fuerte incluso agregan que si la información contenida en tal configuración puede ser transferida a otra forma desde la cual pueda ser recuperada, entonces la individualidad de una persona debiera permanecer intacta. Para ellos, es evidente que la individualidad de una persona puede ser tratada del mismo modo, y que no se pierde nada de ella si su forma física es transferida a algo bastante diferente, e, incluso, la conciencia de una persona persistiría aunque su *información* esté en otra forma.

Desde este punto de vista, una *conciencia humana* debe considerarse en realidad como un elemento del software y su manifestación particular como ser humano material debe considerarse como la ejecución de este software por el hardware de su cerebro y de su cuerpo. (1-43,44). Penrose cree que estas consideraciones son extravagantes aunque hay en ellas algo significativo respecto a la naturaleza física de la conciencia y de la individualidad humana: Proporcionan un indicador sobre la importancia de la mecánica cuántica en la comprensión de los fenómenos mentales. (1-45,46). Para él, son importantes en el modo de operar del cerebro: Los fenómenos cuánticos, los fenómenos electromagnéticos, y la teoría de la relatividad. (1-183,184). Veamos un poco cómo argumenta lo anterior.

V.2 Conciencia, Cerebro y Teoría Cuántica.

La teoría cuántica obliga a cambiar el concepto que se tiene de realidad física. Esta requiere de dicha teoría para una explicación satisfactoria, incluso también del fenómeno de la conciencia, que no puede entenderse satisfactoriamente en términos clásicos. *“Talvez nuestras mentes son cualidades arraigadas en una extraña y misteriosa característica de las leyes físicas que gobiernan ´realmente´ el mundo en que vivimos, en lugar de ser simples características de algún algoritmo ejecutado por los llamados ´objetos´ de una estructura física clásica. Quizás en cierto sentido, esta sea la ´razón´ de porqué debemos vivir; en tanto que seres sensibles, en un mundo cuántico en lugar de en uno enteramente clásico, a pesar de toda esa riqueza y de todo ese misterio que está presente en el universo clásico. ¿Sería necesario un mundo cuántico para que pudieran formarse, a partir de su sustancia, criaturas pensantes y perceptivas como nosotros mismos? Esta pregunta*

parece más apropiada para Dios, intentando construir un universo habitado, más que para nosotros”, dice Penrose. Para él, si la conciencia no puede formar parte de un mundo clásico, entonces nuestras mentes deben depender, de algún modo, de las desviaciones concretas respecto a la física clásica. (1-271). Ni la mecánica clásica, dice, ni la mecánica cuántica, pueden explicar la forma en que pensamos, aunque, creyendo él que la conciencia es algo global, cualquier proceso físico responsable de ella tendría que presentar un carácter esencialmente global, para lo cual, la coherencia cuántica es idónea. (2-106). Con las computadoras, al menos, continúa, se sabe que la acción es algorítmica, pero con los cerebros la situación es diferente: Existe en ellos un ‘ingrediente’ no algorítmico esencial en los procesos de pensamiento consciente. (1-477).

Penrose recurre a experiencias científicas en la investigación de las formas de acción del cerebro para respaldar su posición, particularmente aquellas que se refieren a experiencias de escisión cerebral. Por ejemplo, las desarrolladas por el neurocirujano canadiense-estadounidense Wilder Penfield, quien, en los años 40 y 50 del siglo anterior, detalló gran parte de las regiones sensorial y motora del cerebro humano, argumentando luego que la conciencia no está simplemente asociada a la actividad cerebral, sino más bien que cierta región del cerebro a la que denominó ‘tronco cerebral superior’, consistente principalmente en el tálamo y el cerebro medio, y específicamente la ‘formación reticular’, debería considerarse, en cierto sentido, como la ‘sede de la conciencia’, cuestión que otros neurofisiólogos han ya argumentado también.

Esta ‘formación reticular’ es la responsable del estado general de alerta del cerebro, y si se lesiona, el resultado es la inconsciencia. (1-452). Otros científicos especialistas en la neurología sostienen que la acción del hipocampo es la que más tiene que ver con el estado consciente. Penrose ha señalado que el hipocampo es crucial para el asentamiento de la memoria de largo plazo. Es un argumento muy sostenido que el asentamiento de recuerdos permanentes está asociado con la conciencia, y si esto es correcto, el hipocampo jugaría realmente un papel central en el fenómeno de la conciencia.

Sin embargo, hay también opiniones que van en el sentido de que es la propia corteza cerebral la responsable de la conciencia, y si el cerebro es el orgullo del hombre y sus actividades mentales más estrechamente asociadas con la inteligencia parecen ser llevadas a cabo por dicho cerebro, entonces es ciertamente allí donde reside el alma del hombre. Esta última posición pareciera ser aceptada por la IA fuerte, para quien los algoritmos complicados que ejecuta la corteza cerebral confirmarían a esta región como la más firme candidata a la capacidad de manifestar conciencia. (1-453).

Y respecto a la Teoría Cuántica, La evaluación que hace Penrose sobre sus dificultades tiene evidentes implicaciones filosóficas. Una de ellas se refiere a que la realidad, existiendo en un único y determinado estado, siempre aplica al observador y a sus instrumentos de observación, pero sólo se aplica a los objetos externos después de que han sido observados. El caso del "*gato de Schrodinger*", en el cual el gato está vivo y está muerto al mismo tiempo sólo hasta que la caja es abierta y el gato es observado, es un buen ejemplo utilizado por él. Recordemos que para Penrose, y esta es una tesis ya plenamente aceptada por la ciencia, la teoría cuántica es determinante y precisa, y su indeterminación aparece sólo cuando se realiza lo que se denomina 'hacer una medida', lo que implica amplificar el suceso desde el nivel cuántico hasta el nivel clásico. (2-18). La realidad, tal y como existe aparte de la observación, es la expresión determinística de la Ecuación de Schrodinger para la Función de Onda, que, por su lado, es también interpretada por Heisemberg y Bohr como la suma de todos los posibles estados del sistema físico.

V.3 Escisión del cerebro, individualidad humana y no verbalidad del pensamiento.

Experimentos científicos de escisión del cerebro en humanos muestran que en estos sujetos con cerebro escindido, ambos lados parecen comportarse como individuos prácticamente independientes, cada uno de los cuales puede comunicarse por separado con el experimentador; y, en un nivel más primitivo, con el hemisferio derecho que con el izquierdo, debido a la falta de capacidad verbal del derecho. Resulta tentador plantear la cuestión: ¿Tenemos dos individuos conscientes separados que habitan el mismo cuerpo? Científicamente se ha demostrado que ambos hemisferios son conscientes, y son conscientes por separado, ya que manifiestan gustos y deseos diferentes. Penrose opina que el lenguaje humano ordinario no es necesario para el pensamiento o para la conciencia, y acepta la hipótesis de que las dos mitades de un sujeto con cerebro escindido pueden ser conscientes independientemente. (1-454,455). ¿Cómo podría entonces sentirse un hombre después de haber sido dos personas separadas con distintos 'yo' en algún momento del pasado?, se pregunta Penrose. (1-456).

Esta cuestión de la verbalidad o no verbalidad del pensamiento viene siendo fuertemente debatida en los últimos años. Se ha sostenido denodadamente la tesis de que la verbalización es necesaria para el pensamiento, aunque dicha tesis viene siendo rebatida cada vez más por muchos científicos. El mismo Einstein opinaba, por ejemplo, que las palabras, o el lenguaje, ya sea escrito o hablado, no parecen desempeñar ningún papel en el mecanismo del pensamiento, sino ciertos signos o imágenes más o menos claras que pueden repetirse y combinarse voluntariamente. El genetista Francis Galton,

en similar forma, opina en el sentido que el lenguaje no corre a la par de sus ideas, y son más bien una resistencia para su clara expresión, teniendo que traducir las ideas en un lenguaje que no corre a la par con ellas. Hadamar dice que *"las palabras están totalmente ausentes de mi mente cuando realmente pienso, y todas desaparecen cuando comienzo a pensar"*. Schopenhauer escribe que *"las ideas mueren en el momento en que se encarnan en palabras"*. Ante todo ello, Penrose asume la posición de que no es que no se piense con palabras, sino que estas son casi inútiles para el pensamiento matemático, y que el pensamiento consciente puede ser de carácter no verbal, por lo que de igual manera el pensamiento puede tener también un pensamiento no algorítmico. (1-499, 500,501).

V.4 Manifestaciones pasivas y activas de la conciencia.

Aun y reconociendo lo difícil que resulta definir la conciencia, *"porque no sabemos qué es"*, Penrose, sin embargo, reconoce dos tipos de manifestaciones de la conciencia: Unas, que implican conocimiento, son las que llama *"manifestaciones pasivas"*, y entre ellas se encuentran el color, la armonía, y el uso de la memoria; otras, que implican conceptos, y que llama *"manifestaciones activas"*, entre las que ubica el libre albedrío y la realización de acciones voluntarias y conscientes. Pero Penrose piensa que hay algo esencial que implica la conciencia, y que es diferente de sus aspectos pasivo y activo, estando probablemente en algún lugar entre ambos; es la *"comprensión"*, o bien, la *"intuición directa"*, que, de igual manera, no intenta definir porque no la conoce, así como tampoco conoce o entiende términos como *"inteligencia"* y *"conocimiento"*, (2-83,84), términos estos últimos sobre los que el mismo Hawking ubica el problema y la discusión.

V.5 Un intento de resumen y conclusión.

Podemos tratar de resumir la posición de Penrose en cuanto al problema mente-cuerpo y el origen de la conciencia, frente a la propia posición de los sostenedores y defensores de la IA fuerte: Penrose no cree en la tesis de la IA fuerte de que la simple activación de un algoritmo produciría la conciencia. Sostiene que debe haber un ingrediente esencialmente no algorítmico en la actuación de una conciencia. (1-480,481). Aunque reconoce que es difícil ser muy preciso sobre las distinciones entre los tipos de actividad mental que parecen exigir la conciencia y las que no lo hacen, cree que, utilizando el tipo de terminología actual que distingue nuestra actividad mental consciente de la inconsciente, y esto es muy importante, *"la formación de nuevos juicios sugiere al menos una distinción no algorítmica: Conciencia necesaria: Sentido común, juicio de verdad, comprensión, valoración artística; conciencia no necesaria: automático, seguir reglas despreocupadamente, programado, algorítmico"*. Penrose sugiere que *"mientras que las actuaciones inconscientes*

del cerebro son las que proceden según procesos algorítmicos, la acción de la conciencia es muy diferente y actúa de una forma que no puede describirse mediante ningún algoritmo. La formación de juicios, que afirma que es la impronta de la conciencia, es ella misma', algo sobre lo que la IA fuerte no tendría ninguna idea de cómo programar en una computadora. (1-485,486).

No se trata, pues, para Penrose, de dar respuesta a las clásicas preguntas del problema mente-cuerpo que se vienen debatiendo desde el racionalismo: ¿Cómo es posible que una conciencia pueda surgir realmente a partir de un objeto material, un cerebro?, y ¿Cómo es posible que una conciencia, mediante la acción de la voluntad, influya realmente en el movimiento, (que aparentemente está determinado físicamente), de los objetos materiales?, aspectos pasivo y activo, respectivamente, del problema mente-cuerpo, que indudablemente merecen respuesta. Ahora hay un problema adicional de naturaleza más científica y a la vez más hondamente filosófica: ¿Qué ventaja selectiva confiere la conciencia a quienes realmente la poseen? (1-478).

Penrose plantea además, un problema adicional que implica consideraciones filosóficas, e incluso teológicas, de muy delicado nivel. Él cree en el poder de la evolución natural, a grado tal que considera las teorías de Wallace y Darwin unas de las que él califica como "*supremas*"; pero no ve cómo la selección natural por sí sola pueda hacer evolucionar algoritmos que pudieran tener el tipo de juicios conscientes sobre la validez de otros algoritmos que al parecer tenemos. "*Imaginemos*, dice, por ejemplo, un programa ordinario de computadora. *¿Cómo llegó a formarse? Es evidente que no (directamente) por selección natural. Algún programador humano de computadoras lo habrá concebido, verificando que realiza correctamente las acciones que supone deberá hacer*". (1-488). Este es, en el fondo, un ataque certero a los defensores de la IA, y en el fondo, también una fuerte crítica a la evolución del hombre, y con él, de su conciencia.

En *Lo grande, lo pequeño y la mente humana*, (2), hace Penrose una síntesis de los diversos puntos de vista acerca de la relación entre pensamiento consciente y computación, que estimo es muy útil para aclarar llanamente esta cuestión. (2-84,85). Él reconoce cuatro posiciones:

A) Todo pensamiento es computación; en particular, las sensaciones de conocimiento consciente son producidas por la mera ejecución de cálculos apropiados. Esta es la posición de la IA fuerte, o *funcionalismo computacional*, que afirma que todo pensamiento es simplemente la realización de ciertos cálculos, y, en consecuencia, si uno realiza los cálculos apropiados, el resultado será el conocimiento;

B) La conciencia es una característica de la acción física del cerebro; y mientras que cualquier acción física puede ser simulada computacionalmente, la simulación computacional no puede por sí misma producir conciencia. Es posible, aquí, simular la actividad de un cerebro cuando su propietario es conocedor de algo. Un cerebro compuesto de neuronas y similares podría ser conocedor, mientras que la simulación de la actividad de dicho cerebro no sería conocedora. Esta es, según Penrose, la posición de John Searle, ya señalada antes de alguna manera.

C) La acción física apropiada del cerebro produce conciencia, pero esta acción física no puede siquiera ser propiamente simulada computacionalmente. Esta es su propia posición. Si bien hay algo en la actividad física del cerebro que provoca conocimiento, -en otras palabras, hay algo en la física a lo que tenemos que dirigirnos-, esta actividad física es algo que ni siquiera puede ser simulado computacionalmente. No hay ninguna simulación que pudiera realizarse de dicha actividad. Esto requiere que exista algún factor en la actividad física del cerebro que esté más allá de la computación.

D) La conciencia no puede ser explicada en términos físicos, computacionales, o cualesquiera otros términos científicos.

E) Esta síntesis de posiciones es muy esclarecedora, y sitúa muy bien a Penrose en cuanto a otras posiciones ante este problema.

OBRAS CONSULTADAS.

Roger Penrose, *la mente nueva del Emperador. En torno a la cibernética, la mente y las leyes de la física.* Consejo Nacional de Ciencia y Tecnología. Fondo de Cultura Económica, México, 1996.

Roger Penrose, *Lo grande, lo pequeño y la mente humana.* Cambridge University Press, primera edición, Madrid, 1999.

Roger Penrose, *el camino a la realidad.* Debate, Random House Mondadori, S. A., Barcelona, primera edición, octubre de 2006.

Roger Penrose, *Las sombras de la mente.* Crítica, 1996.