

EL POST-HUMANISMO ROBÓTICO DE RAY KURZWEIL. UN ANÁLISIS CRÍTICO

RAY KURZWEIL'S ROBOTIC POST-HUMANISM A
CRITICAL ANALYSIS

Carlos Beorlegui
Universidad de Deusto

Resumen: *Los avances de la ingeniería informática le llevan al experto en IA R. Kurzweil a proponer un futuro no muy lejano en el que se podrán construir robots inteligentes que llegarán a estar dotados de conciencia, libertad y sentimientos. En su opinión, estas cualidades son las que llevarán a los humanos a tener que reconocer la condición de personas de tales robots, y a elaborar una ciberética que regule las relaciones de los humanos con los robots y de éstos entre sí. Este artículo presenta las tesis centrales de la propuesta de R. Kurzweil, sobre todo la concepción de la condición humana que está latente en ella, para evaluar críticamente lo que de interesante, pero a la vez problemático, hay en esta propuesta de "humanismo extendido" que nos propone el intelectual norteamericano.*

Palabras clave: *Inteligencia Artificial (IA), ciberética, cyborgs, humanismo extendido, filosofía de la mente, emergentismo.*

Abstract: *The advances in computer engineering have lead R. Kurzweil to propose a not too far future in which intelligent robots endowed with consciousness, freedom and feelings may be built. In his opinion, these qualities will force humans to accept the human nature of these robots and to develop cyberethics to govern human-robot and robot-robot interactions. The author of this paper presents R. Kurzweil's proposal, particularly the idea of the human latent in it, to critically assess the idea of "extended humanism" proposed by this North American scholar.*

Keywords: *Artificial Intelligence (AI), cyberethics, cyborgs, extended humanism, philosophy of mind, emergentism.*

1. EL RENACER DE LA INGENIERÍA COMPUTACIONAL

No cabe duda de que la tecnología informática nos está maravillando con sus increíbles avances, que se suceden a ritmos acelerados. De ahí que ya parece que ninguna meta le es imposible, y no nos atrevemos a decir que haya algo que no pueda conseguir. Uno de los exponentes actuales más significativos de este *optimismo tecnológico*, sobre todo en el campo de la Inteligencia Artificial (IA), es R. Kurzweil (Nueva York, 1948). Tras una larga y brillante carrera en varios campos de la ingeniería informática, es el actual jefe de ingenieros de Google, así como autor de varios libros de gran impacto, tanto en el terreno científico como filosófico.

Su influjo y prestigio no se reduce al ámbito norteamericano, sino que también se está dando a conocer en nuestro entorno, debido sobre todo a la traducción reciente de dos de sus libros más importantes: *La Singularidad está cerca. Cuando los humanos trascendamos la biología*¹, y *Cómo crear una mente. El secreto del pensamiento humano*². Esto está haciendo que las propuestas de R. Kurzweil (RK) estén siendo cada vez más estudiadas y analizadas, y causan admiración en muchos y críticas diversas en otros, pero no dejan indiferentes a ninguno de sus lectores.

Aunque las reflexiones de RK se dirigen a un amplio conjunto de temas relacionados entre sí, lo que me interesa en estas páginas es más bien analizar y evaluar críticamente sus ideas sobre la mente y el ser humano, para examinar con detenimiento su consistencia y las limitaciones que en mi opinión contienen. Por consiguiente, tras presentar en un primer apartado un breve panorama de sus aportaciones teóricas, me centraré después en analizar con más detenimiento su concepción del ser humano, en sus diferentes facetas, presentando al mismo tiempo las limitaciones e interrogantes que esta antropología contiene. Las reflexiones se terminarán con unas breves conclusiones críticas finales.

2. LA TRAYECTORIA INTELECTUAL Y LA OBRA ESCRITA DE R. KURZWEIL

El propio RK nos indica que lo que ha guiado siempre su vida es presentar ideas y perseguir retos que puedan aportar algo nuevo y valioso a los retos que se plantea la humanidad³. Esta filosofía le ha orientado al campo de la Inteligencia Artificial (IA), tratando de investigar la estructura del pensamiento humano, con objeto de llegar a construir máquinas cada vez más inteligentes, de tal modo que se pueda en un futuro no muy lejano fabricar robots

¹ Ray KURZWEIL, *La Singularidad está cerca. Cuando los humanos trascendamos la biología*, Berlín, Lola Books GbR, 2012.

² Ray KURZWEIL, *Cómo crear una mente. El secreto del pensamiento humano*, Berlín, Lola Books GbR, 2013.

³ Cfr. Ray KURZWEIL, *La singularidad está cerca*, pp. 1-2. En la p. 2, afirma: "Mi vida se ha guiado siempre por este imperativo. El poder de una idea: esto es en sí mismo una idea".

inteligentes que puedan pasar con éxito el test de Turing y ser considerados totalmente iguales a los humanos.

Para R. Kurzweil está claro que el ser humano es capaz, por su inteligencia, de superarse a sí mismo e ir más allá de su herencia genética. Y todo ello, debido a nuestro extraordinario cerebro y su capacidad de generar artefactos complejos, que, a su vez, de modo reflejo, enriquecen la mente humana. Como defensor de la IA fuerte, está convencido de que las máquinas del futuro podrán llegar a ser conscientes, por lo que no duda en que podrán ser consideradas y tratadas como *personas*. Se trata de un proceso que seguirá tres etapas: en la primera se irán implantando prótesis artificiales en el cerebro humano, que complementarán la capacidad intelectual de los humanos; a continuación, se dará una superación de la inteligencia humana, biológica, por la artificial; y en tercer lugar, se construirán robots tan potentes que podrán reproducirse y construir otros más capaces que ellos, de tal modo que se podrá iniciar la colonización del universo con robots inteligentes, conscientes y libres, de manera que la IA marcará el futuro del universo. Todo el universo está orientado hacia la inteligencia, que comienza por estar encarnada en la biología, para más adelante compartir su dominio y ser superada por la IA.

Kurzweil entiende que el universo conforma una totalidad unitaria, formada por diversos niveles de realidad (el físico, el químico, el biológico y el neurológico), coronados por el último y más complejo, el tecnológico, creador de la IA, de la que es un experto mundial de primera línea. Sus investigaciones actuales se orientan al reconocimiento del habla humana. En su primer libro sobre estos temas, *The Age of Intelligent Machines* (1989), profetizaba que “en la primera mitad del siglo XXI la inteligencia de los ordenadores sería indistinguible de la de sus progenitores humanos”⁴. Es la expresión del convencimiento que tiene RK de que el poder de las ideas para transformar el mundo se está acelerando. A esta tendencia la denomina *principio o ley de los rendimientos acelerados*, consistente en advertir que la tecnología y los procesos tecnológicos progresan de forma exponencial, no lineal. En su segundo libro, *The Age of Spiritual Machines* (1999), pretende mostrar “cómo sería la naturaleza de la vida humana una vez superado el momento en que la máquina y cognición humana se conviertan en una misma cosa”⁵. Advierte RK que progresivamente se va dando una colaboración cada vez más estrecha entre nuestra herencia biológica y la tecnología informática, y está cerca el momento en el que se trascenderá lo biológico, sin dejarlo obsoleto, pero sí superado.

En el libro que concluye su trilogía, *The Singularity is near* (2005), continúa reflexionando sobre las consecuencias de la inminente tendencia a la fusión entre nuestro pensamiento biológico y la inteligencia no biológica que esta-

⁴ *Ibíd.*, p. 3.

⁵ *Ibíd.*, p. 4.

mos creando. Para RK, “la historia se fundamenta en la idea de que tenemos la capacidad de comprender nuestra propia inteligencia (penetrar nuestro propio código fuente, si se prefiere), modificarla y ampliarla”⁶. Frente a quienes dudan de que el cerebro humano sea suficientemente potente como para comprenderse a sí mismo, RK considera que los hechos demuestran que sí tenemos capacidad para comprendernos, crear y ampliar modelos de nuestra propia inteligencia. De hecho, la tesis central de este libro consiste en mostrar que nos estamos acercando a ese momento clave (*la singularidad está cerca*), a esa *singularidad* en la que se dará un salto en el que la inteligencia dejará de estar encarnada en lo biológico para hacerlo en una base tecnológica. Este salto le parece a RK similar a la magia, que saca de la chistera realidades que nadie esperaba ni sabe explicar cómo han surgido. Con las capacidades de esta magia tecnológica moderna “podemos hacer que un ordenador lea un libro en alto, entienda el habla humana, anticipe (y prevenga) un ataque o prediga el movimiento de una acción en bolsa”⁷. RK define este libro como “la historia del destino de la civilización hombre-máquina, un destino al que llamamos Singularidad”⁸.

La principal dificultad que advierte RK para el avance de la IA es el conocimiento de la complejidad de nuestro cerebro. Es el impedimento mayor para hacer una *ingeniería inversa* del cerebro que nos lleve a construir máquinas tan inteligentes o más que nuestros cerebros. Para ello, es fundamental que conozcamos la historia y las estructuras de nuestros cerebros biológicos. La cualidad más destacada del cerebro humano es su *plasticidad*⁹. Tanto en la dimensión filogenética como ontogenética, el cerebro humano sigue cambiando y construyéndose a sí mismo debido a las interacciones permanentes que mantiene con el entorno, y sobre todo con los demás individuos humanos. Tal plasticidad le permite incluso saber cada vez más acerca de sí mismo y puede con ello transformar sus propias potencialidades, todo en una ascendente espiral que no tiene término. Esta progresiva e interminable dinámica en espiral es la que nos está acercando, en opinión de RK, a ese momento de *Singularidad*, en el que las máquinas pensantes creadas por los humanos nos superarán en capacidad intelectual, y será el momento en el que se podrá iniciar la colonización del universo, e incluso la construcción de otros nuevos universos¹⁰. En ese momento, el proceso evolutivo del universo, que ha llegado a ser inteligente y autoconsciente con la aparición de la especie humana, dará el salto de una inteligencia apoyada en bases biológicas y cerebrales a otra apoyada en bases materiales nuevas, de silicio o de otro material. La

⁶ *Ibíd.*, p. 4.

⁷ *Ibíd.*, p. 5.

⁸ *Ibíd.*, p. 6.

⁹ Cfr. Ray KURZWEIL, *Cómo crear una mente*. caps. 4 y 5.

¹⁰ Cfr. Ray KURZWEIL, “Prólogo” a James GARDNER, *El universo inteligente*, Barcelona, Robin Book, 2008, pp. 15-23.

inteligencia humana se habrá librado de sus “cadenas biológicas”, en frase de RK, para apoyarse en nuevos soportes técnicos y artificiales.

El cerebro humano es el resultado de una evolución biológica, que consiste en ir aumentando su complejidad, para después organizarla de modo ordenado, pasando de un tipo de orden a otro más complejo. Por tanto, la evolución, al mismo tiempo que aumenta la complejidad, aumenta también el orden. Así, se va creciendo en complejidad ordenada, atravesando diversos paradigmas, cada uno de los cuales comienza de modo lento, para imponerse en un segundo momento de forma rápida, y asentarse después de forma estable. Está claro para RK que la inteligencia es el instrumento que la selección natural ha ido dotando a los seres vivos en el empeño por sobrevivir, tarea que realiza nuestra mente a base de poner en práctica la técnica de reconocimiento de patrones¹¹. De ahí que discrepe de la tesis de S. Pinker, para quien la inteligencia no sería el objetivo de la selección, mientras que RK defiende que es precisamente la inteligencia la que permite la supervivencia de los individuos y las especies vivas, con lo que queda claro que constituye la tendencia central que ha servido para preservar la selección natural. Pero RK no considera que la evolución haya seguido un ritmo gradualista, aunque tampoco rupturista, sino que más bien se adhiere a la tesis de S. J. Gould y N. Eldredge, denominada de los *equilibrios puntuados*¹². En definitiva, cuando un paradigma se agota, otro más potente y complejo le sustituye.

En el conjunto de la historia del universo RK contempla seis grandes etapas, descritas en *La singularidad está cerca*¹³. En la primera, la *física y química*, etapa prebiótica, el universo se hallaba configurado únicamente por las leyes de la física y de la química; la segunda, la *biológica*, comienza cuando aparecen los seres vivos, basados en compuestos de carbono, estructurados por el ADN como mecanismo para almacenar información; en la tercera etapa, la *cerebral*, aparecen los cerebros humanos, capaces de orientar la evolución en una dirección consciente y predeterminada. A la etapa cerebral le sigue la *tecnológica*, momento en el que la inteligencia humana adquiere la capacidad de construir tecnologías cada vez más inteligentes, complementarias de la inteligencia humana biológica, al mismo tiempo potenciadoras de la misma, como consecuencia de la interacción entre ambas. Es en esta etapa cuando se produce, como vimos, lo que RK denomina el momento de *Singularidad*¹⁴, fase de la historia en la que “el ritmo del cambio tecnológico será tan rápido y su repercusión tan profunda que la vida humana se verá transformada de

¹¹ Cfr. Ray KURZWEIL, *La singularidad está cerca*, cap. 3.

¹² Cfr. *Ibid.*, pp. 48-50.

¹³ Cfr. *Ibid.*, cap. 1.

¹⁴ Cfr. *Ibid.*, pp. 25 y 567.

forma irreversible"¹⁵. RK se arriesga a proponer como fecha del logro de tal *Singularidad* el año 2045¹⁶.

En virtud de lo que denomina *ley de rendimientos acelerados*¹⁷, se producirá en ese momento una aceleración tal del progreso tecnológico que avanzará de forma no lineal, sino exponencial, y la humanidad y el universo entrarán en la quinta etapa, caracterizada por la fusión entre la inteligencia cerebral humana y la tecnología computacional. De este modo, la nueva tecnología nos capacitará a los humanos para trascender nuestras limitadas capacidades intelectuales cerebrales, así como para superar el envejecimiento y ampliar la creatividad humana¹⁸. Claro que este aumento de nuestras capacidades también nos dotará de mayores posibilidades de maldad y destrucción, con lo que "la historia en su conjunto todavía está por escribirse"¹⁹.

A la sexta y última etapa la denomina: *el universo despierta*²⁰, momento en el que la inteligencia, emergida primero en los cerebros biológicos humanos, se irá extendiendo e implantando en artefactos artificiales, en *cyborgs* y robots cada vez más potentes y dotados, que podrán llegar incluso a ser conscientes y libres, así como capaces de ir creando otros robots autoconscientes y libres, de tal modo que serán colaboradores de los humanos, para después formar una cultura robótica autónoma. Además, la existencia de esta IA llegará a tener la capacidad de transformar la materia pasiva e inconsciente del universo, es decir, la materia *tonta*, en materia inteligente, impregnada de la creatividad e inteligencia que podrá tomar prestada de la inteligencia humana. Es el momento en el que el *universo despertará* y pasará de estar formado de materia dormida a estarlo de materia despierta, consciente e inteligente, "en formas de inteligencia exquisitamente sublimes"²¹. En definitiva, el proceso evolutivo de la historia del universo nos hará más listos, tanto a los humanos como al universo entero.

Estas afirmaciones, aunque basadas en los avances concretos de las ciencias y tecnologías de la ingeniería informática y de las ciencias de la información, tienen también mucho de propuestas fantasiosas, de ciencia ficción, que nos exigen grandes dosis de fe y confianza en las inmensas posibilidades de la tecnología. De todas formas, RK no se deja llevar del todo, en su optimismo tecnológico, por una imaginación desbordada y sin soportes concretos, sino que aporta continuamente datos específicos de los avances científicos, así como programas de investigación sobre la estructura y funcionamiento del cerebro, y muestras de los muchos esfuerzos y avances de la denominada

¹⁵ *Ibíd.*, p. 7.

¹⁶ *Cfr. Ibíd.*, p. 148.

¹⁷ *Cfr. Ibíd.*, cap. 2.

¹⁸ *Cfr. Ibíd.*, pp. 22-23.

¹⁹ *Ibíd.*, p. 23.

²⁰ *Cfr. Ibíd.*, p. 23 y cap. 6.

²¹ *Ibíd.*, p. 23.

ingeniería inversa, que persigue construir una IA cada vez más potente, en su empeño por aprender del cerebro humano. Todo esto queda mostrado en su libro *Cómo crear una mente*. Continuando los planteamientos de sus libros anteriores, en éste se detiene RK en investigar los procedimientos que utiliza la mente humana, como es el *sistema de reconocimiento de patrones* (PRTM), procedimiento que, según él, conforma uno de los algoritmos básicos del neocórtex del cerebro humano. Ya hemos visto que llegar a saber cómo funciona el cerebro humano es una estrategia básica para la IA, en la medida en que si ésta no puede avanzar de forma más rápida, no se debe tanto a las deficiencias de la ingeniería informática, sino al enorme desconocimiento que todavía tenemos de los secretos de esa máquina tan compleja que es nuestro cerebro.

3. LAS TESIS DE RK SOBRE LA IA FUERTE

3.1. *Hacia la ingeniería inversa del cerebro*

Las estrategias de la IA tienen que orientarse tanto hacia el avance del *software* como del *hardware*. RK dedica el capítulo tercero de *La singularidad está cerca* a presentar las diversas estrategias que está siguiendo la tecnología informática actual para construir máquinas cada vez más inteligentes, que sepan simular cada vez mejor al cerebro humano. Hasta no hace mucho, la IA seguía estrategias de mejora centradas en las programaciones, en el *software*, buscando rutinas o comportamientos que simulasen la mente humana, menospreciando su base material, el funcionamiento físico del cerebro. En cambio, en la actualidad los esfuerzos van sobre todo en la línea de aprender del cerebro, para llegar a construir lo que RK denomina la *ingeniería inversa* de nuestro cerebro. Se necesita para ello llegar a ver dentro de él, crear modelos del mismo, y tratar de simular el funcionamiento de sus regiones y del propio cerebro en su totalidad. En la medida en que se está progresando en la tecnología de los escáneres cerebrales, y en el futuro se podrán incluso introducir dentro de él *nanorobots*, se podrá después avanzar de forma acelerada en la comprensión del cerebro y en la construcción de robots inteligentes que lo imiten. Ya sabemos que, de momento, los ordenadores son muy diferentes e inferiores al cerebro humano²², pero esta diferencia durará muy poco tiempo, piensa RK, hasta que se produzca la ya nombrada *Singularidad*, hacia el año 2045.

La extraordinaria potencialidad de los diversos escáneres cerebrales permitirá, por tanto, a los humanos entender las diversas regiones del cerebro, para descubrir cómo funcionan y desarrollan las diversas potencialidades de la mente humana. Porque no se trata de imitar sólo las capacidades sensoriales (vista, oído, etc.) y la inteligencia mecánica, sino también las funciones de

²² Cfr. Ray KURZWEIL, *La singularidad está cerca*, pp. 164-170.

más alto nivel, como son la imitación, predicción y emoción²³. Y si conseguimos escanear el córtex cerebral, también se habrá logrado descubrir el funcionamiento del pensamiento consciente, el lenguaje y la capacidad de tomar decisiones libres²⁴. La meta, pues, de la IA es tanto la inteligencia racional como la emocional, que ha constituido hasta ahora la parte más olvidada y menos valorada, y, por eso, menos estudiada, y sin duda la más compleja²⁵.

3.2. *La IA dentro de la filosofía de la mente*

RK es defensor convencido de la IA en su sentido fuerte, pero su postura es suficientemente matizada como para que nos tengamos que detener un momento en analizarla. Cuando a mitad del siglo XX, las teorías cognitivistas y los avances en las ciencias del cerebro arrumbaron el conductismo como teoría dominante²⁶, dentro del cognitivismo se impuso el paradigma filosófico del funcionalismo, partidario de entender las relaciones mente-cerebro como la existente entre programa (*software*) y base material (*hardware*) en un ordenador. Esta imagen tan sugerente y fructífera representó el inicio de una explosión investigadora de la denominada IA. Pero mientras que la versión *débil* se limitaba a defender que muchos estados mentales y comportamientos humanos se asemejan y pueden ser imitados, y superados incluso, por programas de ordenador, la versión *fuerte* de la IA se apoya en el convencimiento de que la mente humana puede ser copiada e imitada, en su totalidad, por programas de ordenadores. Y aunque, es cierto, que reconocen que los ordenadores actuales todavía no llegan a la complejidad de la mente humana, nos aseguran que en el futuro lo conseguirán.

En ese momento clave habrán superado el denominado *test de Turing*²⁷, por lo que ya no será posible distinguir entre la inteligencia de un ser humano y la de un robot²⁸. RK es defensor, como hemos señalado, de la IA en sentido fuerte, y dentro de las tres más grandes revoluciones tecnológicas que se están dando en nuestra época (en el campo de la genética, de la nanotecnología y de la robótica), considera que la revolución en el ámbito de la

²³ *Ibid.*, p. 211.

²⁴ *Cfr. Ibid.*, pp. 211-214.

²⁵ *Cfr. Ibid.*, pp. 214-217.

²⁶ *Cfr.* Howard GARDNER, *La nueva ciencia de la mente. Historia de la revolución cognitiva*, Barcelona, Paidós, 1987 (2ª ed.: 2000); P. F. MARTÍNEZ-FREIRE, *La nueva filosofía de la mente*, Barcelona, Gedisa, 1995); P. F. MARTÍNEZ-FREIRE (ed.), *Filosofía actual de la mente*, Valencia Contrastes, Suplemento 6, , 2001; Manuel LIZ, *Perspectivas actuales en filosofía de la mente*, Tenerife, Gobierno de Canarias, 2001; Fernando BRONCANO (ed.), *La mente humana*, Madrid, Trotta, 1995; Carlos MOYA, *Filosofía de la mente*, Valencia, Universitat de València, 2004; Carlos BEORLEGUI, "Filosofía de la mente. Visión panorámica y situación actual", en *Realidad* (UCA, San Salvador)111 (2007) pp. 121-160.

²⁷ *Cfr.* Alan ROSS ANDERSON (ed.), *Controversia sobre mentes y máquinas*, Barcelona, Tusquets, 1984; John SEARLE, *Mentes, cerebros y ciencia*, Madrid, Cátedra, 1984; B. Jack COPELAND, *Alan Turing. El pionero de la era de la información*, Madrid, Turner, 2013.

²⁸ *Cfr.* R. KURZWEIL, *La singularidad está cerca*, p. 224.

robótica es posiblemente la más significativa²⁹. Desde que se fue imponiendo la teoría de la evolución y se produjeron los avances de la genética durante el siglo XX, el ser humano ha conseguido curar enfermedades, prolongar su vida, soñar con la inmortalidad y conseguir bebés de diseño³⁰. No menos importantes son los avances en la nanotecnología, aunque estemos todavía en los comienzos de este extraordinario campo de investigación. Los límites de la biología serán superados por la nanotecnología, puesto que se podrán introducir microrobots dentro del cuerpo humano, para conocerlo mejor, repararlo y alargar su vida. Es verdad que se trata de una tecnología que tiene grandes y múltiples riesgos y peligros, reconoce RK, pero son mayores aún las ventajas en todos los terrenos, incluso para hacer un mundo menos desigual y más llevadero para los humanos³¹.

Pero donde nos encontramos con la revolución más potente y profunda es en el campo de la robótica, porque supondrá extender todas las potencialidades de los humanos. Los avances más significativos no se darán en el cuerpo físico del robot, sino en el terreno de la inteligencia³². Con estos avances, la IA superará de forma incontestable a la inteligencia humana, porque ésta tiene sus límites biológicos que no puede superar, mientras que la IA no tendrá prácticamente límites. Además, las máquinas inteligentes pueden unir y fusionar sus recursos intelectuales, cosa que no se puede dar entre los humanos, excepto en muy poca medida. Y un capítulo fundamental del avance de la robótica se producirá cuando se unan ésta y la nanotecnología, y emerja la capacidad de construir nanorobots que puedan ser introducidos en el interior del ser humano, sobre todo en su cerebro. Estamos, pues, en un momento en que los avances de la IA van a ser –lo están siendo ya– *galopantes*³³.

Es cierto, reconoce RK, que la IA ha pasado una época de descrédito y desánimo (un *invierno*)³⁴, entre los investigadores, porque se comenzó con un *optimismo prematuro*, pero ese momento se ha superado, en la medida en que después la IA ha progresado ya muchísimo, como es de todos sabido. Y pronto se conseguirá alcanzar las tesis de la IA fuerte, sin que se den ya diferencias esenciales entre la inteligencia humana y la artificial. Muestra RK una larga lista de habilidades de la inteligencia humana que todavía no pueden realizar las máquinas; pero ve que esa lista de diferencias va disminuyendo poco a poco. La dificultad para RK, como ya lo hemos dicho, se halla no tanto en las limitaciones de la ingeniería informática, sino en el gran desconocimiento que todavía tenemos sobre nuestro cerebro. La previsión de RK es que, con las nuevas tecnologías (escáneres cerebrales y nanorobots), se

²⁹ Cfr. *Ibíd.*, cap. 5.

³⁰ Cfr. *Ibíd.*, pp. 230-255.

³¹ Cfr. *Ibíd.*, pp. 225-393.

³² Cfr. *Ibíd.*, pp. 294-341.

³³ Cfr. *Ibíd.*, pp. 297.

³⁴ Cfr. *Ibíd.*, p. 298.

podrán construir máquinas dotadas de sistemas auto-organizativos, capaces de aprender y de reproducirse, dotados de lenguaje y de conocimiento, como lo están los humanos. Y, por supuesto, este tipo de máquinas serán capaces de superar sin problemas el *test de Turing*³⁵. La fecha que RK para este logro es la de 2029³⁶, y constituirá el acontecimiento más importante del siglo, en opinión de RK³⁷.

Como puede comprenderse, el impacto que todas estas tecnologías tendrán en el propio ser humano y en su entorno cultural será tremendo, como se extiende en describirlo RK a lo largo del capítulo 6 de esta obra. Impactos en el cuerpo, en el cerebro, en la longevidad, en el aprendizaje, en el trabajo, en el juego, en la tecnología militar, y, por último, en el conjunto del universo, en la medida en que los humanos, por medio de los robots inteligentes, nos haremos cargo de la historia del universo, colonizándolo en toda su extensión, e incluso construyendo otros universos similares al nuestro. Pero sobre esto volveremos más adelante.

Ahora bien, su postura difiere un tanto de la mayoría de los que se sitúan en esa posición. La mayoría de los defensores de la IA fuerte defienden tesis reduccionistas, considerando la mente humana como un mero programa de ordenador, todo lo potente que se quiera. De ahí que no vean problema en llegar a conseguir sobrepasar el *test de Turing*, porque para ellos es sólo cuestión de tiempo. En cambio, la postura de RK, como indica J. Monserrat³⁸, más que identificarse con el reduccionismo común, se puede considerar como un *humanismo extendido*, en la medida en que, como vamos a ver, no reduce la mente humana a una simple mente computacional, sino que más bien reconoce la especificidad de la mente humana, para extender todas sus características a los robots del futuro.

Además, frente a muchos científicos computacionales de tendencia reduccionista, que consideran que estas cuestiones son exclusivamente competencia de las ciencias (sean las neurociencias o la psicología cognitivista), RK defiende repetidas veces que hay que saber distinguir en este punto lo que le corresponde a las ciencias y lo que es competencia de la filosofía, sea cual sea la postura filosófica que uno quiera defender³⁹. Así, cuando se refiere a la dificultad de captar la vertiente subjetiva de la conciencia, reconoce que muchos científicos y filósofos optan por negar esa dimensión subjetiva, reduciendo la mente a su soporte o expresión objetiva, explicable en toda su extensión desde parámetros científicos. Pero RK entiende correctamente que

³⁵ Cfr. *Ibíd.*, pp. 331-339.

³⁶ Cfr. *Ibíd.*, p. 337.

³⁷ Cfr. *Ibíd.*, p. 339.

³⁸ Cfr. Javier MONSERRAT, "El problema del soporte físico de la sensibilidad-conciencia", en Fco. RODRÍGUEZ VALLS (ed.), *La inteligencia de la naturaleza. Del relojero ciego al ajuste fino del universo*, Madrid, Biblioteca Nueva, 2102, pp. 101-118.

³⁹ Cfr. R. KURZWEIL, *La singularidad está cerca*, pp. 433 y ss.

el que no podamos penetrar en las experiencias subjetivas de los otros, y por tanto no poder objetivarlas, no nos permite concluir que no existen, o que no son verdaderas. No nos queda ante esto más que la vertiente de la primera persona, y, en relación a la mente de los demás, aceptar el testimonio del otro, o intuir empáticamente sus estados subjetivos, sus *qualia*. RK es consciente de que estas reflexiones se mueven todavía en el ámbito de lo político o psicológico, y se necesita profundizar más echando mano de la dimensión filosófica, nivel donde tiene que dirimirse este problema tan nuclear⁴⁰. Por eso dice: "Yo discrepo de aquellos que mantienen que la experiencia subjetiva, o bien no existe, o es superflua y por lo tanto puede ser ignorada con toda tranquilidad. La cuestión sobre quién o qué es consciente o la naturaleza de las experiencias subjetivas con los otros son temas fundamentales para nuestros conceptos de ética, moralidad y ley"⁴¹. La negación o minusvaloración de lo subjetivo de la mente es lo que Jaron Lanier denomina *totalismo cibernético*, postura con la que RK simpatiza y coincide (aunque discrepa en otros aspectos sobre su idea de IA)⁴².

En definitiva, la postura de RK respecto al acercamiento del estudio de la mente es mucho más matizada de lo que suele acostumbrarse entre los defensores de la IA fuerte. Frente a los reduccionistas, que parten de un monismo metodológico, que absolutiza el saber científico y no acepta más demostraciones que las empíricas y objetivas, RK considera que no podemos menospreciar el ámbito del mundo de la vida y de las descripciones fenomenológicas de nuestras experiencias subjetivas, que coinciden con las de los demás, aunque no las podamos someter a una demostración objetiva. Ante esta dificultad, en vez de absolutizar el enfoque metodológico y concluir que la conciencia y las experiencias subjetivas no existen, o no tienen importancia, RK considera que quizás sea más realista y respetuoso con la realidad aceptar el *hecho* de la conciencia subjetiva, reconocer la dificultad de captarla y definirla, y advertir que quizás haya que estudiarla desde otros enfoques diferentes a los científicos. No nos queda, por tanto, más solución que acudir al ámbito de la reflexión filosófica, como lo reconoce repetidas veces RK. "Precisamente porque no podemos resolver las cuestiones sobre la conciencia sólo mediante mediciones y análisis objetivos (es decir, mediante la ciencia), la filosofía juega un papel fundamental. Así, la conciencia constituye la cuestión más importante de la ontología. Después de todo, si de verdad imaginamos un mundo en el que no haya experiencias subjetivas (un mundo en el que haya cosas dando vueltas, pero sin ninguna entidad consciente que tenga experiencia de ellas), dicho mundo puede que tampoco exista"⁴³.

⁴⁰ Cfr. *Ibíd.*, p. 434.

⁴¹ *Ibíd.*

⁴² Cfr. *Ibíd.*, pp. 435-436.

⁴³ *Ibíd.*, p. 436; R. KURZWEIL, *Cómo construir una mente*, p. 192.

Esta cita es diáfana y fundamental, porque hace referencia a dos aspectos centrales del problema que nos ocupa: el tema del tipo de ontología de la conciencia, y la posibilidad y el sentido de universos sin conciencia, frente a su propuesta, que veremos más adelante, de convertir en meta del universo investirlo progresivamente de conciencia e inteligencia. Ahora bien, a pesar del desmarque claro de los reduccionistas, y de su rotunda defensa de la filosofía, su idea de inteligencia computacional es bastante ambigua e insuficiente, razón que le lleva a enfrentarse con varios filósofos de la mente de primer orden, que se oponen a las tesis del funcionalismo computacional y de la IA en sentido fuerte.

3.3. *Entre el funcionalismo y el patronismo (patternism) emergentista*

Los defensores de la IA fuerte se suelen situar en el entorno del funcionalismo computacional, dentro de las diferentes posturas de la nueva filosofía de la mente. Pero RK desborda, en cierta medida esta postura filosófica, situándose también cercano al emergentismo⁴⁴. Cuando analiza su propia postura, se desmarca tanto del dualismo como del *panprotopsiquismo* (teoría que defiende que “todos los sistemas físicos son conscientes, aunque un humano es más consciente que, digamos, el interruptor de la luz”)⁴⁵. Para RK, la conciencia es una propiedad *emergente* de un sistema físico complejo, siendo esta postura más propia del emergentismo que del funcionalismo computacional. Aunque, por otro lado, la metáfora del ordenador que utiliza el funcionalismo, y sigue RK, le hace proclive a caer en un dualismo larvado, en la medida en que separa en exceso el cerebro y la mente, y se centra de modo excesivo, concediéndole la importancia fundamental, a la programación (*software*) en detrimento de la base cerebral de la mente (*hardware*). A diferencia de esta postura, que menosprecia la base material cerebral para centrarse en la mente, como si fueran dos realidades independientes y el cerebro no tuviera mucha importancia, el emergentismo considera que sí importa el cerebro. Es decir, la mente humana tiene las características que posee, precisamente por estar causadas o apoyadas en este tipo de cerebro que poseemos los humanos, resultado de un proceso evolutivo que ha ido madurando desde la sensibilidad animal hasta, tras diversas y sucesivas etapas, posibilitar la emergencia de la mente consciente humana. De ahí que para el emergentismo, en principio sólo este tipo de cerebro puede ser el soporte de una mente como la humana⁴⁶. Las diversas máquinas artificiales, por muy programadas que estén para realizar actividades consideradas *inteligentes*,

⁴⁴ C. BEORLEGUI, “Los emergentismos sistémicos: un modelo fructífero para el problema mente-cuerpo”, *Pensamiento* 62, n. 234 (2006) 391-439; del mismo, “Filosofía de la mente. Visión panorámica y situación actual”; “Emergentism”, *Pensamiento* 65, n. 246 (2009) 881-914.

⁴⁵ Cfr. R. KURZWEIL, *La Singularidad está cerca*, p. 194.

⁴⁶ J. P. NÚÑEZ PARTIDO, *La mente: La última frontera*, Madrid, UPCO, 2012; del mismo, “Más que monos y máquinas: la mente humana se resiste a la reducción”, <<http://www.tendencias21.net>>. Consultado el 10 de noviembre de 2014.

no pasarán de hacer de modo inconsciente las actividades para las que los humanos las han fabricado, pero sus presuntas cualidades mentales (inteligencia, conciencia, emociones, libertad, etc.) no pasarán de ser actividades o conductas que se limitan a simular la conducta humana, pero no la igualan ni la igualarán nunca. Es decir, actúan *como si* fueran inteligentes, conscientes, libres, etc., pero no lo son en realidad. Les separa la ontología de su propia realidad: los humanos son seres biológicos, y los robots, máquinas de una materia diferente.

RK, en su empeño por construir robots semejantes a los humanos, entiende que lo específico de los mismos tiene que ser no tanto las diversas piezas con las que se construyan, sino que lo central para mantener su identidad es el *patrón*, la estructura, con la que están hechos. Ya sabemos que, en nuestro cuerpo, los elementos biológicos que componen nuestras células y órganos cambian continuamente, incluso también se da un cierto recambio dentro de las neuronas del sistema nervioso central. A pesar de ello, seguimos siendo *el mismo yo*, aunque no *lo mismo*. Basta que se mantenga el mismo patrón, la misma estructura de nuestro cuerpo. Incluso podemos cambiar el patrón, porque el yo humano es plástico, creativo, y con sus acciones libres se va haciendo a sí mismo, sin dejar de ser el mismo yo⁴⁷. En definitiva, nos dice RK, “yo soy un patrón de materia y energía que perdura en el tiempo”⁴⁸. Por eso, denomina a su postura como *patronismo* (*patternism*), porque “yo soy principalmente un patrón que perdura en el tiempo. Soy un patrón que evoluciona y puedo influenciar el curso de la evolución de mi propio patrón”⁴⁹.

En *La singularidad está cerca*⁵⁰, define su postura frente a los ataques de Demsky, de *materialista*, pero un materialismo *abierto*, el *materialismo de la capacidad*, o *patronismo de la capacidad*. Lo de *patronismo* ya lo hemos indicado, y lo de *capacidad* se refiere a su idea de que el cerebro no está determinado y cerrado por un programa que no puede ser superado, sino que está abierto a las múltiples interacciones con el ambiente y los demás humanos, que le permiten y le obligan a progresar en sus capacidades. En esas interacciones es donde se desarrollan sus *capacidades*, no sólo intelectuales, sino también emotivas y morales⁵¹.

Como puede verse, su concepción de la mente humana y su relación con el cerebro y el cuerpo entero se halla casi a caballo entre el funcionalismo computacional y un cierto *emergentismo estructurista*, en la medida en que entiende que tales patrones con los que están estructurados los seres vivos, son consecuencia de un proceso evolutivo *emergente*, siendo cada especie el resultado de un nuevo orden biológico, resultado de un nuevo enriqueci-

⁴⁷ Cfr. R. KURZWEIL, *La singularidad está cerca*, pp. 438-443.

⁴⁸ *Ibid.*, p. 439.

⁴⁹ *Ibid.*, p. 442.

⁵⁰ Cfr. *Ibid.*, 544.

⁵¹ Cfr. *Ibid.*, p. 545.

miento y complejidad de la vida. Pero vemos también que, en su visión de los robots inteligentes y conscientes del futuro, está claro para RK que se podrá copiar y traspasar toda la mente de un ser humano desde su cerebro a otra realidad material, pudiéndose apoyar los contenidos mentales del patrón singular de un humano en la estructura material de un robot. Lo importante para RK es que se copie y traspase el programa, el *patrón*. Porque se trata de dos realidades, el *software* y el *hardware*, que aunque están unidas y parecen inseparables, en el caso de los humanos son una unión que puede separarse para conformar otra unión, en este caso, entre mente y silicio, o entre mente y cualquier otro material que utilice la robótica.

Pero está claro que el problema principal en todo esto, y RK es muy consciente de ello, es el problema de la identidad; es decir, cómo demostrar que el robot artificial, que parece consciente (damos por hecho que supera ya el test de Turing), lo es realmente. RK se plantea entonces el problema de la identidad, el yo, y todos los demás elementos que consideramos propios de una persona humana.

3.4. *El problema de la conciencia*

La cuestión, por tanto, que hay que aclarar –lo sabe bien RK– es si cuando se pueda traspasar toda nuestra mente a un soporte artificial nuevo, no biológico, seguiré siendo yo, la misma persona. Además, RK ve que, con la construcción de robots super-inteligentes, en un futuro no muy lejano, este problema ya no será sólo un tema de la filosofía, muy disputado a lo largo de su historia, sino también un problema práctico y concreto, porque estaremos viviendo rodeados de robots y *cyborgs* sobre los que nos cuestionaremos acerca de su conciencia, personalidad y valor ético.

El problema tiene, pues, dos vertientes, pero representan en el fondo la misma dificultad. Por un lado está la posibilidad de que, al igual que puedo copiar un documento en bases materiales informáticas diferentes, puedo traspasar mi mente a un robot. En ese caso, nos tenemos que preguntar si sigo siendo yo; esto es, si soy el original, o la copia; y si me copian y desaparece el original, si sigo existiendo yo o una mera copia mía, que ya no será el yo original⁵². RK quiere ser prudente en esto, aunque se inclina a responder de forma afirmativa. El problema está en que estos soportes digitales de mi nueva personalidad tienen que ser vueltos a copiar de forma recurrente, pues los soportes de información serán siempre provisionales. Y por más que se tenga protección con estas copias, RK tiene que reconocer que la información sólo perdura mientras haya alguien que se encargue de cuidarla⁵³. Así que, aunque parece que seguimos siendo el mismo, y hemos conseguido la inmortalidad, estaremos pendientes de un hilo, de que alguien cuide bien la

⁵² Cfr. *Ibíd.*, p. 439.

⁵³ Cfr. *Ibíd.*, p. 377.

copia y la traspase a tiempo a otro soporte material, cuando el anterior esté deteriorándose.

En definitiva, la cuestión es: ¿quién soy yo? “Esta es la cuestión ontológica definitiva y normalmente nos referimos a ella como la cuestión de la conciencia”⁵⁴. Como hemos indicado más arriba, RK no quiere escapar de las cuestiones difíciles, sino que trata de encararlas con sinceridad y profundidad. Y no quiere darle la espalda, sobre todo, a la cuestión más difícil del problema humano, porque es consciente de que mucha gente, y también científicos y filósofos, al plantearse el problema de la conciencia, se desvían hacia considerarla como meras descripciones comportamentales (conductismo) o sólo funcionamientos cerebrales (teoría de la identidad, reduccionistas cerebrales). Pero este modo de enfocar el problema no pasa, como advierte RK, del enfoque de tercera persona, y no acierta a resolver lo que denomina Chalmers “la difícil cuestión”⁵⁵. Y en esta cuestión, una cosa es preguntarse cómo puede el cerebro producir una mente, y otra, relacionada con la anterior, pero diferente, cómo entender la autoconciencia. Sobre la primera cuestión podemos hablar en tercera persona, pero no sobre la segunda.

Está claro, por tanto, que, como ya vimos antes, a la hora de hablar y demostrar la existencia de la conciencia, no tenemos más evidencias que nuestra propia experiencia interior. Aunque eso no tiene que llevarnos al reduccionismo, o la negación de la conciencia, al no poder demostrarse objetivamente. RK entiende que la dificultad que tendremos de aceptar la conciencia en los robots futuros es similar a la que tenemos a la hora de defender la existencia de la conciencia de los otros humanos. Así como damos por hecho la conciencia de los otros, simplemente porque proyectamos sobre ellos nuestras experiencias interiores, lo mismo debiéramos hacer –opina RK– respecto a los robots. Al observar la conducta de nuestros prójimos humanos, deducimos de ella la conciencia y la libertad, algo que no advertimos ni reconocemos en los animales más cercanos a nosotros⁵⁶.

Aplicado este mismo proceder a las máquinas inteligentes, RK considera que acabaremos por aceptar que esas entidades no biológicas son conscientes, ya que tendrán un comportamiento similar a los humanos a quienes atribuimos conciencia. Es más, los robots llegarán a hablar y a asegurarnos que son conscientes, que entienden lo que les decimos, y que tienen capacidad para interactuar con otras entidades inteligentes y conscientes. ¿Por qué no les vamos a creer, a pesar de que no tendremos nunca acceso directo a su subjetividad, lo mismo que nos pasa con los otros humanos?⁵⁷.

⁵⁴ *Ibíd.*, p. 441.

⁵⁵ *Ibíd.*, p. 441; David CHALMERS, *Conscious Mind: In Search of a Fundamental Theory*, Oxford, Oxford University Press, 1996; J. SEARLE, *El misterio de la conciencia*, Barcelona, Paidós, 2000, c. 6.

⁵⁶ *Ibíd.*, p. 433.

⁵⁷ *Cfr. Ibíd.*, pp. 441 y 434.

Sobre este punto, RK se detiene en diferentes partes de sus obras a recoger las críticas de sus adversarios, y a dialogar con ellos. En relación al tema de la IA fuerte, y sobre todo a la conciencia, sus diálogos más amplios los tiene especialmente con J. Searle, pero también con Penrose y Hameroff. La confrontación con Searle la plantea de forma recurrente, en varios momentos de sus libros, pero la recoge de forma amplia en el capítulo en que dialoga con sus críticos⁵⁸. Como se sabe, Searle, apoyado en las tesis emergentistas, entiende que la mente humana consciente es consecuencia de las propiedades biológicas de nuestro cuerpo, sobre todo del cerebro. Por ello, defiende que los robots que maneja la IA fuerte nunca tendrán conciencia, porque no poseen ni poseerán la dimensión semántica del lenguaje; incluso, más recientemente también niega que puedan tener la dimensión sintáctica, y menos aún, la pragmática⁵⁹. Por eso, como demuestra en su experimento mental denominado *la habitación china*, una máquina programada para simular que habla chino, actuará como si hablara chino, pero no lo habla en realidad, porque no entiende el chino, ni se da cuenta de lo que le preguntan y contesta. Porque, está claro para Searle, no es lo mismo actuar *como si* supiera chino, que hablar realmente chino. La referencia a este experimento mental se ha hecho clásica, y RK se detiene en ella de forma amplia⁶⁰.

RK coincide con otros muchos críticos de Searle en considerar que si la habitación china se hace más compleja, llegará un momento en que estará dotada de dimensión semántica y entenderá chino. Claro que una “habitación” pequeña y poco compleja no podrá entender chino, pero sí lo entenderá si la estructura se hace suficientemente compleja. Lo mismo que en el cerebro, considera RK, las neuronas no piensan ni son conscientes, sino que quien piensa es el cerebro en su conjunto (patrón), también una estructura compleja podrá pensar y ser consciente⁶¹. Eso lo tiene claro Searle, pero también piensa que sólo determinadas estructuras, las de los cerebros biológicos humanos, han demostrado capacidad de pensar y ser conscientes. La carga de la prueba, por tanto, no la tienen los emergentistas que niegan que otras estructuras artificiales puedan llegar a ser conscientes, sino el que defiende la IA fuerte, como RK. Es decir, la cuestión no está en decir: ¿por qué no pueden pensar las máquinas?, sino ¿por qué sí creemos que pueden ser conscientes? Los cerebros humanos ya han demostrado que son conscientes. Los que tienen que demostrarlo son precisamente los robots. Y, de momento, sólo les queda a los que piensan que sí confiar en que la tecnología funcione en el futuro, y pedir esa misma confianza a los demás.

RK nos dice efectivamente que hay que *tener fe* en que habrá en el futuro máquinas tan inteligentes y parecidas a los humanos que tendrán conciencia.

⁵⁸ Cfr. *Ibíd.*, cap. 9; R. KURZWEIL, *Cómo construir una mente*, cap. 11.

⁵⁹ Cfr. J. SEARLE, *El misterio de la conciencia*, cit.

⁶⁰ Cfr. R. R. KURZWEIL, *La singularidad está cerca*, pp. 526-528.

⁶¹ R. KURZWEIL, *Cómo construir una mente*, p. 6.

Hay que tener fe, porque de momento nuestra tecnología está lejana⁶². RK ya la tiene, y está convencido de que en el futuro los humanos podrán construir robots que posean conciencia, libertad, responsabilidad y sistemas morales, por lo que podrán ser tratados como auténticas personas. Con toda claridad, RK concluye afirmando:

mi postura es la de aceptar que entidades no biológicas que sean completamente convincentes en sus reacciones emocionales serán personas, y mi predicción es que el consenso social también los aceptará. Nótese que esta definición va más allá de las entidades que puedan pasar el test de Turing, el cual exige dominar el lenguaje humano. Estas entidades son lo suficientemente parecidas a los humanos como para que yo los incluyera dentro del término persona, y creo que la mayor parte de la sociedad también lo haría, pero también englobo a entidades que evidencien reacciones emocionales como las humanas y no sean capaces de pasar el test de Turing, por ejemplo, niños pequeños⁶³.

También se desmarca RK de los planteamientos de Penrose sobre la IA. Para el físico y matemático inglés, la IA no puede considerarse igual a la inteligencia humana, porque la IA se reduce a un funcionamiento digital, mientras que la inteligencia humana actúa de forma intuitiva, capaz de resolver problemas difíciles sin recurrir a procesos algorítmicos⁶⁴. Esta primera crítica de Penrose a la IA, incluso de tipo débil, ha sido rechazada por la mayoría de los autores, no sólo los funcionalistas (incluso el propio Searle) y defensores de la IA, y de hecho Penrose no ha insistido en ello. RK se refiere sobre todo a las críticas de obras posteriores de Penrose, hechas desde una visión cuántica del funcionamiento del cerebro, y apoyándose en las investigaciones del anestesista Stuart Hameroff⁶⁵. Para Penrose, la cuestión de la emergencia de la mente consciente desde el cerebro no puede explicarse desde un nivel del funcionamiento de las neuronas, sino desde sus componentes, los microtúbulos, realizando, como indica RK, “una exótica forma de computación llamada “computación cuántica””⁶⁶. Este tipo de computación es el responsable, para Penrose, de la aparición de la conciencia, sin que pueda darse sin el funcionamiento del nivel cuántico.

También critica RK la tesis de Penrose sobre la imposibilidad de cargar y traspasar todo el contenido y programa de una mente humana a una máquina inteligente, ya que eso sería imposible, debido a que el *estado cuántico* de una persona cambiará muchas veces durante el tiempo que dure el escaneo. Pero esta dificultad la considera RK inconsistente, porque entiende que esas osci-

⁶² *Ibíd.*, pp. 200-205.

⁶³ *Ibíd.*, p. 213.

⁶⁴ R. PENROSE, *La nueva mente del emperador*, Madrid, Mondadori, 1991.

⁶⁵ Cfr. R. PENROSE, *Las sombras de la mente*, Barcelona, Crítica, 1996; *Lo grande, lo pequeño y la mente humana*, Madrid, Cambridge University Press, 1999; *El camino de la realidad. Una guía completa de las leyes del universo*, Barcelona, Debate, 2006.

⁶⁶ R. KURZWEIL, *La singularidad está cerca*, p. 517.

laciones del *estado cuántico* no sólo se estarían produciendo durante un escaneo, sino que ocurren de forma continua en nuestra existencia ordinaria. Y, a pesar de ello, Penrose no considera que cada oscilación del estado cuántico haga cambiar nuestra identidad personal⁶⁷.

Pero la propuesta más consistente de Penrose, aunque de momento es sólo una propuesta hipotética, se orienta a considerar que la explicación de la emergencia de la conciencia no puede lograrse sólo desde el nivel mecanoclásico, sino desde la conjunción de ese nivel con el cuántico, puesto que sólo la mecánica cuántica nos muestra cualidades de funcionamiento de la realidad similares a las de la conciencia: unificación de sucesos y no sometimiento a los determinismos de la materia en su nivel supra-atómico. De ahí que las investigaciones sobre el origen y el funcionamiento de la conciencia tienen que investigarse en la confluencia balanceada entre la dimensión cuántica y la mecanoclásica⁶⁸.

Las opiniones de los críticos sobre las propuestas de Penrose han sido muy dispares⁶⁹. La de RK es claramente negativa, pues considera que las propuestas de Penrose son difusas y oscuras, al apelar a fenómenos cuánticos que no están comprobados ni demostrados. Pero, aun en el caso de que las tesis de Penrose se demostraran verdaderas, RK considera que no harían más que retrasar algunos años el avance de la IA hacia la reproducción de robots con conciencia. No obstante, las tesis de Penrose se mueven en un terreno tan poco convincente para RK, que recoge las irónicas palabras de Chr. Koch sobre ellas: “la mecánica cuántica es misteriosa y la conciencia es misteriosa. Q.E.D: la mecánica cuántica y la conciencia deben estar relacionadas”⁷⁰. En definitiva, parece más bien, en opinión de RK, que el planteamiento de Penrose está basado simplemente en pedir un acto de fe en unas propiedades cuánticas misteriosas⁷¹.

3.5. *La libertad, la responsabilidad y la ciberética*

Para RK está claro que los robots del futuro no sólo poseerán conciencia, sino también la capacidad creativa que otorga la posesión de libertad, como los seres humanos. Claro que se trata también de una afirmación ambigua, más ambigua incluso que la conciencia. Al menos en el caso de la conciencia, como ya hemos visto, se detenía, aunque de forma poco convincente, en dialogar críticamente con diversos filósofos de la mente que discrepaban de sus planteamientos. No es el caso de la libertad, donde sus referencias son

⁶⁷ Cfr. *Ibíd.*, p. 225.

⁶⁸ Cfr. R. PENROSE, *Las sombras de la mente*, cit.

⁶⁹ Cfr. Javier MONSERRAT, “El problema del soporte físico de la sensibilidad-conciencia”, cit.

⁷⁰ R. KURZWEIL, *La singularidad está cerca*, p. 517.

⁷¹ Cfr. KURZWEIL, *Cómo crear una mente*, pp. 197-198.

más breves y superficiales. En *La singularidad está cerca*⁷² hace referencia al funcionamiento del cerebro, asemejándolo al comportamiento de los *fractales*. Un fractal, nos explica RK, es un diseño en el que un elemento (iniciador) es reemplazado por múltiples elementos (efecto generador); y cada elemento de este segundo nivel da lugar a otros múltiples, y así sucesivamente. Pero hay dos tipos de fractales: determinísticos y probabilísticos, siendo éstos los que añaden un elemento de incertidumbre. Así, el cerebro funciona, según RK, como un *fractal probabilístico*.

Esto se relaciona con lo que hemos indicado más arriba, cuando, en contra de Demsky, se definía como *materialista o patronista de la capacidad*, puesto que entiende que el cerebro, dotado por el ADN, y el proceso epigenético posterior, de una serie amplia de potencialidades, desarrolla sus muchas capacidades en diálogo con el entorno y los demás seres humanos⁷³. Esas capacidades que va adquiriendo, no sólo intelectuales, sino también emotivas y morales, no son consecuencia de una dinámica cerrada, sino abierta, donde se halla presente la libertad. En definitiva, me voy creando a mí mismo a través de las decisiones que voy tomando en esa red de interacciones con el ambiente natural y humano⁷⁴. RK considera que esa capacidad que tenemos los humanos se podrá transmitir a los robots.

El problema está en que una cosa es tener capacidad de realizar una variedad de posibles modos de acción, y otra entender que eso equivalga a tener libertad. La idea de libertad que maneja RK es un tanto problemática. Al igual que la consciencia, la idea de libertad no es fácil de definir ni de demostrar. Junto a quienes la niegan y quienes la defienden, están los *compatibilistas*, que intentan conjugar el ámbito de la realidad material, sometida a leyes inalterables, con el ámbito interno de subjetividad, donde se experimenta la libertad. Pero RK considera que ese modo de entender las cosas no resulta adecuado para defender la libertad, tal y como la experimentamos en nuestra vida cotidiana. Ahora bien, por otro lado, parece que su idea de libertad se reduce a la propuesta de Wolfram, para quien el mundo es como un gran *autómata celular*, de tal modo que, aunque todo está determinado, es tan complejo, que no puede saberse el futuro. Todo queda, por tanto, abierto a la indeterminación. Y esto, según RK, es suficiente como para que se dé el libre albedrío. En definitiva, RK considera que los humanos tienen libre albedrío, y, aunque no se pueda demostrar, tenemos que actuar como si fuésemos libres⁷⁵. Necesitamos, por tanto, ejercer un permanente acto de fe sobre nuestra libertad. “Mi propio acto de fe, dice RK, es que creo que los humanos poseen libre albedrío, y como actúo como si fuese el caso, me encuentro fuertemente condicionado a la hora de encontrar ejemplos entre mis propias decisiones que ilustren este

⁷² R. KURZWEIL, *La singularidad está cerca*, pp. 51-52.

⁷³ *Ibíd.*, pp. 544-545.

⁷⁴ *Cfr. Ibíd.*, p. 437.

⁷⁵ *Cfr. R. KURZWEIL, Cómo construir una mente*, p. 227.

hecho”⁷⁶. Así, cuando piensa en la decisión de escribir un libro, ve que no ha sido una decisión totalmente libre, sino que “la idea del libro lo decidió por mí”, viéndose a merced de ideas que parecen implantarse en mi mente por sí mismas, por encima y antes de mi decisión. A la vista de la complejidad con la que se muestra la toma de decisiones, grandes o pequeñas, RK se muestra escéptico respecto a la libertad. Y afirma: “aunque comparto con Descartes su confianza en que soy consciente, no estoy tan seguro del libre albedrío”⁷⁷. De ahí que le resulte más convincente la idea de Schopenhauer de que “podemos hacer lo que deseamos, pero en todo momento de nuestras vidas podemos *desear* solamente una cosa en concreto y absolutamente nada más que dicha cosa”⁷⁸. De todos modos, acaba concluyendo RK: “continuaré actuando como si tuviera libre albedrío y continuaré creyéndomelo mientras no tenga que explicar por qué”⁷⁹.

Como puede verse, esta idea de libertad es confusa y poco convincente, porque una cosa es la indeterminación (como se da en el juego de dados, o en la posición de un electrón alrededor de su núcleo, y no por eso decimos que los dados o los electrones tienen libertad) y otra es la libertad humana. La indeterminación nos abre simplemente a una variación de opciones no dirigidas por un ser consciente, mientras que en el ámbito de la libertad, por más que la libertad humana siempre es limitada y mundana, cada sujeto decide sus actos y los encamina hacia el objetivo elegido. De todos modos, estas dudas sobre la libertad no le impiden a RK defender claramente la responsabilidad de los humanos sobre los demás y sobre la obligación de construir un mundo justo y solidario, digno del ser humano. De ahí que en varias páginas de sus libros se detenga a reflexionar sobre las importantes consecuencias éticas y sociales de las nuevas tecnologías, tanto en el ámbito de la genética como de la nanotecnología y de la robótica. Basta seguirle en el análisis exhaustivo que realiza, en el capítulo ocho de *La singularidad está cerca*⁸⁰, dedicado a las promesas y peligros de los avances en GNR (genética, nanotecnología, robótica).

En la medida en que, para RK, el mundo del futuro próximo de los robots inteligentes estará formado por entes que serán iguales que los humanos, estarán dotados no sólo de consciencia, sino también de libertad. Por lo que habrá que construir entonces una *ciberética*, o *biocibernética*, dedicada a regular y a orientar las relaciones entre hombres y robots, y entre los mismos robots⁸¹. El punto central de la ética humana es no hacer daño al otro, a cualquier realidad consciente, y, por tanto, tenemos una normativa para evitar

⁷⁶ *Ibíd.*, pp. 227-228.

⁷⁷ *Ibíd.*, p. 228.

⁷⁸ Cfr. Arthur SCHOPENHAUER, *On the Freedom of the Will*, 1839 (cita tomada de R. R. KURZWEIL, *Cómo construir una mente*, p. 228).

⁷⁹ R. KURZWEIL, *Cómo construir una mente*, p. 228.

⁸⁰ R. KURZWEIL, *Las singularidades están cerca*, pp. 449-490.

⁸¹ Cfr. *Ibíd.*, p. 434.

hacer daño a los otros, en la medida en que nuestros actos son capaces de hacer sufrir y herir sus sentimientos. Pues esos mismos principios, nos dice RK, tienen que valer para las relaciones con los robots, de ellos hacia nosotros y entre sí. En la medida en que los robots afirmarán que son conscientes y poseen sentimientos, tendrán todo el derecho del mundo a que no se les haga sufrir, no se les desconecte, y no se les destruya. En relación a ello, RK hace referencia al caso del abogado Martine Rothblatt, que propuso en 2003 una moción fingida para evitar que una corporación desconectara a un ordenador consciente, simulando también un juicio fingido de *biociberética* sobre las razones a favor o en contra acerca de la moralidad de dicha pretendida desconexión⁸².

3.6. *La identidad personal*

En definitiva, como ya vimos que reconocía RK, el problema más importante en relación con la mente y la conciencia es la cuestión de nuestra identidad, la cuestión de ¿quién soy yo? “Esta es la cuestión ontológica definitiva y normalmente nos referimos a ella como la cuestión de la conciencia”⁸³. Claro que esa afirmación, en el sentido profundo de la ontología antropológica, no es del todo correcta. La respuesta a la cuestión de *quién soy yo* no se reduce a la cuestión de la conciencia. En algunos filósofos como Descartes y Sartre sí ocurría eso (la esencia, o lo más valioso del yo, es la conciencia), pero para Husserl y la mayoría de los filósofos, la cuestión de mi identidad se refiere al núcleo central de mi personalidad, que tiene como uno de sus rasgos fundamentales ser consciente. La conciencia es una cualidad del yo, no la esencia del yo. Por tanto, lo central de este problema es conjugar la base ontológica que me hace ser yo, alguien insustituible, y la cualidad de ese núcleo personal que me permite ser consciente de mí mismo. Las dos preguntas que focalizan este problema son *qué soy yo* y *quién soy yo*. La primera (¿qué soy yo?) parece referirse al núcleo ontológico de mi persona, mientras que la segunda (¿quién soy yo?) se refiere a la conciencia, a la capacidad refleja de ser consciente de *que soy* (constatación existencial), y, al menos en parte, de *qué soy* (contenido esencial de mi persona).

RK dedica muchos momentos de reflexión en sus libros a esta cuestión, y una sección concreta en *La singularidad está cerca*⁸⁴ a tratar de resolverla. Ya hemos visto que, para RK, lo que caracteriza a una conciencia humana no es tanto la base biológica con la que está hecha, sino la estructura o *patrón* con la que está organizada su biología. De ahí que, aunque cambiemos poco a poco las diferentes partes de nuestro cerebro y cuerpo, si no cambia el patrón, permanece invariable mi conciencia y mi personalidad. De hecho, nuestra base

⁸² Cfr. *Ibíd.*, p. 435.

⁸³ *Ibíd.*, p. 441.

⁸⁴ Cfr. *Ibíd.*, pp. 438-443.

biológica está cambiando continuamente, como nos dice la biología. También cambia nuestro patrón, en cierta medida, porque nuestro yo tiene la capacidad de hacerse a sí mismo a través de sus decisiones existenciales. Pero, a pesar de ello, nuestra personalidad no cambia. Es decir, somos el mismo yo, pero de diferente forma; somos el mismo *quién* pero diferente *qué*; somos *el mismo*, pero no *lo mismo*. Hay, por tanto, una continuidad dentro de la variación constante. Aunque no sepamos discernir qué es lo que permanece y qué cambia, ni sepamos distinguir entre esas dos caras de la realidad⁸⁵. Zubiri denomina a estas dos facetas la *personidad* y la *personalidad*; la *personidad* sería la estructura de notas esenciales que nos constituye como humanos, siendo similares en todos los humanos; y la *personalidad*, la diferente figura o forma que toma la *personidad* en cada individuo, y en cada momento concreto de su existencia⁸⁶.

RK no se detiene en hacer estas distinciones, y por eso se mueve dentro de un lenguaje ambiguo, aunque implícitamente las tiene en cuenta para defender su propio planteamiento. Lo que a RK le interesa es la cuestión de si, cuando se puedan construir robots inteligentes, se podrá copiar y traspasar también la conciencia (para él, la persona entera) de un individuo humano a un robot, con lo que podría decirse que permanece la misma persona, aunque entonces su *patrón* o *estructura* estén apoyados en una base material diferente, ya no biológica⁸⁷. La prueba definitiva para ver si es la misma conciencia o persona consistirá, según RK, como ya hemos visto en páginas anteriores, en si ese robot es consciente y hay en él una continuidad personal con el individuo humano del que supuestamente es copia. Claro que nunca estaremos del todo seguros que será el mismo yo, puesto que una cosa es la copia y otra el original; como tampoco sabremos si una vez borrado el original, y manteniéndose la copia, podremos decir que no se ha perdido nada. RK está convencido de que no; seguirá existiendo la misma persona. Y además, la copia robótica tiene la ventaja, para RK, de que, aparte de mantener la identidad, pueden construirse diversas copias de mi yo, diversas personalidades, lo que nos va a permitir enriquecer nuestra personalidad, más que perderla o empobrecerla.

No es cuestión de volver a las disquisiciones anteriores sobre si esa conciencia robótica es auténtica o no, puesto que para RK no queda más prueba que el propio testimonio del robot, que, recordemos, tendrá una potente inteligencia, será consciente y podrá hablarnos y tratar de convencernos. Su testimonio será tan creíble como el de cualquier persona humana. ¿Por qué no tendríamos que creerlo, si creemos a los demás humanos? De esta total igualdad entre la conciencia humana y la robótica RK deduce que no habrá ningún obstáculo para atribuir a los robots la condición de *personas*, con todo el valor

⁸⁵ Cfr. Paul RICOEUR, *Sí mismo como otro*, Madrid, Siglo XXI, 1996.

⁸⁶ Cfr. Xavier ZUBIRI, *El hombre y Dios*, Madrid, Alianza/Sociedad de Estudios y Publicaciones, 1984; *Sobre el hombre*, Madrid, Alianza/Sociedad de Estudios y Publicaciones, 1986.

⁸⁷ Cfr. R. R. KURZWEIL, *La singularidad está cerca*, pp. 222-229.

y la dignidad que atribuimos a las personas humanas. De ahí la necesidad de construir unas reglas éticas (*ciberética* o *biociberética*) que rijan las conductas entre ellos, y de los robots con los humanos, como hablamos más arriba.

El problema está en cómo solucionar dos aspectos. El primero se refiere a cómo estar seguros de que esos robots tendrán conciencia y libertad, y, por ello, podrán ser considerados personas, con todas las consecuencias. Ya vimos que, en nuestra opinión, la carga de la prueba no está en los que lo niegan, sino en los que lo afirman. Y un segundo problema se refiere a la ontología de la persona humana. Lo que constituye a una persona no es sólo la estructura o el patrón con los que está hecha, sino también la base material estructurada por el correspondiente patrón. No se trata de un patrón que se independiza de su materia, y por eso puede trasplantarse. En algunos tipos de estructuras quizás se podrán separar ambos aspectos, pero en el caso de la personalidad humana no parece razonable imaginar dicha separación, como opina el dualismo.

En esta línea se sitúan las objeciones de M. Denton, biólogo de la Universidad de Otago (Nueva Zelanda), que acusa a RK de poseer un concepto muy superficial de la naturaleza de lo biológico. Denton considera que la ontología de la biología no será nunca superada por la ontología de las máquinas artificiales materiales, en la medida en que los humanos, por ser entidades biológicas, a diferencia de las máquinas, somos realidades *holísticamente organizadas*. Las máquinas son realidades compuestas de partes, organizadas de abajo a arriba, de las partes al todo. En cambio, los seres vivos son totalidades que se van conformando por saltos evolutivos, dando lugar a especies o totalidades nuevas, en las que el todo está presente en cada una de las partes. De ahí que Denton defiende que la ontología de la biología no será nunca superada por la ontología de las máquinas artificiales materiales, en la medida en que la condición humana, a diferencia de las máquinas, y por su específica estructuración emergida en el proceso evolutivo, está dotada para abrirse al misterio, a la dimensión religiosa y trascendente, tanto en el ámbito de lo humano como del conjunto del universo.

Pero RK se defiende, y aunque está de acuerdo con Denton en admirar el misterio de la condición humana, no cree que haya tanta diferencia entre los hombres y las máquinas del futuro. RK entiende que los sistemas biológicos no son totalmente holísticos, ni tampoco las máquinas actuales completamente modulares, sino que son dos modos extremos de funcionar que admiten estados mixtos e intermedios. Hay aspectos dentro de los humanos que funcionan como sistemas modulares, y en las máquinas del futuro se darán procesos de naturaleza holística, al estilo de los seres vivos. De ahí que considere que "es engañoso sugerir que todos los aspectos del cerebro humano interactúan con todo el resto de aspectos y que por eso es imposible comprender sus métodos"⁸⁸. Y, por otro lado, los robots del futuro serán

⁸⁸ *Ibíd.*, p. 551.

capaces de funcionar con estructura holística, de modo que podrán tener las mismas cualidades que Denton considera que son exclusivas de los humanos (seres autoorganizativos, autorreferenciales, autorreplicantes, recíprocos, autoformativos y holísticos), puesto que estos ordenadores serán construidos siguiendo los mismos procedimientos que se han seguido en la formación de los humanos⁸⁹.

En uno de los diálogos tenidos entre Bill Gates y RK, y recogidos en *La singularidad está cerca* con permiso del primero, Gates le plantea la dificultad de que una máquina pueda tener conciencia y libertad. Para él es evidente que la conciencia de una máquina no será igual que la de los humanos. En concreto, no cree que una inteligencia de silicio (o de cualquier material con el que se hagan las máquinas del futuro) “vaya a ser consciente en sentido humano”⁹⁰. Aunque a su inteligencia se la pueda denominar conciencia, será de otro tipo distinto al humano. Y, en ese sentido, B. Gates considera que “cada uno de nosotros poseemos una individualidad diferenciada que no puede ser evitada”⁹¹. Pero esta supuesta dificultad es considerada por RK una ventaja para las máquinas, en la medida en que esa es precisamente una limitación de las mentes biológicas, ser sólo de una manera. En cambio, “la inteligencia “de silicio” puede comportarse de las dos maneras”⁹², del modo robot y humano, mientras que éstos, sólo de una. Además, los robots pueden fusionarse con otros robots y poseer múltiples personalidades, si así lo desean, mientras que los humanos sólo pueden ser ellos mismos, y sólo de forma efímera podemos fusionarnos con otros. B. Gates advierte: “Todo lo que tiene valor es efímero”, a lo que RK le responde que esas realidades de valor efímero (los humanos) serán sustituidas por otras (los robots) de valor todavía más grande⁹³.

3.7. *¿Post-humanismo o humanismo extendido?*

A pesar de las razones que presenta contra todas las objeciones, RK piensa que el debate seguirá, y no cesará hasta que los ordenadores acaben superando el test de Turing. Y quizás ni siquiera entonces. Pero el avance de la tecnología robótica seguirá imparabile, y conseguirá realizar tareas cada vez más difíciles y antes impensables, relevando a los humanos en muchos aspectos. Ahora bien, esta sustitución no se llevará a cabo haciendo desaparecer a los humanos de forma violenta ni trágica, sino pacífica y muy fructífera, según la opinión de RK. No se tratará de una invasión alienígena que venga de Marte.

Estamos creando estas herramientas para convertirnos a nosotros mismos en seres cada vez más inteligentes. Estoy convencido de que la mayoría de

⁸⁹ Cfr. *Ibíd.*, pp. 550-551.

⁹⁰ *Ibíd.*, p. 430.

⁹¹ *Ibíd.*, p. 431.

⁹² *Ibíd.*, p. 431.

⁹³ *Ibíd.*, p. 431.

los analistas estarán de acuerdo conmigo en que esto es lo que hace única a la especie humana, el hecho de que construimos estas herramientas para aumentar nuestro propio alcance⁹⁴.

De este modo se habrá iniciado la etapa post-humana o trans-humana, aunque a RK no le gustan estos términos, porque tienen el peligro de suponer una superación de los humanos ante el empuje de los robots. Esta IA que se dará en el futuro, “continuará formando parte de la civilización humana, que ya es una civilización humano-máquina. En otras palabras, las máquinas futuras serán humanas aunque no sean biológicas. Este será el próximo paso evolutivo, el próximo cambio de paradigma de alto nivel, el siguiente nivel de indirección”⁹⁵. En definitiva, no se trata, para RK, de enfrentar a robots y humanos, y de sustituir a éstos por aquéllos, sino de componer una cultura en la que se complementen los humanos con los robots, conformando un *humanismo extendido* (Monserrat), en que el concepto de persona se ampliará y se enriquecerá con los nuevos matices que aporten las máquinas inteligentes y los *cyborgs*. Y aunque éstos sean más capaces y potentes que aquéllos, les tendrán un gran respeto y veneración⁹⁶, y se impondrán las nuevas normas de la *ciberética*, como ya vimos.

Es importante hacer ver el esfuerzo que RK pone en dejar claro que esta época supuestamente *post-humana* no supondrá un dejar atrás los valores y cualidades de la época humanista, sino que tenemos que ser conscientes de que uno de los aspectos más centrales de la condición humana es poder traspasar sus propios límites, *extender su humanidad*, en este caso a las máquinas⁹⁷. Así, RK nos dice que un humano modificado no ha de ser considerado como no humano, porque ¿dónde pondríamos la frontera o la línea divisoria entre lo humano y lo no humano en los avances de la tecnología de mejora y de transformación humana en el ámbito de la medicina y otros ámbitos? En definitiva, las máquinas del futuro no sólo no dejarán de ser humanas, sino que “serán todavía más humanas que los humanos de hoy”⁹⁸. Y lo serán porque tendrán más capacidad intelectual y emotiva, y una moral más profunda que los humanos, en la medida en que serán humanos mejorados, diseñados para ser mejores, y no sólo en el ámbito de la inteligencia instrumental, sino, en opinión de RK, también en el terreno de lo emotivo y moral. Y a quienes consideren que eso será lo mismo que hacer emerger una nueva *especie*, RK responde que no es correcto utilizar ese concepto, puesto que es sólo válido en el mundo de la biología, y esta nueva época, con humanos de otra manera,

⁹⁴ R. KURZWEIL, *Cómo construir una mente*, pp. 262-263.

⁹⁵ R. KURZWEIL, *La singularidad está cerca*, p. 33.

⁹⁶ Cfr. *Ibíd.*, p. 36.

⁹⁷ Cfr. *Ibíd.*, p. 429.

⁹⁸ *Ibíd.*, p. 433.

los robots, habrá superado el ámbito de la biología para adentrarse en una nueva era en la historia de la tierra y del universo⁹⁹.

3.9. *El futuro del universo y la dimensión religiosa: trascendencia y espiritualidad*

Este paradigma post-humano no se limita al ámbito de nuestro planeta, sino que tiene dimensiones cosmológicas. En oposición a S. Pinker, RK entiende que la inteligencia es el objetivo último que persigue no sólo la evolución biológica, sino también el dinamismo profundo del universo. Parece claro que la evolución persigue ir construyendo seres vivos con cerebros cada vez más grandes, plásticos e inteligentes, y eso se advierte en el terreno tanto de la hominización como de la humanización, es decir, en el proceso del avance de la cultura, en el que se da una dinámica dialéctica en la que el ser humano va construyendo artefactos, que son expresión de la inteligencia humana, al mismo tiempo que potenciación de la misma. De este modo, como culmen del desarrollo de nuestra inteligencia, terminaremos construyendo inteligencias complementarias a nuestros cerebros, como son los ordenadores, y toda la tecnología inteligente. En el momento de la *Singularidad*, seremos sustituidos por la nueva inteligencia de los robots, y será el momento, en opinión de RK, en el que podremos también salir a poblar el universo, superando las dificultades que tal empresa lleva consigo, como son salvar las grandes distancias estelares, sin poder superar la –al parecer– barrera insalvable de la velocidad de la luz. Pero el optimismo de RK le lleva a imaginar que podemos encontrar algún atajo, como es viajar a través de los elementos que nos proporcionan la teoría de cuerdas o de gusanos. En ese momento es cuando se producirá, como ya hemos indicado, un momento singular, una singularidad inteligente, en la que la capacidad de inteligencia artificial se disparará en potencialidad, de tal modo que daremos inicio a una nueva era, la post-humanidad o trans-humanidad.

RK valora el esfuerzo de los investigadores de la Astrobiología, y la organización (SETI) que lidera estos esfuerzos por hallar vida, y vida inteligente, en planetas fuera de nuestro planeta y en el conjunto del universo, apoyándose en la denominada ecuación de Drake¹⁰⁰. Todos los esfuerzos que se han hecho han resultado hasta ahora baldíos, haciéndose más evidente la denominada paradoja de Fermi (si hay vida inteligente en el universo, ¿por qué no la vemos?). RK, sin negar que en el futuro se pueda encontrar algún modo de toparse con vida inteligente en algún lugar del universo, se inclina más bien a pensar que estamos solos en el universo¹⁰¹, en lo que coincide con otros muchos pensadores¹⁰². Sin embargo, nuestra capacidad de crear máquinas

⁹⁹ Cfr. *Ibíd.*, 429.

¹⁰⁰ Cfr. *Ibíd.*, pp. 395 y ss.

¹⁰¹ Cfr. *Ibíd.*, pp. 392-394; 409-410.

¹⁰² Cfr. Paul DAVIES, *Un silencio inquietante. La nueva búsqueda de inteligencia extraterrestre*, Barcelona, Crítica, 2011; John GRIBBIN, *Solos en el universo. El milagro de la vida en la tierra*, Barce-

inteligentes más potentes que nuestros cerebros nos capacitará para salir de nuestro planeta e ir poblando progresivamente el resto del universo. Pero no se tratará, piensa RK, de mandar al espacio humanos biológicos, sino “sondas robóticas, ya que en ese momento la civilización humana será, en todos los sentidos prácticos, no biológica”¹⁰³. Se tratará de sondas robóticas más bien pequeñas, porque “estarán compuestas mayoritariamente de información”, y podrán moverse con el tiempo a una velocidad mayor que la luz, burlando esa barrera a través, quizás, de los agujeros de gusano o cualquier otro resquicio que la cosmología nos muestre en el futuro¹⁰⁴.

Así, más que decantarse por la teoría de los multiversos, con la que se explicaría de modo científico el *principio antrópico* (PA), sin tener que recurrir a una solución metafísica basada en una hipótesis religiosa, RK defiende más bien, en la línea que estamos mostrando, que el universo en el que vivimos tiene su propia tendencia esencial y evolutiva que le lleva inevitablemente a producir vida inteligente. Así, recuerda que ya en su libro *The Age of Spiritual Machines*¹⁰⁵, defendía que “la inteligencia acabaría por permeear el universo y que decidiría el destino del cosmos”¹⁰⁶. Frente a la opinión generalizada de que la inteligencia humana es sólo una pequeñísima burbuja, que ha emergido por azar en un planeta periférico del cosmos, RK señala que “mi conjetura es que la inteligencia acabará por demostrar que es más poderosa que estas grandes fuerzas impersonales”¹⁰⁷. De ahí que el futuro del universo no esté sometido a las leyes impersonales y azarosas del cosmos, bajo el mero juego de la probabilidad, sino que “el destino del universo es una decisión que está por tomar, una decisión que consideraremos inteligentemente cuando llegue el momento de hacerlo”¹⁰⁸.

La historia del universo está conformada, según RK, como un proceso de impregnar de inteligencia todo el cosmos. Se trata de que esta inteligencia, que aparece en los humanos como una pequeña burbuja sin importancia, vaya conformando la materia y energía según sus planes; esto es, “convertir la materia tonta en materia lista”¹⁰⁹. No es que la materia lista se sitúe por encima de las leyes de la materia, sino que será tan inteligente que dominará dichas leyes para “manipular la materia y la energía a su antojo”, porque es evidente que “la inteligencia es más poderosa que la física”, “más poderosa que la cosmología”¹¹⁰. De esta forma, la inteligencia llegará a diseñar el

lona, Pasado y Presente, 2012.

¹⁰³ R. R. KURZWEIL, *La singularidad está cerca*, p. 403.

¹⁰⁴ *Cfr. Ibid.*, pp. 404-409.

¹⁰⁵ R. KURZWEIL, *The Age of Spiritual Machines*, Nueva York, Viking, 1999, pp. 258-260.

¹⁰⁶ R. KURZWEIL, *La singularidad está cerca*, p. 413.

¹⁰⁷ *Ibid.*, p. 414.

¹⁰⁸ *Ibid.*, p. 414.

¹⁰⁹ *Ibid.*, p. 417.

¹¹⁰ *Ibid.*, p. 417.

universo que quiera, y como plantea también J. Gardner, basándose en los planteamientos de RK, la inteligencia humana llegará a conseguir construir universos nuevos, distintos al nuestro, *universos bebés*¹¹¹. De esta forma, la inteligencia se irá imponiendo sobre la materia para ir haciendo del universo “un ordenador a escala universal”¹¹², al mismo tiempo que “un holograma gigantesco”¹¹³, en el que el todo no será la simple suma de sus partes, sino una estructura inteligente, que hace presente el todo en cada uno de sus elementos y partes, como los seres vivos. Y, siguiendo la propuesta de J. Gardner, “si fuera posible diseñar nuevos universos y establecer contacto con ellos, este hecho proporcionaría todavía más medios para que una civilización inteligente continuara expandiéndose”¹¹⁴. Pero RK considera que la expansión de la inteligencia humana a otros universos es problemática, en la medida de que la teoría de los multiversos da por hecho que no puede haber comunicación entre universos diferentes. Si la hubiera, ya no sería universos diferentes, sino mera expansión del único existente.

A la vista de estos planteamientos futuros tan fantasiosos y ambiciosos, no es de extrañar que a más de un lector le suene a pura ficción, y, sobre todo, vea en ello los signos de una nueva religión. Esa es la reacción de B. Gates en el ya citado diálogo con RK. Así, le objeta: “Tu optimismo es casi una fe religiosa”¹¹⁵. Y RK no lo niega, sino que afirma que “necesitamos una nueva religión”. Claro que la idea que tiene RK de la religión es un tanto especial y ambigua. Por de pronto, esta nueva religión seguirá cumpliendo algunas de las funciones que, según RK, han cumplido las religiones tradicionales, como es la de “racionalizar la muerte” (la nueva religión, según RK, como ya hemos visto, será capaz de superar la muerte, hacernos inmortales), además de dotarnos de normas de comportamiento (la regla de oro: “trata a los demás como te gustaría que te trataran a ti”) basadas en el respeto a la conciencia humana¹¹⁶. Pero la nueva religión tendrá un principio nuevo, secular, que consistirá, según RK, en “la importancia del conocimiento”¹¹⁷, que va más allá de la mera información, y que abarcará el respeto y el cultivo de todos los ámbitos del saber.

El modo de introducirse en esta nueva religión parece consistir en convertirse en un *singularista* (*Singularitarian*), creyente, defensor y contagiador de estas ideas (“alguien que entiende la Singularidad y que ha incorporado su significado a su propia vida”¹¹⁸). La comunidad o “iglesia” de los *singularistas*

¹¹¹ Cfr. J. GARDNER, *El universo inteligente. Una auténtica revolución: la inteligencia propia del cosmos*, Barcelona, Robin Book, 2008.

¹¹² R. R. KURZWEIL, *La singularidad está cerca*, p. 418.

¹¹³ *Ibid.*, pp. 418-420.

¹¹⁴ *Ibid.*, p. 420.

¹¹⁵ *Ibid.*, p. 429.

¹¹⁶ Cfr. *Ibid.*, p. 429.

¹¹⁷ *Ibid.*, p. 430.

¹¹⁸ *Ibid.*, p. 424.

no tendrá, con todo, un clero o conjunto de líderes que orienten a la masa de los nuevos creyentes, porque eso sería mantener algo que, en opinión de RK, "forma parte del antiguo modelo", sino algo más impersonal y eficaz, una especie de "ordenador carismático"¹¹⁹.

¿Y contará esta nueva religión con un Dios? Depende de qué entendamos por tal. RK sueña con un modelo panteísta de universo, que parte de estar, como hemos dicho, compuesto de materia tonta, física, para convertirse progresivamente de materia lista, inteligente. Y ese universo convertido en un ordenador inteligente vendrá a ser como un dios omnisapiente y omnipotente. Así pues, "una vez que saturemos la materia y la energía del universo con inteligencia, éste 'despertará', se volverá consciente y excelsamente inteligente. Eso es lo más cercano a Dios que puedo imaginarme"¹²⁰. Como puede verse, la dimensión espiritual y trascendente de los humanos y del universo se descubre en la capacidad de superar sus propios límites, trascender su realidad, y saber pasar de una inteligencia y conciencia biológica a otra artificial, así como ir trascendiendo de un universo físico a otro progresivamente inteligente, que en su meta final se convertirá en divino.

Este es el Dios en el que RK cree, como responde cuando un supuesto robot inteligente le pregunta por su fe¹²¹. No cree en Dios como "una persona consciente y todopoderosa que nos vigila", ni tampoco en el Dios del deísmo, que "lo crea todo y luego se retira", sino más bien en un universo que se va haciendo progresivamente inteligente. En definitiva, "podemos considerar que Dios es el universo y te he dicho que creo en la existencia del universo"¹²². Así, aunque Dios no es visto por RK como una realidad personal y trascendente al universo, sí parece tener cualidades similares a las personas, porque

el universo no es consciente (todavía). Pero lo será. Para ser precisos deberíamos decir que a día de hoy solo una pequeña parte de él es consciente. Pero eso va a cambiar, y pronto. Preveo que el universo se volverá enormemente inteligente y se despertará en la sexta era. Así, la única creencia que estoy expresando es que el universo existe. Si hacemos este acto de fe, la expectativa de que despertará no es tanto una creencia como una comprensión fundada basada en la misma ciencia que sostiene la existencia del universo¹²³.

¹¹⁹ *Ibíd.*, p. 430.

¹²⁰ *Ibíd.*, p. 430.

¹²¹ *Cfr. Ibíd.*, p. 446.

¹²² *Ibíd.*, pp. 446-447.

¹²³ *Ibíd.*, p. 447.

4. UNAS VALORACIONES CONCLUSIVAS

Tras el recorrido que hemos hecho de las ideas centrales de RK, sobre todo su concepción de ser humano, nos corresponde hacer un balance conclusivo de sus diferentes propuestas y aportaciones. Decíamos al principio que, a pesar de no estar de acuerdo con aspectos clave de su concepción antropológica y metafísica, en general, tenemos que reconocer que la obra de RK está llena de interesantes sugerencias, presentadas con un talante ponderado y cuidadoso, algo que no suele estar muy presente en pensadores que defienden ideas similares a las suyas.

Vimos en su momento cómo respetaba claramente la distinción de ámbitos entre la ciencia y la filosofía, sobre todo a la hora de hablar de la esencia de la conciencia, de la libertad y de la persona. Tenemos que repetir que no nos interesa entrar a calibrar y valorar las ideas de RK sobre los avances de la IA en su sentido más técnico, ni sus atrevidas propuestas de futuro en relación a este campo, así como a la construcción de programas tecnológicos que consigan simular los procedimientos de la inteligencia humana. Nos interesa sobre todo dilucidar sus ideas sobre la IA fuerte, y, por tanto, sus afirmaciones sobre la conciencia, la libertad y la esencia de la persona, así como sus propuestas metafísicas sobre el futuro del universo y su idea de Dios y la religión.

Indicamos ya en su momento que su idea de conciencia es demasiado ambigua, en la medida en que no se detiene suficientemente en analizar lo que entienden por tal los diversos filósofos de la mente, y, por otro lado, utiliza siempre la idea de conciencia como si fuera algo similar a la idea ontológica de persona. Ya vimos que la conciencia no es el núcleo sustancial u ontológico de la persona, que sería el yo, sino una cualidad de la persona, del yo. Pero en la medida de que lo que nos hace que seamos personas, con nuestra especificidad individual irreplicable, es al mismo tiempo lo que nos hace ser consciente y libres, en el fondo las carencias de sus propuestas hace que no sea demasiado relevante la distinción entre el yo (el núcleo sustancial de la persona) y la conciencia. La pretensión de RK es que en el futuro se podrán construir robots inteligentes que, además de superar el test de Turing, tendrán conciencia, libertad y responsabilidad ética, y, por ello, defiende que no ve por qué no podrán ser considerados y ser tratados como personas humanas. Personas humanas, aunque de otra manera, siguiendo su idea de *humanismo extendido*, en expresión de J. Monserrat.

Esta es la tesis más fuerte de RK, y precisamente es aquí donde se produce nuestra más fuerte discrepancia. Entendemos que, aunque RK pueda presentarnos muchas muestras del avance de la ingeniería informática y robótica, no hay razones suficientes para pensar que lo que constituye la raíz ontológica de cada yo, de cada persona humana, pueda reducirse a un mero programa informático, un patrón o estructura mental, al margen de su base biológica. Entendemos que, en la línea de la tesis emergentista, la mente humana es el resultado de la especial estructuración de la masa cerebral y del cuerpo

humano en general, pero tal estructuración no es una realidad al margen e independiente de su base biológica. Aunque estamos muy lejos de saber y de entender la relación entre el funcionamiento cerebral y el mental, entre la base fisiológica del cerebro y las experiencias subjetivas de la mente, parece evidente que se tienen que dar los dos elementos unidos para que emerja la conciencia. Y parece también claro que la materia cerebral es de naturaleza tan diferente al resto de los demás tipos de materia, que sólo ella es capaz de producir, dar de sí, la aparición de lo mental. De ahí la propuesta de Penrose de situar la explicación de esa emergencia de la mente a partir de entender el funcionamiento cuántico de la materia cerebral, en conjunción con el funcionamiento de la dimensión mecanoclásica del mismo, aunque con esta simple afirmación se esté todavía muy lejos de explicarnos cómo se produce este fenómeno tan complejo.

Por ello, por mucho que avancemos en perfeccionar los programas robóticos, lo máximo que podemos hacer es “implementar en la máquina ciertas rutinas de razonamientos o toma de decisión, pero no la capacidad en sí misma”¹²⁴. De todas formas, pensamos que, si afirmamos que una de las diferencias entre hombres y máquinas está en que los primeros podemos dudar ante diferentes opciones de actuación, mientras que las máquinas no tienen capacidad de dudar, sino que se limitan a ejecutar las tareas para las que están programadas (y lo pueden hacer mejor y más rápido que los humanos); y si seguimos afirmando que, como señala J. P. Núñez, “la actividad consciente ha sido ‘diseñada’ por la evolución, precisamente, para poder afrontar estas situaciones en las que desconocemos qué respuesta dar y, por tanto, tenemos que diseñar ad hoc, así como evaluar, sin criterio claro, su eficacia sobre la marcha”¹²⁵, seguramente que RK nos responderá enseguida que esa diferencia es adecuada para los robots actuales pero no para los del futuro, que tendrán capacidad de ser conscientes y de dudar, y de todo lo que los humanos posean.

Pero, en definitiva, desde nuestro punto de vista, los robots inteligentes no llegarán a entender las actividades para las que están programados, sino que seguirán, como los actuales, pero de forma más perfecta, las rutinas previstas en los programas con los que han sido construidos, como señala J. Searle en su experimento mental de la *habitación china*. La diferencia entre hombres y máquinas se parece a la que hay entre un actor ciego y otro que ve¹²⁶. Podrá el primero actuar de tal forma que parezca que ve. Pero si le cambian algún elemento del decorado, teniendo entonces que improvisar, no podrá hacerlo, porque no tiene la capacidad de captar el cambio que se ha producido y adaptarse a él, con lo que su incapacidad visual quedará al descubierto. RK nos podrá argumentar que las máquinas del futuro, en las que

¹²⁴ J. P. NÚÑEZ PARTIDO, “Más que monos y máquinas: la mente humana se resiste a la reducción”, p. 5.

¹²⁵ *Ibid.*, p. 8.

¹²⁶ *Cfr. Ibid.*, p. 8.

él piensa, también tendrán la capacidad de ver y de improvisar. Pero mucho nos tememos que, si se trata de máquinas que hay que programar, no tendrán la capacidad de ser autónomas y de autoprogramarse a sí mismas, sino de seguir las rutinas conductuales para las que han sido programadas (por más complejas y sutiles que sean).

En definitiva, el punto clave está, como ya lo dijimos en su momento, en darnos cuenta de que parece que la subjetividad, la conciencia y la libertad de los humanos son el resultado de un largo y complicado proceso evolutivo, en el que tan importante es la complejificación del entramado neuronal como la naturaleza biológica del mismo. Sólo los entes, como los humanos, de naturaleza biológica, parece que tienen capacidad de ser un yo, poseer *qualias*, experiencias subjetivas, tomar conciencia de sí y programar su vida a base de decisiones suficientemente libres, en medio de múltiples circunstancias que recortan nuestra libertad, pero no la suprimen, y que proceden tanto de nuestra naturaleza corpórea e inconsciente como del entorno natural y cultural. Creemos que la carga de la prueba para demostrar lo contrario le corresponde a quienes nos piden fe en que las máquinas del futuro serán igual que los humanos, para extender sobre ellas la condición de personas, y no tanto a los que no creemos, o dudamos seriamente, que eso pueda darse. No es suficiente argumentar ¿por qué no?, sino mostrar por qué sí, en qué se apoya esta pretensión y cómo se consigue.

Entendemos que la pretensión de RK de poder construir robots inteligentes que sean iguales a los humanos supervalora la capacidad de la ingeniería informática, e infravalora el crisol en el que se conforman los humanos: por un lado, el proceso evolutivo, en el que se han ido acumulando innumerables experiencias instintivas de supervivencia en nuestras habilidades corpóreas, tanto sensitivas, como emotivas, inconscientes y conscientes; y, por otro lado, el entorno cultural, interpersonal y social. Y es muy difícil entender que las máquinas puedan asimilar en su proceso de formación y creación un proceso tan complejo y tan específico como éste. De ahí que la sospecha que nos viene en muchos momentos al leer las obras de RK es que no tiene claro, ni da en el clavo de lo que nos hace humanos. No es un proceso informático e ingenieril, sino un proceso de hominización y, sobre todo, de humanización, que nos capacita para tomar conciencia de nosotros mismos y e ir haciéndonos a través de actos libres, pero en diálogo con los demás. De ahí que lo más específico de lo humano no es ser inteligentes (sobre todo en su sentido instrumental) y capaces de solucionar los difíciles problemas técnicos que la vida nos plantea, sino la capacidad de empatía y de compasión hacia el otro, que nos lleva a construir un mundo más acogedor y humano, y cómo al servicio de este proyecto se subordinan todas nuestras capacidades intelectuales y técnicas¹²⁷.

¹²⁷ Cfr. Michael TOMASELLO, *Los orígenes culturales de la cognición humana*, Buenos Aires, Amorrortu, 2003; del mismo, *Los orígenes de la comunicación humana*, Buenos Aires, Katz Editores, 2013.

Y si resulta problemática su idea de persona, también produce escepticismo y distanciamiento sus propuestas sobre el futuro del universo y su concepción del misterio y de la divinidad. Sus teorías sobre la posibilidad o no de hallar vida extraterrestre, la capacidad de los humanos y robots inteligentes para dominar y poblar el resto del universo, así como, en la línea de J. Gardner, la posibilidad de construir nuevos universos, universos bebés, son algo que corresponde discutir y desmontar o apoyar a los diversos ámbitos científicos, tanto a la astrobiología como a la cosmología. Lo que parece claro que estas hipótesis tan atrevidas y optimistas no son seguidas más que por un grupo más o menos amplio de científicos, frente a la indiferencia de la mayoría¹²⁸.

En cambio, resulta ya más criticable su concepción sobre la dimensión religiosa del ser humano, así como su idea de Dios. El talante comedido y defensor del humanismo le lleva a entender como elementos específicos de la condición humana la capacidad de abrirse a la trascendencia, entendida en múltiples direcciones, y percibir en el trasfondo último de la realidad la sombra del misterio y de la divinidad. De ahí que afirme que la dimensión religiosa es algo que pertenece a la naturaleza de lo humano. Pero está claro su empeño en desmarcarse de las diversas religiones que los humanos han cultivado a lo largo de su dilatada historia. Ya vimos que su idea de Dios se adscribía a un panteísmo cósmico, que atribuye al mundo una condición divina. A esta divinidad no le atribuye la cualidad de ser persona, pero sí la capacidad de ser omnipotente y omnisciente, puesto que el universo, al final de su historia, habrá sido convertido, como vimos más arriba, en un gran ordenador universal, cuando la inteligencia, emergida de forma inicial y precursora en la especie humana, haya conseguido imponerse sobre toda la materia física. Esta es para RK la gran meta de la historia, impregnar con un manto de inteligencia todo lo que hay y lo que pueda existir, por acción creadora de nuestro universo inteligente, meta y objetivo al que los humanos tienen que contribuir, como sublime vocación que justifica la existencia de la especie entera. No otra cosa propuso en su momento, aunque sin el despliegue de estas referencias ingenieriles ni cosmológicas, el filósofo alemán Max Scheler, para quien el universo en su totalidad es divino, pero su divinidad imperfecta (*Deitas*) y todavía no consciente tiene que convertirse a lo largo de la historia, y por empuje de los humanos, en una divinidad perfecta y consciente (*Dios*)¹²⁹. De este modo, RK se acerca a la teoría denominada *panprotopsiquismo*, que hemos visto que rechazaba a la hora de definir y entender la naturaleza de la consciencia humana. Para él, de forma similar a lo que defiende esa teoría, el psiquismo estaría ya hasta cierto punto presente en el trasfondo del universo desde el principio, para emerger más adelante en la

¹²⁸ Cfr. David, JOU, "La evolución cósmica, biológica y antropológica", en Emili MARLÉS (ed.), *Trinidad, universo, persona. Teología en cosmovisión evolutiva*, Estella, Verbo Divino, 2013, cap. 3.

¹²⁹ Cfr. Max SCHELER, *El puesto del hombre en el cosmos*, Buenos Aires, Losada, 1938 (1976), cap. VI.

especie humana, la cual lo extiende al resto del universo, consiguiendo como resultado la divinización del universo, o la emergencia de la condición divina del universo. La cuestión que siempre se plantea a los panteísmos, cuestión que no saben responder, es cómo a una realidad imperfecta se la puede llamar Dios; y cómo una realidad imperfecta, el universo en sus primeros momentos, puede llegar a convertirse por su propio dinamismo y autosuficiencia en algo perfecto, digno de ser considerado Dios.

En definitiva, en medio de propuestas interesantes sobre la ingeniería informática, y sus aplicaciones al servicio de los humanos, RK se adscribe a una concepción de la IA en sentido fuerte, con una noción del ser humano, así como del universo y de lo divino, que resultan poco convincentes. Y si nos referimos sobre todo a su concepción de lo humano, entendemos que, cegado y encandilado por la lógica de la IA, donde se sitúan sus investigaciones profesionales, creemos que no acierta a situar en sus justos términos la específica condición de lo humano, que no se halla constituida tanto por sus dotes intelectuales e instrumentales, sino por el diálogo humanizador entre la naturaleza, el individuo y la sociedad, como ya lo vio Horkheimer¹³⁰; es decir, entre los elementos biológico-genéticos de la especie humana, la subjetividad irreplicable de cada uno y el ámbito interpersonal y social, que nos hace despertar a nuestro yo, ir madurándolo y orientándolo hacia nuestra propia realización. Por ello, esta específica ontología humana resulta tan difícil, por no decir imposible, de reproducir artificialmente, a no ser que se haga a través de la reproducción biológica, dentro de un contexto social, tal y como nos reproducimos los seres humanos.

Carlos Beorlegui
Universidad de Deusto, Facultad de CC. Sociales y Humanas
Avda. de las Universidades, s/n, Apdo. 1
48080 Bilbao
carlos.beorlegui@deusto.es

¹³⁰ Cfr. Max HORKHEIMER, "Observaciones sobre la Antropología filosófica" (1935), en *Teoría Crítica*, Buenos Aires, Amorrortu, 1974, pp. 50-75.