

EL TRANSHUMANISMO DE RAY KURZWEIL. ¿ES LA ONTOLOGÍA BIOLÓGICA REDUCTIBLE A COMPUTACIÓN?*

JAVIER MONSERRAT
Universidad Autónoma de Madrid

RESUMEN: Los programas de computación, ante todo la ingeniería de la visión artificial y la programación de los sensores somáticos, ya han permitido, y lo harán con mayor perfección en el futuro, construir con alta perfección androides o cyborgs que colaborarán con el hombre y abrirán sin duda nuevas reflexiones morales sobre como respetar en su dignidad ontológica las nuevas máquinas humanoides. Además, tanto los hombres actuales como los nuevos androides estarán en conexión con inmensas redes de computación externa que harán crecer de forma casi increíble la eficacia en el dominio del propio cuerpo y de la naturaleza. Sin embargo, nuestro conocimiento científico actual, por una parte, del hardware y del software que sostendrá tanto las máquinas humanoides como las redes de computación externa hechas con la ingeniería existente (y también la previsible a medio e incluso largo plazo) y, por otra, nuestro conocimiento científico sobre el comportamiento animal y humano desde las estructuras biológico-neuronales que producen un sistema psíquico, nos permiten establecer que no existe fundamento científico que permita hablar de una identidad ontológica entre las máquinas computacionales y el hombre. En consecuencia, diversas ontologías producirán sistemas funcionales diversos. Podrá haber simulación, pero nunca identidad ontológica. Estas ideas son esenciales para valorar el pensamiento transhumanista de Ray Kurzweil.

PALABRAS CLAVE: Kurzweil, transhumanismo, androides, cyborgs, ontología biológica, ontología computacional, conciencia, redes externas de computación.

The transhumanism of Ray Kurzweil. Is biological ontology reducible to computation?

ABSTRACT: Computer programs, primarily engineering machine vision and programming of somatic sensors, have already allowed, and they will do it more perfectly in the future, to build high perfection androids or cyborgs. They will collaborate with man and open new moral reflections to respect the ontological dignity in the new humanoid machines. In addition, both men and new androids will be in connection with huge external computer networks that will grow up to almost incredible levels the efficiency in the domain of body and nature. However, our current scientific knowledge, on the one hand, about hardware and software that will support both the humanoid machines and external computer networks, made with existing engineering (and also the foreseeable medium and even long term engineering) and, on the other hand, our scientific knowledge about animal and human behavior from neural-biological structures that produce a psychic system, allow us to establish that there is no scientific basis to talk about an ontological identity between the computational machines and man. Accordingly, different ontologies (computational machines and biological entities) will produce various different functional systems. There may be simulation, but never ontological identity. These ideas are essential to assess the transhumanism of Ray Kurzweil.

KEY WORDS: Kurzweil, transhumanism, androids, cyborgs, biological ontology, computational ontology, consciousness, external computer networks.

* De acuerdo con la política editorial vigente en la Revista Pensamiento se publican exclusivamente artículos originales. No obstante, la Dirección de la Revista autoriza aquí la publicación de este artículo que es una readaptación de uno similar publicado en el volumen XXXI de la colección ASINJA, editado en Publicaciones de la Universidad Comillas. La Dirección autoriza esta readaptación porque se trata de un texto perteneciente a la UPComillas, que aquí adquirirá más divulgación, y porque su temática completa el hilo lógico conductor de los artículos de la sección [Tecnología y Naturaleza] de este volumen VII de la serie Ciencia, Filosofía y Religión.

Ray Kurzweil, nacido en Nueva York en 1948, de familia judía emigrada de Europa por la persecución nazi, es una personalidad relevante en la ingeniería, en la filosofía, en la futurología, pero sobre todo en la teoría computacional de los seres vivos y del hombre. Kurzweil ha sido, y es, un extraordinario profesional en la ingeniería informática. Fundó compañías que diseñaron importantes programas informáticos que se aplicaron, y se siguen todavía hoy aplicando, en tecnologías de reconocimiento de voz, de textos escritos o de imágenes (aplicaciones tan conocidas como *Omnipage* o *Siri*, usado en *Iphon* o en *Ipad* de Appel, van unidas al nombre de Kurzweil). Después de haber fundado, administrado, vendido, importantes compañías de ingeniería informática, después de haber pasado de unos sitios a otros, en la actualidad es Jefe de Ingenieros de *Google*. Ha sido nombrado doctor honoris causa en más de veinte universidades y ha recibido los máximos honores del gobierno americano y del mundo empresarial privado, incluyendo los tres premios concedidos por tres diferentes presidentes americanos. Ha publicado siete libros (cuatro que tengan en realidad relación con el pensamiento científico-filosófico, ya que otros son sobre la salud y el bienestar), pero son innumerables sus intervenciones públicas en conferencias y debates de todo tipo. Kurzweil ha sido uno de los promotores principales de la fundación en 2009 de la *Singularity University* en *Moffett Field CA*, apoyada por *Google* y la *Nasa Ames Research Center*, entre otras empresas y particulares señalados (incluyendo a Arnold Schwarzeneger). Eludimos exposiciones más amplias de su biografía personal, intelectual, y de su actividad profesional, ya que son fácilmente asequibles por Internet¹.

Pero, ¿en qué consiste entonces el pensamiento de Kurzweil? Necesitamos, en efecto, una cierta visión global de lo que, a nuestro entender, constituye su obra científico-filosófico-futurista, de tal manera que podamos encuadrar en ella el orden y el sentido de nuestro comentario. A) En primer lugar, Kurzweil aporta una teoría computacional del hombre y de los seres vivos. Esta teoría se funda en una teoría sobre el ordenador como producto de ingeniería humana y en una teoría de los seres vivos, de su sistema neuronal, y en especial del neocortex humano, así como en una teoría tendente a mostrar la identidad funcional entre los computadores y la mente animal y humana. B) De acuerdo con esta idea de la IA (inteligencia artificial) y de la mente animal/humana Kurzweil entra en el terreno de la futurología anunciando que el desarrollo tecnológico, al que atribuye un crecimiento exponencial por la *Ley de los Rendimientos Acelerados*, llevará a la humanidad a entrar en una *Singularidad* futura, o momento en que la historia se introducirá en una fase nueva, esencial y cualitativamente distinta a lo anterior. En qué consiste la teoría computacional/humana de Kurzweil, así como las razones que apoyan el anuncio de la *Singularidad* y de su naturaleza, deben ser precisamente expuestas y valoradas por nuestra parte.

¹ Los libros fundamentales de Kurzweil a que hacemos referencia en este escrito son estos. En 1989 publicó *The Age of Intelligent Machines*. En 1999 apareció *The Age of Spiritual Machines*, que completó en 2002 con una respuesta a sus críticos: *Are we Spiritual Machines? Ray Kurzweil versus the Critics of Strong IA*. En 2005 aparece su obra más importante: *The Singularity is Near: When Humans Transcend Biology*. Por último, en 2012 se publica su tercera obra antropológico-filosófica: *How to Create a Mind: The Secret of Human Thought revealed*. Además quizá deba añadirse: *Virtually Human: The Promise—And the Peril—Of Digital Immortality*, escrito por Martine Rothblatt y Ray Kurzweil (aparecido el 9 septiembre 2014). Los otros libros que pueden verse en el mercado tratan sobre el envejecimiento y la salud, no teniendo por tanto especial importancia desde el punto de vista científico, antropológico y filosófico. La obras de 2005 y 2012 han sido traducidas y publicadas en español por LolaBooks, Berlin. La presentación de *Cómo crear una mente. El secreto del pensamiento humano*, se hizo en la Universidad Comillas en 2013, con la presencia de José Luis Cordeiro (de la Singularity University), de Adolfo Castilla y del Editor en español.

Abordaremos aquí el estudio y discusión de la teoría computacional, aplicada a las máquinas y a los humanos, que constituye el fundamento del pensamiento de Kurzweil. Primero hablaremos de su «humanismo extensivo» y de su forma de entender la «conciencia». Por último, de forma conclusiva, apuntaremos algunos perfiles referentes a su teoría sobre el funcionamiento de la mente como sistema neurológico de reconocimiento de patrones y a su idea de la *Singularidad*, como pieza crucial de su futurología.

No queremos comenzar, sin embargo, sin declarar nuestra admiración y respeto por la obra de Kurzweil. Nos sumamos al reconocimiento que ya ha sido hecho por la sociedad misma. Pero además, y esto es más especial, queremos también declarar nuestra simpatía por Kurzweil, simpatía para con su carácter y su personalidad, muy americana, tal como traslucen en su obra y en su actuación pública. Pero simpatía también con sus ideas, aunque en muchos aspectos debamos también disentir, y no en cuestiones triviales, tal como expondremos. No obstante, intuimos que Kurzweil no es un transgresor atolondrado (pensemos en su discutida predicción de la inmortalidad derivada del dominio tecnológico futuro), sino un pensador que cree en valores, que es valiente para meterse en la futurología, que se deja llevar por la lógica de los hechos que se imponen, según su criterio, por la evolución tecnológica, y que trata —dudamos de que siempre lo consiga— de hacerlos conciliables con los valores tradicionales del humanismo.

1. EL «HUMANISMO EXTENSIVO» DE RAY KURZWEIL

Kurzweil es agnóstico en cuanto a las creencias metafísicas. Pero es respetuoso para con las creencias religiosas. Ser agnóstico es una forma de reconocer el misterio del más allá metafísico. Si se tuviera certeza de la existencia de Dios o de su no existencia, o de cualquier otra naturaleza que debiera atribuirse al más allá, no se estaría en el agnosticismo. Se estaría en el teísmo dogmático, en el ateísmo dogmático, o en cualquier otro tipo de dogmatismo. Si Kurzweil se declara agnóstico es porque considera que la argumentación teísta acerca de un más allá entendido como Dios no sería descartable. Tampoco sería descartable el ateísmo. Su posición moderada es, pues, muy diferente del ateísmo radical excluyente, dogmático, y explícito de autores como Richard Dawkins, Daniel Dennett, San Harris o Christopher Hitchens. Esto es evidente. Kurzweil no es un ateo porque está abierto al enigma de un más allá que se desconoce. Por esto es agnóstico.

Aunque de religión judía familiar, Ray Kurzweil no tuvo una formación judía fundamentalista, sino extraordinariamente abierta. «Mi formación religiosa se produjo en una iglesia unitaria en la que estudiábamos todas las religiones del mundo. Nos pasábamos seis meses con, pongamos el caso, el budismo. Íbamos a “misas” (las comillas son nuestras) budistas, leíamos sus libros y formábamos grupos de discusión junto con sus líderes. Después pasábamos a otra religión, como por ejemplo el judaísmo. El tema primordial era “muchos caminos hacia la verdad”, además de la tolerancia y de la transcendencia. Esta última idea significaba que el resolver las aparentes contradicciones entre las diferentes tradiciones no implicaba decidir que una era correcta y otra equivocada. Así, la verdad sólo puede ser descubierta encontrando una explicación que supere (trascienda) las aparentes diferencias, especialmente en lo que respecta a cuestiones fundamentales de significado y propósito»². He citado este acertado texto de Kurzweil porque muestra su identificación con la convergencia y no por la exclusión. No se trata

² KURZWEIL, R., *Cómo crear una mente* (CCM), edición española, LolaBooks 2013, página 211.

de rechazar nada, sino más bien de buscar la convergencia. Este mismo talante es el que parece tener al estudiar la naturaleza del ordenador, de la mente humana, y de la relación entre ordenador/mente.

1.1. «Humanismo extensivo» frente a reduccionismo

Por contraste, en orden a explicar lo que, a nuestro entender, es el «talante» de la obra de Kurzweil, pensemos en el «reduccionismo». Para éste lo que, de acuerdo con nuestra experiencia fenomenológica, personal y social, es más importante (conciencia, sensibilidad, pensamiento, lenguaje, motivaciones, sujeto psíquico, mente racional, decisiones libres y responsables...) queda «reducido» a los sistemas de interacción deterministas y mecánicos del mundo mecanoclásico. En último término, los seres vivos y el hombre quedaban «reducidos» a la condición de puros robots; muy complejos, ciertamente, pero robots. La mayor parte de las teorías computacionales del hombre son en la actualidad «reduccionistas» y, a lo más, reducen los fenómenos psíquicos a puros «epifenómenos» (un producto residual, marginal, de la evolución que no es un factor causal que influya en los actos humanos). Para el robotismo computacional y para el determinismo neural nuestra vida es sólo una consecuencia determinista y mecánica de cadenas causa-efecto ciegas que se explican por las interacciones físicas de la mecánica clásica. Pero el sujeto psíquico, apoyado en los procesos psíquicos que nos dan la apariencia falsa de ser protagonistas de la vida, no tiene en realidad ninguna «causalidad descendente controladora de lo físico-químico». El sujeto es un puro testigo epifenoménico (falsamente sentido como protagonista) de los productos de un mundo absolutamente determinista³.

Sin embargo, este no es el punto de vista de Kurzweil. No se identifica con el «reduccionismo» de tantos colegas. ¿Dónde se sitúa entonces su pensamiento? En principio digamos que quiere salvar al hombre, según la experiencia que este tiene de sí mismo, según la experiencia y convicciones «humanistas» que están en la base de la vida social, religiosa, filosófica, jurídica, económica y política. Kurzweil es un hombre con «sentido común» y sabe que no puede ofrecer una idea del hombre «reduccionista» que, en el fondo, es incompatible con nuestra vida personal y social, haciéndola imposible. Por ello, Kurzweil se esfuerza por mantener una idea «humanista» del hombre. Para él es claro que el hombre tiene conciencia, identidad personal, es libre y responsable de sus actos (hasta un cierto punto), construyendo su vida mediante una serie de decisiones que ponen en juego el uso de un psiquismo emocional y racional. Esto quiere decir que para Kurzweil, tal como yo lo entiendo, el mundo psíquico no es epifenoménico (ni reduccionista). La mente humana tiene una *causalidad real* descendente controladora de lo físico químico.

³ La idea computacional del hombre (y de los seres vivos) puede responder a la metáfora fuerte del ordenador o a la metáfora débil. La metáfora fuerte entiende que ordenador y ser vivo son ontológica y funcionalmente lo mismo, aunque evidentemente se reconozca que el ordenador no tiene un hardware biológico. No obstante, la biología animal y humana, el sistema neurológico es una implementación evolutiva de un puro sistema computacional. Este puede entenderse de dos maneras, serial o PDP conexionista. Las consecuencias de esta manera de pensar llevan a un reduccionismo robótico. El reduccionismo serial clásico comenzó en los años sesenta con la obra de Herbert Simon y Allen Newell. El conexionismo PDP con Rosenblatt inicialmente para acabar despegando con la obra de McClelland y Rumelhart en los años ochenta. La visión computacional del hombre puede verse en autores como Jakendoff, Kosslyn o Dennett, así como en el gran diccionario del MIT. La versión neurológica del determinismo causal se halla en el llamado determinismo neural que en España ha sido defendido, entre otros, por J. Rubia.

Por tanto, Kurzweil no pretende negar aquellas propiedades psíquicas del hombre en que se funda el «humanismo» en nuestra vida personal y social. En su perspectiva no se trata de negar sino de «extender». La cuestión no es negar, o «reducir», la mente (*mind*), sino de afirmarla y «extenderla» (*extended mind*). Para él, la tecnología moderna, fundada en los sistemas de computación, en el desarrollo de su *hardware* y de los programas de *software* que puedan ser implementados, está haciendo nacer nuevas formas de realidad que supondrán una extensión de las propiedades de la mente: la conciencia, la identidad, la condición de sujeto libre y responsable de sus acciones. Lo humano, pues, se extenderá más allá de lo que hasta ahora era «humano», introduciéndose en nuevos ámbitos y dimensiones de la realidad⁴.

1.2. *Las dos dimensiones principales del «humanismo extensivo» (extended mind)*

- 1) *La creación de nuevas entidades cuasi-humanas.* A no mucho tardar, los complejos programas de ordenador encaminados a simular la vida psíquica, y a producir un comportamiento en ella fundado, se perfeccionarán hasta un punto casi inverosímil. Se construirán robots cuasi-humanos que actuaran en el medio en función de una vida psíquica simulada que incluirá percibir el medio, el razonamiento y los planes de conducta, las emociones, la elección en un abanico de posibilidades abiertas, la autoprogramación, el diálogo y la interacción con los humanos —casi como si se tratara de un humano más—, y la configuración incluso de una biografía fruto de la personalidad propia del robot. Este tipo de «humanoides», cyborgs o androides, nos son hoy familiares por muchas obras de ciencia-ficción, divulgadas en literatura, cine y TV, como por ejemplo *Star Trek*.

Kurzweil nos plantea la cuestión de fondo que esto supone. «Imagínese que usted se encuentra con una entidad futura (un robot o un avatar) que es absolutamente convincente en lo que se refiere a sus reacciones emocionales. Se ríe convincentemente de sus chistes y además le hace reír y llorar (y no solamente pellizcándole). Le convence de su sinceridad al hablar de sus miedos y anhelos. Parece consciente en todas sus formas. De hecho, tiene el aspecto de una persona. ¿Aceptaría usted a esta entidad como si fuera una persona consciente?»⁵. «Ciertamente, existen desacuerdos sobre si alguna vez nos encontraremos con una entidad así. La predicción que yo mantengo es que esto ocurrirá en 2029 y se convertirá en algo habitual en la década de 2030. Pero dejando a un lado el marco temporal, creo que eventualmente acabaremos por considerar a estas entidades como conscientes. Considere cómo las tratamos ya cuando nos las encontramos bajo la forma de personajes de historias y películas. R2D2 de las películas de *Stars War*, David y Teddy de la película *A.I.*, Data de la serie de televisión *Star Trek: The next Generation...*. «Empatizamos con estos personajes pese a que sabemos que no son biológicos. Los consideramos personas conscientes, tal y como hacemos con los personajes biológicos humanos. Compartimos sus sentimientos y tememos por ellos cuando se encuentran en

⁴ Creemos que Kurzweil, en el fondo, entiende el sistema neural de acuerdo con la metáfora fuerte del ordenador. ¿Por qué? Porque para Kurzweil el cerebro funciona, aunque de forma biológica, según los mismos algoritmos de análisis de patrones propios del funcionamiento de la máquina. Sin embargo, Kurzweil ofrece una valoración moderada de su metáfora fuerte, de acuerdo con los principios del humanismo extensivo que seguidamente exponemos.

⁵ Cfr. CCM, o.c., p. 200.

problemas. Si es así como a día de hoy tratamos a los personajes de ficción no biológicos es que así es como en la vida real trataremos a inteligencias futuras que no posean un sustrato biológico»⁶ (CCM, pág. 201).

- 2) *Extensión de la mente humana a sistemas de computación externos.* Ya en la actualidad tenemos experiencia de lo que esto significa: a través del ordenador y de las redes estamos abiertos «extensivamente» a un mundo de información, de percepciones, de conocimientos, de imágenes... que potencian la capacidad de acción, de adaptación al mundo y de supervivencia. Hoy estamos ya conectados inalámbricamente a estas redes (como fue predicho hace años por Kurzweil) y es posible que en el futuro no necesitemos teclados, u otros «medios» que entonces parecerán «primitivos» para comunicarnos con las redes, que serán algo así como una extensión de nuestros cerebros. Podremos comunicarnos por el lenguaje (vg. Siri) o por el mismo pensamiento (como hoy se investiga). Los robots conducirán nuestros coches, trenes o aviones, realizarán las tareas industriales o del hogar, incluso gran parte de nuestra actividad intelectual. Conoceremos nuestro ADN con un poco de saliva en cualquier farmacia y un análisis médico exhaustivo. La nanotecnología hará posible que multitud de nanorobots biológicos vigilen nuestro organismo desde dentro, conectados a un sistema médico externo dirigido por un robot, o médico-androide, que podría tener incluso su propia personalidad, siendo capaz también de aprendizaje y de retro-programación. En fin, la supervivencia por el conocimiento y por la toma de decisiones habrá dejado ya de depender exclusivamente de nuestros cerebros biológicos, ya que habrán aparecido multitud de mentes externas que nos ayudarán en interacción con nosotros, habiéndose «extendido» además nuestra misma mente a inmensas plataformas externas de computación, de virtualidades multiformes para el apoyo individual y social.

Conclusión: talante del «humanismo extensivo» de Kurzweil. El hecho que se impone por la dinámica misma del conocimiento y de la tecnología, que no podrán frenarse, es la aparición futura de un estado, individual y social, de *Singularidad*, en el que la forma de vivir dará un salto cualitativo y se entrará en una nueva Era (la quinta y sexta Era, según la clasificación de Kurzweil en SEC, cap. Uno). Esta nueva Era no supondrá para Kurzweil «reducir» (reduccionismo) la mente humana a lo robótico, mecánico, determinista y ciego, sino, muy al contrario, la extensión de la mente humana hacia posibilidades nuevas y en sumo grado enriquecedoras. Las propiedades de la mente humana, tales como la conciencia, la condición de sujeto psíquico, la identidad personal, la libertad, el conocimiento, la racionalidad, la construcción biográfica y la especificidad de la propia personalidad, no desaparecerán. Al contrario, aparecerán nuevas formas de realidad que harán posibles nuevas formas de conciencia, identidad, libertad, personalidad... Se hará posible un nuevo «humanismo extensivo»⁷.

⁶ Cfr. CCM, o.c., p. 201.

⁷ El humanismo extensivo de Kurzweil, a nuestro entender, aunque considere el cerebro según la metáfora fuerte, no es reduccionista, ya que mantiene todas las propiedades existenciales que confieren al hombre su dignidad: la conciencia, la condición de persona, la identidad, el libre albedrío... Pero las mantiene en un plano similar al de una máquina. La nueva idea del hombre y de la máquina coincidirán en una nueva dimensión en que la conciencia humana y la conciencia de la máquina será similares. Kurzweil, por ello, no niega el humanismo, sino que lo reinterpreta y lo extiende al mundo de las máquinas «espirituales».

3. LA CONCIENCIA EN RAY KURZWEIL

No parece caber duda de que Kurzweil sea una personalidad optimista. Tiene experiencia profesional de los logros producidos en poco tiempo y su confianza en la *Ley de los Rendimiento Acelerados* le induce a suponer que los logros futuros van a crecer en una aceleración exponencial. Vislumbra el panorama futuro que supondrá el enriquecimiento de un *humanismo extensivo* y toma la actitud *voluntarista* de creer que la «imagen vislumbra» se hará realidad. El pensamiento de Kurzweil es, pues, un voluntarismo, es obra de un «creyente». Existen sin duda razones objetivas, científicas y filosóficas, para admitir que lo afirmado en su voluntarismo se hará realidad. No es un voluntarismo arbitrario, gratuito, sino argumentado científicamente y filosóficamente. Pero, en el fondo, su prognosis del futuro no es segura, y aceptar ese futuro supone lo que el mismo Kurzweil llama en diversos lugares un «acto de fe».

En el fondo, la belleza de ese futuro, en el que Kurzweil cree, depende de que pueda hacerse realidad el *humanismo extensivo*. Por tanto, no se trata de negar la experiencia humanista del hombre que constituye nuestro ser personal y social, sino de extenderla. Por tanto, si nuestra realidad humana como entidad biológica se funda en propiedades tales como conciencia, identidad personal, libertad, racionalidad, etc., la extensión del humanismo deberá asumir una «extensión» de esas propiedades. Pero aquí es donde comienza el problema para Kurzweil.

3.1. Una estrategia para la extensión de las propiedades de la mente

- 1) Si nos referimos —primera dimensión del *humanismo extensivo*— a la creación de *nuevas entidades cuasi-humanas* se presenta el problema de si a esas entidades cabe atribuirles la extensión de propiedades. La entidad biológica del hombre tiene «conciencia». Ahora bien, ¿cabe atribuir conciencia a un robot, a un avatar, cyborg o androide, cuyos programas simulan casi perfectamente la conciencia humana? El hombre es una entidad personal que se mantiene en el tiempo y que constituye una biografía específica. ¿Tiene sentido decir que un robot sea también una entidad personal? El hombre, y en su nivel, los animales gozan de grados de libertad. ¿Puede decirse que un robot humanoide sea también una entidad libre que construye su historia a través de un conjunto de elecciones libres derivadas de su personalidad? El problema de Kurzweil es que hay muchos científicos y filósofos que niegan que esas propiedades puedan aplicarse legítimamente a los robots. Es la cuestión de si será posible crear un nuevo tipo de entidades que supongan la extensión de las propiedades de la entidad biológica humana. Es el problema esencial de *How to Create a Mind*, el libro de 2012⁸.
- 2) Si nos referimos —segunda dimensión del *humanismo extensivo*— a la *extensión de la mente humana por redes externas* de percepción e información, de conocimiento, de ingeniería y mecánica, de biología y de nano-tecnología médica, ¿será todo eso compatible con las propiedades «humanistas» de la condición humana? ¿Se perderá la conciencia como propiedad individual de ciertas entidades biológicas como el hombre? ¿Será posible la libertad y el proceso libre de creación de una personalidad específica a través de una historia biográfica

⁸ Cfr. *How to Create a Mind*, 2013 en inglés. *Cómo crear una mente*, en español, 2013.

individual? ¿Seguirá siendo el hombre independiente para mantener su identidad personal a través del tiempo sin instancias deterministas externas que la limiten? El establecimiento de la *Singularidad*, ¿acabará arrastrando la existencia de la dignidad humana personal tal como hasta ahora ha sido conocida? Este es el gran problema que para el hombre individual supone la llegada de la humanidad al momento de una futura *Singularidad* que supondría vivir como hombre en un nuevo marco de condiciones objetivas para la vida. La *Singularidad*, por tanto, ¿permitirá la continuidad de la dignidad individual del hombre? Esta es la gran cuestión tratada por Kurzweil en *The Singularity is Near*⁹.

Como decíamos, abordamos más extensamente aquí sólo la cuestión de la conciencia, conectada con el humanismo extensivo, a saber, la creación de nuevas entidades cuasi-humanas (*How to Create a Mind*, 2012). Al concluir este artículo haremos una alusión breve a la cuestión del análisis de patrones y a la noción de Singularidad (*The Singularity is Near*, 2005).

La estrategia de Kurzweil para defender el *humanismo extensivo* es obvia. Por una parte, a nuestro entender, es claro, como decíamos, que no es ningún iconoclasta atolondrado que pretenda reducirnos a robots, negando todo aquello que da sentido individual y social a la vida. Por otra parte, es también claro que se ha dejado atrapar por la belleza de una nueva sociedad en que lo maravilloso del hombre pueda ser extendido a nuevas entidades emergentes de sorprendentes posibilidades tecnológicas. Kurzweil está atraído por la hermosura estética de la aparición de un mundo transhumano en que máquinas y hombres formen una dimensión unitaria y nueva. Ahora bien, ¿cómo armonizar estos dos valores? ¿Cómo extender lo humano a las máquinas, sin deteriorar lo humano de las entidades biológicas ya existentes? A mi entender, la estrategia de Kurzweil consiste, usando algunas expresiones del lenguaje español, en «marear la perdiz», o escaparse del problema echando la «tinta del calamar». Kurzweil argumenta que la idea de propiedades humanas como conciencia, identidad, libre albedrío..., lejos de estar definidas y ser evidentes, han sido objeto de interpretaciones muy diferentes y a ellas han sido atribuidos significados polisémicos. Por tanto, en este estado de confusión, sin contornos precisos, mareados como la perdiz o enmascarados por la oscuridad de la tinta del calamar, ¿quién puede negar la legitimidad de atribuir a las máquinas la conciencia, la identidad, la personalidad o la libertad, en el mismo sentido con que se atribuyen al hombre?

2.1. *El hecho de la conciencia*

Kurzweil sabe perfectamente que el hombre real y la sociedad están contruidos a partir de un hecho radical incuestionable: la *conciencia*. Los individuos se hacen a sí mismos personalmente, y la sociedad es lo que es, a partir de una propiedad de la mente que hace posible vivir una biografía humana responsable (existe un sujeto de atribución) que funda el orden social: es la conciencia. Por consiguiente, si se trata de argumentar que en la *Singularidad* se producirá una extensión del humanismo a las máquinas (*spiritual machines*), difícilmente será esto posible si a las máquinas no se les confiere *conciencia*. No sólo conciencia, sino también las otras propiedades de la mente constituidas en una entidad biológica como la nuestra. Es evidente que los cyborgs no son «biología» y esto lo sabe perfectamente Kurzweil.

⁹ *The Singularity is Near: When Humans Transcend Biology*, 2005. *La singularidad está cerca: cuando los humanos transcendamos la biología*, LolaBooks 2012.

Veamos un texto del mismo Kurzweil. «De manera que, en lo que respecta a la conciencia, ¿cuál es exactamente la cuestión una vez más? Es esta: ¿quién o qué es consciente? En el título de este libro (CCM, 2013) hago referencia a la «mente» (*mind*) en vez de al «cerebro» (*brain*) porque una mente es un cerebro consciente. También podríamos decir que una mente posee libre albedrío e identidad. La propia aseveración de que estas cuestiones son filosóficas no es por sí evidente. Así, yo mantengo que estas cuestiones nunca podrán ser completamente resueltas por la ciencia. En otras palabras, sin la aceptación de presuposiciones filosóficas, no existen experimentos falsables que podamos realizar para resolverlas. Si tuviéramos que construir un detector de conciencia, Searle querría estar seguro de que chorreara neurotransmisores biológicos. Por su parte, el filósofo norteamericano Daniel Dennett (nacido en 1942) se mostraría más flexible en cuanto al sustrato [es decir, el fondo ontológico, biológico o no-biológico], pero es posible que quisiera saber si el sistema contiene o no un modelo de sí mismo y de su propio rendimiento. Esta visión se acerca más a la mía, pero en el fondo sigue tratándose de una asunción filosófica. Regularmente se presentan propuestas que pretenden ser teorías científicas que enlazan la conciencia con algún atributo físico medible, a lo cual se refiere Searle como el mecanismo causante de la conciencia»¹⁰.

Por consiguiente, cuando Kurzweil escribe su obra *How to Create a Mind* tiene la intención explícita de explicar cómo se construye una «mente» que incluye la «conciencia». La mente es biológica, pero se puede construir una mente computacional. En otras palabras, Kurzweil pretende explicar cómo construir una «máquina» que sea «mente», es decir, incluyendo que tenga «conciencia». Dado el objetivo de Kurzweil (conferir a las máquinas «espirituales» aquellas propiedades que hagan posible un *humanismo extensivo*, es decir, hacerlas *spiritual machines*) es explicable que no le guste que algunos (como Searle, Hameroff o Penrose) hagan depender la conciencia de condiciones biológicas muy precisas, ya que entonces se podría hacer depender la atribución, o la no-atribución, de conciencia a la presencia de tales o cuales condiciones biológicas objetivas ya establecidas. En este caso, ya que las máquinas no tienen «ontología» o *harware* biológico, sería imposible detectar en ellas, o atribuirles, estas condiciones biológicas requeridas que justificaran hablar de conciencia. En otras palabras: no podríamos atribuir conciencia a las máquinas, en contra precisamente de lo que Kurzweil quiere conseguir en su humanismo extensivo. Por ello, es lógico que mencione a Dennett que es abierto en cuanto al sustrato ontológico y parece no exigir lo biológico, limitándose a postular que el sistema (consciente) debería tener al menos un modelo de sí mismo y de su propio rendimiento. Es claro que esto sí podríamos tratar de encontrarlo en un robot (aunque no en todos).

2.2. *El filosóficamente difuso concepto de conciencia*

Parece, pues, que a Kurzweil le interesa que la naturaleza de la conciencia, qué significa tener conciencia, cuál sea la ontología de la conciencia, es decir, quién o qué tiene conciencia, es decir, todo lo que atañe al concepto de conciencia no sea una cuestión científica precisa y bien definida, ante la que quepa tomar una posición igualmente

¹⁰ Cfr; CCM, o.c., pags. 196-197. Searle es uno de los autores que insisten en que la conciencia es una propiedad esencial del hombre y de los seres vivos que emerge de su ontología biológica. Por ello, se ve claramente que Kurzweil lo considera uno de los adversarios principales, de los que debe distanciarse. En cambio, Dennett considera que el sistema neural es un computador y, por ello, estaría más abierto a que su sistema formal (computacional) tuviera una coincidencia «formal» con el sistema formal de un cyborg (a esto apunta precisamente Kurzweil).

precisa. El concepto de conciencia es, al contrario, filosófico y, por ello mismo, difuso, impreciso, abierto a la especulación y a posiciones teóricas, en ocasiones incluso contradictorias, que permanecen en disputa desde hace siglos.

Nos dice Kurzweil que «el filósofo británico Colin McGinn (nacido en 1950) escribe que “discutir sobre la conciencia puede reducir incluso al pensador más meticuloso a un parloteo incoherente”. La razón de esto es que a menudo la gente tiene opiniones no contrastadas e inconsistentes sobre lo que el término conciencia significa exactamente»¹¹ (CCM, pág. 192).

Muchos analistas consideran que la conciencia es una forma de rendimiento, como por ejemplo, la capacidad de autoreflexión, que es la capacidad para comprender los pensamientos propios y explicarlos. Según esto, un bebé, un perro, ¿tendrían conciencia? Para otros, como Searle, la conciencia se reduce a mecanismos biológicos, pero esto llevaría, según Kurzweil, al reduccionismo como muchos critican. David Chalmers ha acuñado la expresión «el duro problema de la conciencia». La conciencia, ¿ nombra simplemente los atributos observables y medibles asociados al hecho de tener conciencia? Hay analistas que asocian la conciencia con la inteligencia emocional o inteligencia moral. Comprender cómo el cerebro realiza este tipo de tareas y lo que sucede en el interior del cerebro cuando las realizamos es lo que configura la «sencilla» cuestión de la conciencia planteada por Chalmers. Por el momento la «dura» cuestión de Chalmers es tan dura que entra dentro de lo inefable, comenta Kurzweil. Chalmers imagina un experimento mental: una reunión de humanos y zombis (cuyo comportamiento no contendría nada que pudiera distinguirlos de los humanos). Por ello, por definición, no habría manera de identificar a un zombi. Chalmers especula también con que la conciencia podría existir en una dimensión distinta del mundo físico. Este «dualismo» (mundo de la conciencia y mundo físico) no permitiría que la conciencia afectara causalmente al mundo físico o fuera afectada por él (epifenomenalismo). Otra especulación de Chalmers es el pamprotopsiquismo (todo lo físico es consciente). Un enchufe, un hombre, una hormiga, pero también una colonia de hormigas como sistema, deberían ser conscientes en mayor o menor grado. Otra manera de conceptualizar la idea de conciencia es como un sistema que posee «qualia» (o «experiencias conscientes» de todo tipo, un color, una imagen, un sentimiento, un pensamiento). Aquellas «cualidades» (qualia) que son susceptibles de ser percibidas por un sujeto, al ser percibidas, constituirían la conciencia. Kurzweil argumenta también que esta definición tampoco es fiable, crea confusiones que pueden aplicarse de forma generalizada.

Abundando en la idea de Searle de que la conciencia es un estado que deriva propiedades biológicas precisas, se fija Kurzweil en las propuestas de Hameroff y de Penrose, que trata sucesivamente, para acabar descartando su propuesta por presentar inconsistencia interna y por no poder aportar evidencias empíricas, de tal manera que quedaría reducida a un puro «acto de fe».

Al igual que hace con el hecho de la conciencia, presenta también Kurzweil los conceptos de libre albedrío e identidad. Son igualmente confusos y distan de constituir un concepto cerrado que sólo pudiera aplicarse a entidades exclusivamente biológicas. Sobre el libre albedrío discute y muestra las contradicciones entre los experimentos clásicos de Sperry sobre cerebro dividido, las aportaciones de Gazzaniga, del determinismo neural que parte de Libet, o las ideas de Ramachandan de que se debería hablar, más bien, de un *libre no-albedrío* (la mente sería libre para elegir entre un abanico de propuestas deterministas). Presenta también cómo se vería el libre albedrío en la propuesta

¹¹ Cfr. CCM, o.c., p. 192.

del autómatas universal del Dr. Wolfram Alpha, así como la discusión del libre albedrío que estaría presente en el robot Watson, capaz de derrotar a cualquier humano en el juego *Jeopardy* o sea, de preguntas y respuestas sobre tema universal¹² (sobre todo esto ver: CCM, págs.. 191-234).

La conclusión de Kurzweil. Es obvio que interesa que el concepto de conciencia sea algo difuso e impreciso, al que no podemos acercarnos «sin marear la perdiz» o sin meternos dentro de la «tinta del calamar». Es claro que sólo así se podrá hallar una vía para atribuir «conciencia» a los robots. Por tanto, nos dice Kurzweil, la pregunta sobre si una entidad es o no es consciente no es una cuestión científica. Basándose en esto, algunos analistas proceden a cuestionar si la propia conciencia tiene alguna base real. La filósofa y escritora inglesa Susan Blackmore (nacida en 1951) habla de la gran «ilusión de la conciencia». Reconoce la realidad del meme (idea) de la conciencia. En otras palabras, la conciencia existe como idea y existen muchísimas estructuras neocorticales que tienen que ver con esta idea, por no mencionar las palabras que han sido proferidas y escritas sobre ella. Sin embargo no está claro que haga referencia a algo real. Blackmore continúa diciendo que ella no está necesariamente negando la realidad de la conciencia, sino que se limita a exponer los tipos de dilemas con los que nos encontramos al intentar fijar el concepto. Tal y como escribió el psicólogo y escritor británico Stuart Sutherland (1927-1998) en el *International Dictionary of Psychology*, «la conciencia es un fenómeno fascinante, pero evasivo. Es imposible especificar lo que es, lo que hace o el por qué de su evolución»¹³. Ciertamente en este párrafo evidencia Kurzweil lo que significa «marear la perdiz» o «echar tinta de calamar».

3. LA CONCIENCIA EN LA EXPERIENCIA INDIVIDUAL Y SOCIAL, Y EN LAS CIENCIAS HUMANAS

No ponemos en duda que hay muchos autores que han «mareado la perdiz» en torno al hecho y al concepto de conciencia. Son autores provenientes en su mayor parte de la filosofía, de la filosofía analítica, de la teoría de la mente, del mecanicismo determinista clásico, de las modernas versiones computacionales de la mente, y de algunos otros. Kurzweil está interesado en echar «tinta de calamar» sobre el concepto de conciencia y, por ello, recurre a estos autores para avalar su idea que la «conciencia» es algo muy confuso. Tiene todo el derecho a unirse a este grupo minoritario de autores.

Sin embargo, debemos también poner de relieve que para la mayor parte de autores y profesionales en psicología, filosofía, neurología, medicina, biología evolutiva, etología, ciencias naturales y humanas en general, existe un concepto de conciencia general, más obvio, derivado directamente de la experiencia. Pero, sobre todo, hay que insistir en que ese concepto es intuitivamente el que todos tenemos personal y socialmente. Es la idea de que la conciencia es un hecho real, dado en el hombre y en animales superiores. Este hecho funda los procesos y la actividad de la mente, jugando además un papel causal descendente en la producción, control de las acciones, en el comportamiento, en la construcción de la propia biografía y en las acciones libres responsables que fundan el orden social. La idea fenomenológica de la conciencia que funda la ciencia, así como el orden personal y social, no es «epifenomenalista»; para la mayor parte de los científicos, neurólogos, psicólogos y para la gente en general, el hombre no es un robot.

¹² Sobre todo esto ver: CCM, o.c., págs.. 191-234.

¹³ Cfr. CCM, o.c., p. 202.

De la misma manera que Kurzweil, o cualquier otro pensador, tiene derecho a considerar que no sabemos de qué hablamos cuando hablamos de la conciencia, así igualmente la mayor parte del consenso social tiene derecho a mantener que la conciencia es un hecho definido cuyos perfiles fenomenológicos forman parte de la evidencia social. En nuestra perspectiva, la *naturaleza fenomenológica* de la conciencia es evidente como hecho (es decir, la forma en que aparece o se manifiesta en nosotros este hecho). Es un hecho que, además, debe ser descrito en sus características fenomenológicas (y no todos lo harán insistiendo en los mismos rasgos fenomenológicos). Además, junto a lo fenomenológico, debemos distinguir el problema de la *naturaleza ontológica* de la conciencia que sigue siendo una cuestión oscura tanto para la ciencia como para la filosofía. En todo caso, la cuestión sobre la *ontología de la conciencia* ha sido planteada en las modernas ciencias humanas en dependencia de la *fenomenología de la conciencia*. Es obvio que sea así, ya que la ciencia pretende siempre explicar los fenómenos. Se parte de un *explicandum* (el fenómeno que debe ser explicado, aquí la «conciencia») y se busca un *explicans* a medida del fenómeno que se quiere explicar (un conjunto de causas que permitan entender cómo y por qué el fenómeno en cuestión es real).

3.1. Una idea de la conciencia alternativa a Kurzweil

Defendemos, pues, que la forma en que Kurzweil presenta el «duro problema de la conciencia» (Chalmers) no responde a la forma en que ordinariamente se plantea hoy la cuestión en la ciencia, en la filosofía y, sobre todo, en la vida personal y social. Hay una experiencia fenomenológica de la conciencia que establece el marco en el que se accede a plantear la discusión del espinoso problema de la ontología de la conciencia. Hacemos seguidamente algunas observaciones que permitan distinguir con precisión la cuestión fenomenológica de la conciencia en relación con la cuestión de su explicación ontológica.

- 1) Lo que designamos con el término «conciencia» forma parte de la experiencia primordial que constituye el origen de nuestro mundo humano. Todo surge del hecho de que «advertimos» nuestra auto-experiencia psíquica. Lo que llamamos «conciencia» es una propiedad básica y fundamental, esencial, de nuestra psique. El término psique designa, pues, nuestro cuerpo con dos propiedades esenciales: tener «conciencia» (sentirse a sí mismo, ser una entidad *cum-scien-**cia*, con sensación o conocimiento de sí mismo, o sea, con conciencia) y ser un «sujeto» (una entidad que, a partir de la conciencia, es capaz de impulsar acciones para la adaptación óptima al medio). Esta experiencia primordial de conciencia se nos da de una forma integrada o unitaria (si se quiere «molar»): es la integración de la conciencia visual, auditiva, propioceptiva, gravitatoria, etc. La conciencia es el resultado de la integración unitaria, armónica y coordinada, de todos los sistemas sensitivo-perceptivos, internos y externos. Por ello, el sujeto psíquico («psíquico» porque impulsa y dirige la adaptación del medio a partir de la «conciencia») es el que, según nuestra auto-experiencia actúa en el mundo. Conciencia y sujeto constituyen la psique, pero la conciencia aislada (esto en realidad no sería posible) no es sujeto, no tiene por sí misma «sujetualidad», aunque sea una propiedad esencial de la psique (ya que, sin conciencia, no habría ni psique ni sujeto).
- 2) La experiencia primordial de todo hombre es, pues, un cuerpo consciente que constituye un sujeto activo. Ahora bien, a su vez, la experiencia del cuerpo es una experiencia psicobiofísica. El cuerpo es psico-bio-físico: «físico» porque

es parte del mundo físico objetivo (la conciencia está abierta al mundo por los sentidos externos); «bio» porque se siente un cuerpo que pertenece a la clase especial de los cuerpos físicos «vivientes», con vida; «psico» porque el cuerpo se siente a sí mismo, tiene «psique», o sea, «conciencia» y es un «sujeto psíquico» (que actúa a partir de la conciencia y desde ella se constituye).

[Experiencia primordial, *deícsis*, fenomenología]

- 3) La experiencia de la conciencia —así integrada en la unidad psicobiofísica de nuestro cuerpo— es una experiencia primordial que tiene un carácter deíctico: es decir, no podemos describirla o conceptualarla en sí misma en función de alguna otra experiencia más primordial, sino recurriendo a palabras que sólo apuntan, describen, señalan (*deíksis*), hacia la conciencia como única experiencia primordial. Así, al decir que sentimos, que tenemos *cum-sciencia*, conciencia o conocimiento de nosotros (o, por igual, que sentimos imágenes visuales, auditivas, propioceptivas, o tenemos emociones...) no hacemos sino introducir palabras que apuntan, señalan, a la experiencia primordial de sensación-percepción-conciencia. Si no tuviéramos esa experiencia primordial seríamos incapaces de entender qué es la conciencia. En otras palabras: a alguien que no tuviera conciencia sería imposible explicarle qué es la experiencia real de tener conciencia.
- 4) La experiencia primordial, deíctica, de la sensación-percepción-conciencia es individual e intransferible. El hombre sólo tiene experiencia de la conciencia en su experiencia psicobiofísica individual. Ahora bien, tras la intercomunicación humana primitiva en los grupos sociales (por signos), la aparición del lenguaje, la cultura y las diversas formas de conocimiento, la experiencia individual da lugar al establecimiento de los grandes ejes de consenso sobre la conciencia (y la entidad psíquica): se asume que lo que hombre individual advierte en sí mismo, y señala deícticamente, existe también en los demás hombres. Pronto aparecen en el lenguaje los términos que nombran la actividad psíquica, presente en todos los hombres, de tal manera que se hace entonces posible intercambiar los contenidos de esa actividad psíquica (sensaciones, percepciones, recuerdos, conocimientos, emociones...). La posesión de una psique y la actividad psíquica consecuente se atribuyen por extensión a otros hombres y a los animales, cada una dentro de sus propiedades específicas y diferenciales.
- 5) La auto-experiencia de la actividad psíquica, consensuada socialmente, nombrada hasta ahora en general, se especifica como experiencia del conjunto de los procesos psíquicos (de un cuerpo consciente dirigido por un sujeto): sensación, percepción, conciencia, memoria, subjetualidad psíquica, atención, conocimiento, pensamiento, lenguaje, emoción, etc. Ahora bien, se trata de funciones, procesos y estados psíquicos que se producen en la actividad psíquica del cuerpo, a saber, como producto propio de su unidad psicobiofísica. El término «mente» podría servir para designar el conjunto de funciones y procesos psíquicos que hace posible la psique animal y humana (la mente animal y la mente humana). En lo específico de la mente humana aparece la actividad racio-emocional específica del hombre. El sujeto humano advierte que dirige un cuerpo consciente por medio de la aplicación de los recursos de su mente. [Causalidad descendente de la sensibilidad-conciencia]
- 6) Hay un elemento importante para describir la auto-experiencia psíquica: la causalidad descendente con que la psique (a través de la actividad del sujeto psíquico) controla y dirige las acciones del cuerpo. Si este es una entidad

psico-bio-física, lo bio-físico tiene una acción causal sobre lo psíquico, pero, a su vez, lo psíquico (mediando los procesos psíquicos) tiene una acción causal sobre lo bio-físico. El hombre advierte que por medio de sus pensamientos, intenciones, conocimientos, planes de conducta, emociones, decide, controla y dirige la actividad bio-física del cuerpo. Es la experiencia de la interacción psico/bio-física. Todos los grupos humanos han vivido siempre en el consenso «obvio» de que la vida es producto de las acciones humanas. En las modernas sociedades este consenso ha crecido hasta llegar al concepto de persona y responsabilidad moral que funda el orden social, jurídico y político de todas las naciones.

- 7) Debe advertirse que la descripción de esta experiencia primordial «psíquica», que incluye la conciencia, tiene un carácter exclusivamente fenomenológico. Por «fenómeno» (de *fainein*, en griego, «aparecer», «manifestarse») se entiende la forma en que se manifiesta la realidad. La auto-experiencia psíquica, como conciencia y como sujeto, con sus funciones, sus procesos y sus estados, tal como hemos descrito, es la experiencia fenomenológica primordial de la realidad. Es, por una parte, la experiencia de la unidad psico-bio-física del cuerpo, y, por otra parte, el mundo objetivo, externo, del que el cuerpo forma parte. El mundo y el cuerpo sólo existen, o son reales, a través de la actividad psíquica primordial que los presenta en su experiencia fenomenológica.
- 8) Por consiguiente, teniendo de hecho el hombre una constitución psíquica racio-emocional, ha ejercido la razón para preguntarse de dónde viene, por qué, cuáles son las causas de la existencia del mundo físico, del mundo biológico y del mundo psíquico, es decir, del hecho de la unidad psico-bio-física del cuerpo en el mundo. La razón ha sido ejercida en el conocimiento ordinario, en la filosofía y en la ciencia, como muestra la historia. Así aparecen la física, la biología y la psicología (que se corresponden, en definitiva, con los tres mundos de Popper). [La construcción racional de una idea de la realidad]
- 9) La primera gran inferencia hecha por la razón, a partir de la experiencia de la psique, es que el cuerpo humano, en su unidad psicobiofísica, existe en un mundo real objetivo externo, que existe independientemente del hombre y de su psiquismo, en el que el hombre ha sido producido. El consenso social sobre la existencia del mundo externo es la base de todas las culturas, del conocimiento ordinario, de la filosofía y de la ciencia. El gran movimiento de la ciencia actual, por ejemplo, carece de sentido si le quitamos su intención de conocer el mundo y de dominarlo.
- 10) El hombre, pues, ha ejercido la razón para preguntarse cuáles son las causas que han producido la existencia física del mundo externo, que se infiere como independiente. Se pregunta también cómo y por qué ese mundo físico que está ahí ha producido los seres biológicos.
- 11) Se ha preguntado además cómo y por qué el mundo bio/físico ha producido la emergencia de los seres vivos con sensibilidad y con conciencia. En el marco de la ciencia moderna, la física (en la microfísica clásica y cuántica, y en la macrofísica o cosmología), la biología y la psicología, explican la existencia de la psique a partir de las propiedades primordiales de la materia que constituye el universo. La materia debe tener unas propiedades ontológicas tales (es decir, un modo de ser realidad) que hagan posible que en los organismos vivientes hayan emergido la sensibilidad y la conciencia. Sin atribuir a la materia la «capacidad ontológica de producir sensibilidad-conciencia» no sería posible una

explicación racional de cuanto el hombre, y el consenso social, advierten en su experiencia fenomenológica primordial.

- 12) Ahora bien, ¿por qué la materia del universo tiene una ontología «capaz de producir sensibilidad-conciencia» y, más bien, no la tiene? La materia podría no ser apta para producir conciencia. La ciencia no puede argumentar la necesidad de que la materia produjera la conciencia. La materia primordial podría no ser sensible, capaz de producir la conciencia superior. Simplemente debe admitirlo como un hecho porque lo impone la experiencia fenomenológica primordial de la psique de hecho existente.

[El factor sensibilidad-conciencia en la biología evolutiva y en la neurología]

- 13) Las ciencias humanas, por tanto, han tratado de explicar la emergencia de la sensibilidad, o sentisciencia primitiva, en los organismos vivientes. El universo fue puramente físico durante más de diez mil millones de años. Es probable que la vida comenzara como puro mecanicismo determinista. Pero ya en los seres vivos unicelulares, en algún momento evolutivo, debió de comenzar a producirse una «sentisciencia primitiva». Desde entonces la vida evolucionó mediante una coordinación balanceada entre determinación y sensibilidad. Los organismos pluricelulares especializaron un subsistema de células para constituir el sistema nervioso que organizaba la sensibilidad externa e interna en orden a una adaptación óptima al medio. Sin embargo, los seres vivos evolucionaron por medio de un equilibrio entre la determinación (mecanicismo, vg. el ADN) y la sensibilidad-conciencia, de acuerdo con el interaccionismo antes señalado. La información (estímulo) produjo reacciones adaptativas (respuesta). El hecho de que hubiera emergido la capacidad de sentir, ya en el mundo celular, constituyó un cambio evolutivo favorable para que los organismos vivientes consiguieran una mejor información, una mejor evaluación del estado del propio cuerpo y una mejor respuesta adaptativa al medio en que se debía supervivir. Poco a poco, por tanto, en paralelo y coordinadamente, el perfeccionamiento del sistema sensitivo de información y la perfección consecuente del sistema de respuestas, fue dando origen al nacimiento del sujeto psíquico.

- 14) En estados primitivos de sentisciencia no había, evidentemente, un sujeto que fuera capaz de «advertir el hecho de sentir». Pero la sentisciencia era algo real que tenía consecuencias favorables en la adaptación al medio. La neurología evolutiva explica cómo poco a poco los diversos sistemas sensitivo-perceptivos fueron coordinándose por medio de un sistema nervioso central. Esta integración sensitiva dio origen a la «conciencia» y, en consecuencia, a la aparición de un sujeto psíquico cada vez más complejo. Los organismos primitivos, como un anfibio, serían «autómatas sensitivos», aunque el anfibio estuviera produciendo ya la emergencia incipiente de un sujeto psíquico (que permanecería latente, sin causalidad descendente, como factor sólo epifenoménico). Esta forma de respuesta automática (determinista) sería lo que llamamos el instinto (o también «respuestas signitivas»). Pero en los animales superiores la hipercomplejidad de los sistemas estímulo/respuesta, es decir, de la información (señales) y de los programas de respuesta (adaptación consecuente al medio) habría ido forzando poco a poco la aparición de unas nuevas funciones del sujeto psíquico que supondrían la «ruptura de la signitividad de las respuestas». Frente a un puro determinismo en las respuestas, el animal superior comienza a deliberar sobre los sistemas de señales y sobre las posibles respuestas. Lo vemos en animales superiores. Esta ruptura de signitividad es la que desemboca evolutivamente

en la aparición de la razón en la especie humana. La neurología moderna ha descrito cómo las sensaciones se registran en el cerebro en engramas o patrones neurales, en módulos específicos (visual, auditivo...), de tal manera que entre ellos se establecen un conjunto de links que hacen posible la memoria, la combinatoria de imágenes por el *remember present* de Edelman, hasta el nacimiento de las representaciones en la mente animal (en el segundo artículo relacionaremos esto con la idea de reconocimiento de patrones de Kurzweil).

[El problema de la conciencia como problema de su ontología biofísica profunda]

- 15) La conciencia es, y ha sido, para la ciencia algo evidente y sin problema: el hecho de que existe una sensibilidad-conciencia, que de ella tenemos una directa experiencia primordial, déctica, fenomenológica, incuestionable, personal y socialmente, que ha aparecido en el proceso evolutivo, que existe en los seres biológicos en una coordinación armónica con factores mecánico-deterministas, que existe dentro de un interaccionismo causal biofísico-psíquico, y que es un factor causal descendente que influye en la conducta dirigiendo y controlando los mecanismos físico-químicos que permiten la acción y la adaptación al medio físico objetivo. En la ciencia lo normal es considerar que los seres vivos no son robots (aunque en ellos haya mucho de determinación mecánica); esto quiere decir que la sensibilidad-conciencia no tiene un puro papel epifenoménico, sino que influye en la acción con una causalidad descendente que controla los mecanismos físico-químicos. Ahora bien, la sensibilidad-conciencia tiene muchas manifestaciones: desde la sentisciencia de organismos unicelulares, a la sensibilidad de autómatas sensitivos (como los anfibios), a la conciencia de animales superiores. Pero si nos referimos a la capacidad de reflexión (o imagen racional de sí mismo), a la conciencia racional o moral, no podemos hablar sólo de conciencia, puesto que en la reflexión, la razón, la moral, juega un papel el sujeto psíquico (la psique armoniza la conciencia con la subjetualidad).

Esta manera de ver las cosas, admitiendo como evidencia real el hecho de la sensibilidad-conciencia y su función causal en la conducta, es común en la mayor parte de los científicos, en física, en biología, psicología, filosofía, pero sobre todo en la praxis profesional de médicos, neurólogos y psicólogos que, si consideraran que sus pacientes son robots, no sabrían cómo proceder. Este es el mismo consenso que existe en la vida real: en la manera de entender y ejercer los individuos su propia vida y en la forma en que se ha organizado la sociedad como marco de las relaciones de unos seres humanos con otros.

- 16) ¿Dónde está, por tanto, el problema de la conciencia? El problema no está en reconocer que la conciencia (mejor, la psique que incluye conciencia y sujeto) exista, sea real e influya como protagonista en la conducta de los seres vivos. El problema consiste, más bien, en conocer las causas que producen la aparición de la sensibilidad-conciencia en el proceso evolutivo. Hay un problema irresoluble, ya mencionado: por qué la ontología de la materia del universo es susceptible de producir «conciencia». No cabe sino aceptar el hecho de que la materia es así, aunque pudiera ser de otra manera. Ahora bien, el problema sigue. Consiste en que la sensibilidad-conciencia (de acuerdo con las propiedades que podemos describir fenomenológicamente) tiene que haber sido producida desde el mundo físico y desde el mundo biológico. ¿Cómo concebir la imagen física y biológica del universo de tal manera que podamos explicar satisfactoriamente

el hecho real de la existencia de la sensibilidad-conciencia, según la experiencia fenomenológica que de él tenemos?

Este es el auténtico problema de la conciencia, tal como en efecto ha sido planteado en la ciencia. En el fondo es el problema del «soporte físico» del psiquismo animal y humano. Las teorías clásicas sobre la conciencia (que aquí voy sólo a nombrar) son, en el fondo, una discusión y propuesta sobre cómo entender el «soporte físico» del psiquismo a partir del mundo bio-físico: agnosticismo psicofísico interaccionista, dualismo interaccionista, fisicalismo, fisicalismo lógico y computacional, teoría de la identidad, también llamada identismo, epifenomenalismo, pampsiquismo y su matización en la forma de pamprotopsiquismo, funcionalismo, la forma computacional del funcionalismo, la teoría marxista de los saltos cualitativos (Luria), o, finalmente, las teorías que pretenden explicar el psiquismo a partir de las propiedades cuánticas de la materia (Hameroff y Penrose, rechazados por Kurzweil como antes indiqué).

La explicación del «soporte físico» de la conciencia ha sido un serio problema que dependía de la imagen del mundo físico en la ciencia. Cuando en esta era predominante, y casi exclusiva, una explicación solo mecanoclásica (mecánica clásica), mecánica y determinista, fue muy difícil y se cayó en el reduccionismo. Muchos llegaron a pensar que era imposible una explicación unitaria de la condición psicobiofísica del hombre y de los seres vivos (muchos recurrieron al dualismo). Sin embargo, nadie en la ciencia dudaba de que la conciencia era realmente existente y que se producía dentro de la unidad psicobiofísica que mostraba la experiencia fenomenológica.

17. CONCIENCIA BIOLÓGICA Y CONCIENCIA COMPUTACIONAL

No ponemos en duda, por tanto, como antes decíamos, que sobre el hecho de la conciencia es posible «marear la perdiz» y echar sobre ella «tinta del calamar». Kurzweil se hace eco de estas visiones desconcertadas ante el hecho de la conciencia porque cree que esta oscuridad le interesa en orden a su «humanismo extensivo». No obstante, el desconcierto de que Kurzweil se hace eco no es lo ordinario, ni en el mundo de la ciencia, de las ciencias humanas, ni mucho menos en el mundo de la experiencia ordinaria de la vida personal y social. Es lo que el mismo Kurzweil entiende y, por ello, sabe que su intento de unificar la conciencia biológica con la conciencia computacional está bailando en la cuerda floja. Es muy difícil negar que el hecho de la conciencia sea algo real que va unido esencialmente a la experiencia fenomenológica de nuestra realidad psico-bio-física.

17.1. *El mecanicismo «espiritual» de Kurzweil*

«La palabra “espiritual”, nos dice Kurzweil, suele usarse para hacer referencia a las cosas que poseen una significación fundamental o última. A mucha gente no le gusta utilizar la terminología procedente de las tradiciones espirituales o religiosas, ya que ello implica unos conjuntos de creencias con los que puede que no estén de acuerdo. Sin embargo, si dejamos a un lado las complejidades místicas de las tradiciones religiosas y nos limitamos a respetar “lo espiritual” como un término que implica algo cuyo significado para los humanos es muy profundo, entonces el concepto de conciencia sí que encaja, ya que refleja el valor espiritual último o fundamental. De hecho, la propia palabra “espíritu” suele ser usada para indicar conciencia».

«Así, la evolución puede ser vista como un proceso espiritual en el que se crean seres espirituales, es decir, entidades conscientes. Así mismo, la evolución se desplaza hacia una mayor complejidad, un mayor conocimiento, una mayor inteligencia, una mayor belleza, una mayor creatividad y hacia la capacidad de expresar emociones de mayor trascendencia como el amor. Todas estas son descripciones que la gente ha utilizado para referirse al concepto de dios, aunque en estos aspectos a dios se lo describe como carente de limitaciones».

«La gente suele sentirse amenazada por discusiones que implican la posibilidad de que una máquina pueda ser consciente, ya que este tipo de consideraciones son interpretadas por ellos como denigrantes con respecto al valor espiritual de las personas conscientes. Sin embargo, esta reacción refleja una mala comprensión del concepto de máquina. Tales críticas abordan la cuestión basándose en las máquinas que conocen a día de hoy, y aun con todo lo impresionantes que se están volviendo, estoy de acuerdo en que los ejemplos contemporáneos de la tecnología todavía no merecen que los respetemos como seres conscientes. Mi predicción es que se harán indistinguibles de los humanos biológicos, a los cuales sí que consideramos seres conscientes, y que por tanto se beneficiarán del valor espiritual que otorgamos a la conciencia. No se trata de una denigración de las personas, sino de un aumento de nuestra comprensión sobre ciertas máquinas futuras. Es probable que para estas entidades tuviéramos que adoptar una terminología diferente, ya que serán un tipo diferente de máquinas»¹⁴.

17.2. *Funciones, procesos y estados de un sistema: identidad ontológica e identidad funcional*

Podemos extraer ciertas conclusiones sobre las funciones, procesos y estados, de un sistema, así como de la identidad ontológica y funcional entre sistemas.

- 1) En principio, podemos imaginar la estructura de las funciones, procesos y estados de un sistema, prescindiendo de cómo pudiera estar construido en la realidad (es decir, de su ontología, bien fuera biológica o no lo fuera). Así, sería posible imaginar un sistema capaz de recibir información por diversas vías externas al sistema e internas al sistema mismo; capaz de analizar esa información, de registrarla, retenerla, recuperarla y de relacionarla entre sí; capaz de ponderar esa información en orden a ciertos valores relativos al óptimo mantenimiento y funcionamiento del sistema; capaz de actuar en el medio motrizmente en beneficio del propio sistema; capaz de aprender y capaz de construir formas de información perfeccionadas y seleccionadas; capaz de aprender y de construir sistemas de valores perfeccionados y seleccionados; capaz, mediante ese aprendizaje, de construir un sesgo de propiedades global en la captura de información, en su valoración y en el diseño de las formas de actuación motriz en el medio («personalidad»); capaz de un análisis reflexivo (sobre sí mismo) de la propia información, de las formas de aprendizaje, de los sistemas valorativos, de los procesos de respuesta al medio por la motricidad del sistema (programas); capaz de construir diversos programas o formas de información, valoración y respuesta, poseyendo formas de elección entre las diversas opciones posibles... Ponemos «puntos suspensivos» porque podríamos seguir la enumeración y matización de las propiedades y atributos de este «sistema

¹⁴ Cfr. CCM, o.c., p. 212.

- imaginario». Ahora bien, un sistema de estas características imaginarias, ¿cómo podría estar construido realmente?
- 2) Una primera respuesta es obvia y nos viene dada por nuestra experiencia de la entidad psico-bio-física que nos constituye y que podemos describir de una forma fenomenológica, tal como antes explicábamos. La ontología psico-bio-física de la realidad ha hecho posible la existencia de los seres vivos, y del hombre, como sistemas equilibrados entre determinación (interacciones mecano-clásicas) y sensibilidad-conciencia-sujeto. Estos sistemas de ontología psico-bio-física (más en breve, de ontología biológica) permiten, de hecho, construir seres vivos con conciencia, subjetualidad psíquica, conocimiento, valores, identidad, libre albedrío y las otras muchas propiedades que advertimos en nuestra vida psíquica. Es lo que hemos venido describiendo.
 - 3) Una segunda respuesta viene dada hoy al extraer las consecuencias de la ingeniería y lógica de la computación. Un sistema como el que imaginábamos en el punto uno podría ser hecho realidad a partir de la ingeniería y la lógica del ordenador. Hemos visto cómo Kurzweil predice que pronto tendremos máquinas robóticas que harán realidad los cyborgs o androides que hoy vemos ya en las películas de ciencia ficción. Podrán ser construidos de manera tal que no sólo tengan las propiedades del sistema imaginario antes descrito, sino que tengan propiedades similares a las de una implementación psicobiofísica de ese sistema: conciencia, identidad, personalidad, emociones, libre albedrío, etc. No ponemos en duda que pudiera ser posible en el futuro la construcción de ese tipo de androides que, en la práctica fueran indistinguibles de un humano (robots que pasarían finalmente el llamado test de Turing).
 - 4) Ahora bien, entre los sistemas psicobiofísicos, producidos por la evolución del universo real, y los sistemas computacionales, producidos por la ingeniería computacional de los androides, ¿existe una identidad ontológica? Es decir, ¿son idénticos por el *hardware* que los constituye internamente? Demos puntualizar que no albergamos, ni nadie puede albergar, ninguna duda sobre la respuesta. Las entidades de ontología biológica están hechas de tejidos vivos en los que ha emergido la sensibilidad-conciencia, una psique constituida por conciencia y subjetualidad psíquica, tal como hemos explicado y hoy describe la ciencia. Las entidades de ontología computacional, en cambio, están hechas de chips, de silicio, de unidades de memoria digital, de CPSs, de programas o algoritmos de operaciones, de redes neurales o unidades de procesamiento físicas (PDP), etc. Pero todo ello es objetivamente un sistema lógico-matemático de algoritmos «mecánicos y ciegos». No tenemos ningún dato objetivo que nos permita dudar que sea así. En otras palabras, es evidente que no existe una «identidad ontológica» entre el hombre (seres vivos) y el ordenador (máquinas).
 - 5) Ahora bien, entre las funciones, procesos y estados de ambos sistemas —las entidades biológicas y las computacionales—, ¿existe identidad? El que dos sistemas sean capaces de llegar a resultados similares (resultados inteligentes, análisis racionales, reflexión sobre sí mismo, personalidad, identidad, libre albedrío...), no justifica afirmar que los dos sistemas lleguen a esos resultados a través de funciones, procesos y estados idénticos. Es decir, la similitud en los resultados no permite hablar de identidad funcional. Sabemos que no existe una identidad ontológica entre entidades biológicas y entidades computacionales. Nadie puede ponerlo en duda sensatamente (aunque sabemos, eso sí, que marcar la perdiz y echar tinta de calamar siempre es posible). La mente humana

funciona poniendo en juego la sensibilidad-conciencia, la psique (conciencia más sujeto) en una coordinación armónica con muchos procesos mecánicos y deterministas, también regulados por el cerebro. La «mente» computacional, en cambio, funciona por algoritmos lógico-matemáticos mecánicos, deterministas y ciegos. Lo que el robot/androide hace (o hará) nunca podrá ser otra cosa en realidad que «simular» propiedades funcionales, procesos y estados de las entidades biológicas: conciencia, subjetualidad, identidad, libre albedrío, etc. Pero no hace esa simulación ni a partir de una «identidad ontológica», ni a partir de una «identidad funcional». El computador puede llegar a resultados similares (por ejemplo, libre albedrío en la máquina Watson que juega al *Jeopardy*), pero llega a ellos por procesos ontológicos, y en consecuencia también funcionales, diferentes. De ahí que, en nuestra opinión, la «metáfora fuerte» del ordenador no tiene argumentos serios a su favor (no así la «metáfora débil» que en muchos sentidos puede ser admitida).

- 6) Por consiguiente, que el computador pueda «simular» funciones, procesos y estados de las entidades biológicas (psico-bio-físicas), tales como conciencia, subjetualidad, identidad, personalidad, emotividad, libre albedrío, etc., no quiere decir ni identidad ontológica ni identidad funcional. No parece, pues, apropiado calificar con términos idénticos las funciones, procesos y estados de una entidad biológica y de una entidad computacional. Nuestra propuesta sería obvia: hablar, por una parte de sensibilidad-conciencia, o simplemente conciencia, biológica, o psico-bio-física, y, por otra parte, de conciencia computacional, de identidad computacional, de libre albedrío computacional, etc., de tal manera que al añadir el adjetivo «computacional» quedara explícitamente señalado que se trata de una forma especial de esas propiedades que no es idéntica a la de los seres vivos, aunque pueda ser similar¹⁵.

6.1. Reconocimiento de patrones en la mente biológica y en la mente computacional

La teoría de la mente fundada en el reconocimiento de patrones es el núcleo del pensamiento antropológico de Kurzweil. La mente computacional (con su hardware propio) y la mente biológica, el cerebro animal y humano (con su hardware biológico-neuronal) funcionan como un análisis de patrones. Por tanto, funcionan exactamente igual: el cerebro es un computador biológico (esta es la tesis de la IA fuerte). Es, en definitiva, la tesis que defiende en *How to Create a Mind*. Entre el funcionamiento de un computador y el sistema neural hay una similitud total: el análisis de patrones permite entender cómo funciona mente biológica y en ello se muestra su identidad con la mente computacional.

La mente computacional. Mucho antes de que Kurzweil se interesara en serio por la mente biológica había desarrollado su amplio trabajo profesional, siempre orientado al análisis y reconocimiento computacional de patrones en tres campos definidos: análisis y reconocimiento de palabras en textos escritos; análisis y reconocimiento de voz y del significado de los sonidos; así como análisis y reconocimiento de imágenes (de cualquier orden). Esto le llevó a desarrollar temáticas tan importantes como el estudio de los rasgos simples y complejos, las escalas e identidad de rasgos, la jerarquía y organización estructural de rasgos, la eliminación, selección y conexión de patrones por los modelos de Markov, los procesos bottom up y top down, los sistemas de links y las redes

¹⁵ Nota sobre ingeniería computacional.

de patrones, las gramáticas de patrones, el tratamiento de los análisis por algoritmos matemáticos, etc.

Las aportaciones de Kurzweil, y su equipo, en forma de algoritmos matemáticos que analicen y reconozcan textos escritos, voz o imágenes multimodales, son una forma de análisis de rasgos (patrones) en la línea de la visión artificial por computación serial similar a la escuela clásica de David Marr. Kurzweil lo complementa con análisis computacional por PDP conexionista (*paralell distributed processing*), aunque siempre implementado en la simulación de esos procesos PDP en algoritmos seriales. Teorías como la aplicación multimodal de las series de Fourier para el análisis de los sentidos, o la alternativa de la teoría de la percepción directa de J.J. Gibson, son completamente ignoradas por Kurzweil.

La mente biológica. Kurzweil no comenzó por la biología, sino por la ingeniería del conocimiento. Ahora bien, consciente de que los programas de IA y simulación del ordenador funcionaban por el análisis de patrones, establece una de las hipótesis fundamentales de su pensamiento: a saber, que también la mente biológica, animal y humana, funciona por análisis y reconocimiento de patrones. Es evidente que esta *identidad funcional* es básica para argumentar que máquinas y hombres son en último término lo mismo y, al margen de la ontología, coinciden en las propiedades de esa dimensión transhumana nueva que permite el humanismo extensivo. Para argumentar esta similitud funcional Kurzweil emprende una revisión —y reinterpretación— de nuestros conocimientos sobre el sistema nervioso: cerebro antiguo, neocortex, células hipercomplejas, hipercolumnas... hasta llegar al cálculo de los trescientos millones de reconocedores de patrones, registrados en una jerarquía de niveles como ocurre en el computador¹⁶. ¿Es esto así? Prescindiendo de los detalles de la interpretación neurológica de Kurzweil (que tiene mucho cuestionable o, al menos, matizable) pensamos que su interpretación de la mente humana como análisis y reconocimiento de patrones es fecunda, y probablemente correcta.

¿Por qué lo decimos? Para abreviar, pongamos un ejemplo: la mente animal de un mamífero, un perro, por todos conocido. Su cerebro primitivo sensorio-motriz (sensaciones y respuestas adaptativas), por ejemplo el de un anfibio como *autómata sensitivo*, se ha ido ampliando al crear las zonas temporales del cortex, de tal manera que los *engramas* (o *patterns* neurales) que registran las experiencias en tiempo real se asientan en una memoria más eficaz que permite *links* entre el pasado y el presente. Cuando el perro observa, por vista y olfato, que su dueño se acerca por el camino su registro en tiempo real activa una serie de links que reactivan por memoria registros del pasado que se hacen presentes difusa y enmascaradamente: su amo acariciándolo, dándole de comer, llevándolo a pasear, dirigiéndole la palabra... Entonces la valencias emocionales del animal se activan y el rabo comienza a moverse de alegría. El pasado se hace presente: es el *remembered present* de Edelman. El perro tiene un concepto de su amo (el conjunto de todos esos engramas sobre él). El perro tiene una mente compleja que funciona por una combinatoria de imágenes (patrones) que están ordenadamente registradas en su sistema neural. Es capaz categorización, abstracción, inferencias lógicas, imaginación y de anticipación del futuro¹⁷. La mente humana funciona de forma similar, aunque con una mayor complejidad por la emergencia de la razón. Su cerebro de conexión se ha ampliado en forma considerable al crecer las zonas temporales y las zonas frontales

¹⁶ Cfr. CCM, o.c., pags. 31-105.

¹⁷ Véase Monserrat, Javier; *La percepción visual. La arquitectura del psiquismo desde el enfoque de la percepción visual*, Biblioteca Nueva, Madrid 2008, segunda edición. Capítulo XIV, págs.. 533ss.

y prefrontales. Las ideas, el pensamiento, el conocimiento, que construye la razón humana son también engramas o patterns que conectan con imágenes de otros módulos cerebrales. Una inmensa cantidad de buses conducen los links que vinculan todos los engramas del cerebro (el mapa completo de estos links es lo que quiere trazar el *Brain Project* de financiación europea). Aunque deberíamos matizarlo más, puede decirse, por tanto, que la mente racional del hombre funciona también como una compleja combinatoria de patrones.

Los patrones computacionales, ¿son idénticos a los patrones neurales? Esta es, en definitiva, la cuestión más importante. Los patrones computacionales que la máquina registra por sus sensores propios y los procesos de análisis, registro y reconocimiento, así como la combinatoria entre ellos, ¿son idénticos que los que tienen lugar en la mente humana, biológico-neural, de animales y hombres? Lo que hemos explicado en este artículo hasta el momento no permite afirmarlo. El análisis de patrones en el computador y en el hombre, pueden quizá ser similares, pero no son idénticos. Entre ordenador y hombre no existe identidad ontológica y, en consecuencia, tampoco existe identidad funcional. Así, el computador funciona por procesos (programas de algoritmos matemáticos) que son mecánicos y ciegos. En cambio, el hombre, la mente humana, aunque construida sobre una base de procesos inconscientes que son automáticos, tiene una cúspide en que aparecen los *qualia*, es decir, sensaciones biológicas, imágenes sensibles, imaginación, percepción, recuerdo de las propias ideas registradas en la mente, etc. La razón humana superior funciona por una combinatoria y análisis de imágenes sesibles que constituyen la esencia de la mente y de la conciencia del «yo» como sujeto psíquico. Por tanto, cuando Kurzweil se pregunta cómo crear una mente (es decir, cómo crear una mente biológica en el ordenador) plantea una expectativa falsa porque el computador no puede ni ontológica y funcionalmente crear una mente. En términos de ingeniería del conocimiento sólo puede *simularla*.

7. LA SINGULARIDAD ESTÁ CERCA

«Entonces, ¿qué es la Singularidad? Es un tiempo venidero en el que el ritmo del cambio tecnológico será tan rápido y su repercusión tan profunda que la vida humana se verá transformada de forma irreversible. Aunque ni utópica ni distópica, esta Era transformará los conceptos de los que dependemos a la hora de dar significado a nuestras vidas ya sea en lo que se refiere a modelos de negocios o al ciclo de la vida (incluyendo la muerte). Así, comprender la Singularidad cambiará la perspectiva sobre la relevancia de nuestro pasado y sus repercusiones en el futuro, ya que su comprensión cambia intrínsecamente la visión que cada uno tiene de la vida en general y de su propia vida en particular. A las personas que han entendido la Singularidad y que han reflexionado sobre sus implicaciones en su propia vida yo las llamo «singularistas»¹⁸.

Como sabemos, la prognosis de que la evolución tecnológica hará entrar a la humanidad en un nuevo tiempo transhumano, en que incluso llegará a vencerse la amenaza ancestral de la muerte sobre el ser humano, es la propuesta de Kurzweil **más arriesgada y más sometida a críticas**. Kurzweil se ha sentido interpelado y ha respondido honestamente a sus críticos¹⁹. Su predicción está fundada en la *Ley de Rendimientos Acelerados*

¹⁸ Cfr. SEC, o.c. p. 17.

¹⁹ Por ejemplo en: SEC, o.c., 491ss.

(*The Law of accelerating Returns*) y en otras de menor importancia como la *Ley de Moore*. Propone un análisis de la historia en seis etapas que culminan en la sexta etapa en que el universo está completamente conocido y codificado en inmensas redes de computación.

El avance tecnológico que hace posible la singularidad. Nos viene expuesto en *The Singularity is Near*, y se ha repetido en multitud de lugares. Respondería a estos diversos niveles de avance tecnológico.

- 1) Crecimiento exponencial de la tecnología de la computación. Crecerá el tamaño de los circuitos, así como la velocidad de procesamiento. Se llegará también al abaratamiento exponencial coste/rendimiento. Pero lo principal será que aparecerán *máquinas capaces de autocontrol*. Capaces de diseñar sus propios objetivos. Capaces de diseñar sus propios programas. Capaces de control de la información, del conocimiento y del universo, a través de inmensas redes de computación. Uno de sus aspectos será el control externo de los cyborgs, de los cerebros humanos y de la biología humana en general.
- 2) Existirán cyborgs perfectamente contruidos, inspirados en la simulación del hombre: de inteligencia, programas de simulación, percepción, registro de datos, procesamiento, infinitamente superior del hombre, con su inteligencia biológica. Serán cyborgs identificados con la especie humana, formando con ellos una dimensión existencial unitaria transhumana.
- 3) Se producirá un aumento y perfeccionamiento significativo de la condición biológica humana por crecimiento de la inteligencia y de la emotividad. Implantes que conectarán directamente el cerebro a las redes externas y a los otros cyborgs darán una mejor eficiencia biológica. Se alcanzará una mejora de la velocidad de procesamiento neural por efecto de la inserción de robots nanotecnológicos.
- 4) De esta manera habrá un control, mejora y corrección de errores biológicos: por conocimiento y control de los códigos genéticos, por conocimiento anticipado del futuro biológico, por inserción de nanorobots de control, mejora y corrección de los errores existentes, por control de la biología individual por redes externas que conocerán perfectamente el estado del organismo y darán las instrucciones para su control.

¿Inmortalidad? Esta es, en efecto, la arriesgada predicción hecha por Kurzweil a través del concepto de Singularidad. ¿Qué pensar? Obviamente las valoraciones han sido variadas y de diferente signo. ¿Cuál es nuestra opinión? La presentamos conclusivamente en los puntos que consideramos esenciales.

- 1) Ante todo establecemos un Presupuesto General para cualquier valoración: no ponemos en duda los avances sorprendentes que puedan producirse por la tecnología en computación, control del universo, cyborgs, nanotecnología, longevidad...
- 2) Supuesto este presupuesto debemos establecer también un criterio general que depende de cuanto hemos dicho en este escrito: todo debe valorarse desde los principios de no-identidad ontológica y no identidad funcional. Por tanto, en consecuencia, la psique humana no podrá nunca estar conectada ni ontológica ni funcionalmente a los cyborgs o a las redes externas. Entre lo mecánico y ciego y lo biológico con sensibilidad-conciencia habrá siempre un abismo ontológico y funcional.
- 3) La tecnología es un diseño humano cuya construcción y propiedades está siempre sometida lógicamente a los intereses humanos y de la especie por criterios morales y éticos. En este sentido el futuro de la Singularidad está sometida a

la ciencia política y a la filosofía de la historia, campos que son completamente ignorados por Kurzweil. Por tanto, lo que pueda suceder en el futuro estará siempre sometido a la valoración socio-política de lo que interese a la especie humana.

- 4) A la especie humana no interesa un universo de máquinas independientes: de **máquinas, redes**, cyborgs, capaces de autoprogramación y de ingeniería propia, independientes, al margen del control humano. Ya Stanley Kubrick, con una gran clarividencia histórica, planteaba en la *Odisea en el Espacio 2001*, la eventualidad de un universo de **máquinas que se revelan y autonomizan frente al hombre y la especie humana**.
- 5) ¿Inmortalidad? Una creencia posible, pero objetivamente incierta, para el individuo y mucho más para la totalidad de la humanidad. Son muchas las variables intervinientes que sabemos que existen y que Kurzweil no parece haber tomado en cuenta.
- 6) La Singularidad, en un sentido u otro, no es problema para la fe cristiana. Esta no es nunca un freno para el progreso, ya que entiende que Dios ha puesto el universo en manos de la creatividad humana. La regulación del progreso depende de razones éticas y morales, así como de la gobernanza socio-política de las naciones a favor de la especie humana. El cristianismo está, pues, a la expectativa de lo que pueda suceder. Entiende que el hombre, ya desde la imagen de la fruta del árbol de la ciencia del bien y del mal, está en busca de aquel estado en que nuestro dominio del universo nos haría como Dioses. La historia es una lucha entre la apertura a Dios y la clausura en el puro mundo. Pero no sabemos hasta dónde se desarrollará la imagen del Misterio de Iniquidad o cuando se producirá la Segunda Venida de Cristo que supondrá el fin de la historia en el mundo.

7. CONCLUSION: EL «HUMANISMO EXTENSIVO» POSIBLE

No ponemos en duda que se pueda llegar a la construcción de robots-androides o cyborgs tan perfectos que apenas puedan ser distinguidos de los humanos (que pasaran el test de Turing). No lo afirmamos, porque no tenemos las creencias futurólogas de Kurzweil. Pero podría hacerse realidad. No ponemos límite a la capacidad de programar una simulación de los seres vivos. Estas máquinas podrían tener, no lo negamos, lo que en el párrafo anterior denominamos, específicamente, conciencia computacional, identidad computacional, libre albedrío computacional, emoción computacional, etc. Pero estas máquinas no tendrían ni identidad ontológica ni identidad funcional con las entidades que llamamos biológicas, o psico-bio-físicas. Pensar así es una exigencia de los hechos objetivos y de la ciencia. ¿Es posible argumentar en contra? No lo creemos.

No obstante, es por otra parte evidente que la construcción de estos nuevos robots-androides cuasi-humanos, representa un salto cualitativo importante en la historia de la ingeniería. Estas *spiritual machines*, como Kurzweil suele decir, no serán para nosotros igual que un ventilador o una nevera, por ejemplo. Darán comienzo a una nueva forma de comportamiento moral y ético para con ellas. De la misma manera que un animal es biológico, pero no humano, y existe una moral en el comportamiento ante los animales, así igualmente las nuevas *spiritual machines* no tendrá sensibilidad-conciencia biológica (que los animales tienen), pero tendrán más cercanía al hombre porque poseerán mejor ciertas propiedades similares (o sea, computacionales) como conciencia, identidad,

razón y diseño existencial,, emotividad, personalidad, etc. ¿Cómo se integrarán estas máquinas en nuestras vidas? Sin duda que se deberá tener ante ellas un respeto y reconocimiento de su condición cuasi-humana. La simpatía, respeto, solidaridad existencial, con que los humanos tratan a los cyborg androides en la ciencia ficción que vemos en novelas, en películas y series de televisión, son un anticipo de lo que probablemente será la integración moral de esas *spiritual machines* en la sociedad del futuro.

Es también evidente que la aparición de estos androides supondrá la aparición de un incuestionable «humanismo extensivo». La condición humana se habrá sin duda «extendido» a máquinas creadas por el mismo hombre (con las cautelas y restricciones que hemos explicado). Pero, aunque esas máquinas no sean biológicas (no tengan una ontología y una funcionalidad psico-bio-física), serán ciertamente una forma de extensión artificial de la condición humana. Por tanto, habrá un «humanismo extensivo» posible, fundado en la similitud entre estas *spiritual machines* y el hombre. Pero un «humanismo extensivo», entendido tal como parece hacerlo Kurzweil, no creemos que fuera posible. Como decía, en otros dos artículos, deberemos matizar también otros aspectos importantes del pensamiento de Kurzweil.

Universidad Autónoma de Madrid,
Cátedra CTR, Escuela Técnica Superior de Ingeniería,
Universidad Comillas, Madrid
jmonserrat@res.upcomillas.es

JAVIER MONSERRAT

[Artículo aprobado para publicación en diciembre de 2015].