La propuesta de inmortalidad tecnoinducida y el "totalismo cibernético" de Raymond Kurzweil. Algunas consideraciones antropológicas

Francisco T. Baciero Ruiz

Universidad de Salamanca

XVIII Congreso Ciencia y Vida

Universidad San Pablo CEU. Madrid, 22-24 julio 2013

"En aras de la autonomía humana, de la dignidad que exige, de que nos poseamos a nosotros mismos y no nos dejemos poseer por nuestras máquinas, tenemos que poner el galope tecnológico bajo control extratecnológico"

H. Jonas, *Técnica, medicina y ética. La práctica del principio de responsabilidad,* p. 39

Introducción: La Inteligencia Artificial y el "totalismo cibernético":

A mediados del siglo XVIII, uno de los fundadores del materialismo ilustrado, el médico Julien O. de La Mettrie afirmaba en 1748 en *El hombre-máquina*, una obra polémica desde su publicación, en línea con el mecanicismo cartesiano, que el hombre podía ser descrito como una máquina o bien como un animal, o lo que es lo mismo, que su comportamiento era explicable por las mismas leyes que las de cualquier otro objeto de la naturaleza, puesto que estaba hecho de la misma materia, aunque quizás más sofisticada:

"¿Hace falta más … , para probar que el hombre no es más que un animal o un conjunto de resortes, que se montan unos sobre otros, sin que pueda decirse por qué punto del círculo humano empezó la naturaleza? … . Por consiguiente el alma no es más que un principio de movimiento o una parte material

sensible del cerebro, que se puede considerar, sin temor a equivocarse, como el resorte principal de toda la máquina, el cual tiene una influencia visible sobre todos los demás"[1]

La Mettrie, pararójicamente, como P. Singer dos siglos y medio más tarde, borraba al mismo tiempo cualquier frontera precisa entre hombres y animales, elevando a los animales a la categoría de los primeros y degradando a los hombres disminuidos a la de los animales:

"Sigamos las operaciones del mono, del castor, del elefante, etc. Si es evidente que éstas no pueden hacerse sin inteligencia, ¿por qué rechazársela a estos animales? … . Os persuadiréis también de que el imbécil o el estúpido son animales de aspecto humano, al igual que el mono lleno de inteligencia, es un hombre en pequeño bajo otra forma, y finalmente, ya que todo depende absolutamente de la diversidad de la organización, os daréis cuenta de que un animal bien construido, al que se le haya enseñado astronomía, puede predecir un eclipse, así como la curación o la muerte, cuando ha ejercitado durante un tiempo su talento"[2].

La Mettrie tenía en alto grado las excelencias morales de su teoría, pues a pesar de denigrar con ella la autoimagen del hombre, ello no afectaría para nada a su conducta, antes al contrario:

"En fin, el materialista convencido, por mucho que su propia vanidad murmure que sólo es una máquina o un animal, no maltratará a sus semejantes, por estar demasiado instruido acerca de la naturaleza de estas acciones, … , y por no querer, en una palabra, según la ley natural dada a todos los animales, hacer al prójimo lo que no quisiera que se le hiciera a él"[3].

---

[1] J. O. de la Mettrie, *Obra filosófica,* Ed. preparada por Menene Gras Balaguer, Editora Nacional, Madrid, 1983, "El hombre-máquina", pp. 197-251, pp. 240-1; "Concluyamos osadamente que el hombre es una máquina, y que en todo el universo no existe más que una sola sustancia modificada" (ibid., p. 250).

[2] Ibid., p. 249.

[3] Ibid., p. 250, cf. asimismo en pp. 217 ss. la teoría –reivindicada por ciertos primatólogos desde hace medio siglo- de que el orangután es una especie tan próxima al hombre, que podría aprender a hablar con la educación adecuada: "¿Por qué, pues la educación de los monos habría de ser imposible? ¿Por qué finalmente, a fuerza de cuidados, no podría imitar, a ejemplo de los sordos, los movimientos necesarios para pronunciar?" (ibid., p. 218), puesto que, anticipando tesis darwinistas: "La transición de los animales al hombre no es violenta; los verdaderos filósofos lo reconocerán. ¿Qué era el hombre, antes de que se inventaran las palabras y se conocieran las lenguas? Un animal de su especie, el cual, con mucho menos instinto natural que los demás, de los que entonces no se creía rey, no se distinguía del mono y de los restantes animales más de lo que el propio mono" (ibid., p. 219).

La autocomprensión del hombre como un animal más o menos evolucionado en la escala filogenética ha conocido una aceptación creciente en las últimas dos décadas gracias sobre todo a las obras de P. Singer y al movimiento social que han generado en defensa de los supuestos "derechos" de los animales, que se pretenderían equiparar con algunos de los derechos humanos, comenzando por el derecho a la vida. Se trata a la postre de un discurso inaugurado en lo esencial por Darwin, quien insistió reiteradamente en que las diferencias que separaban al hombre de los animales eran sólo "de grado," no de esencia[4].

La equiparación del hombre con una máquina, sin duda mucho más contraintuitiva que su equiparación con el animal, ha tenido y tiene menos adeptos que el *animalismo* de Singer y sus seguidores, si bien ha logrado una plausibilidad creciente en determinados ambientes relacionados con la ingeniería informática y merced a la creciente difusión de las tecnologías informáticas.

El primer jalón en esta autocomprensión del hombre como una máquina, más allá de la mitología o la literatura[5], hay que buscarlo sin duda en la llamada "conferencia de Darmouth College" del verano de 1956, en la que un grupo de expertos en computación acuñó el término "Inteligencia Artificial" (A.I. de acuerdo con sus siglas inglesas), para describir el objeto de sus investigaciones[6]. El término era lo suficientemente ambiguo como para suscitar la idea que podrían existir artefactos o "máquinas" pensantes (en la medida en que parecería denotar objetos inanimados de los que se podría predicar inteligencia, una capacidad hasta entonces reservada exclusivamente a los seres vivientes racionales: bien el hombre, bien los ángeles o, en un sentido especial, Dios mismo). Hay que decir que el término obtuvo bastante éxito, como confirma el hecho de que hoy

---

[4] "no hay diferencia esencial en las facultades del hombre y mamíferos superiores" (Ch. Darwin, *El origen del hombre y la selección en relación al sexo*, Edaf, Madrid, 1989, p. 71), "La mayoría de las más complejas emociones que experimentamos son comunes a los animales superiores" (ibid., p. 76), "El naturalista no puede comparar ni clasificar las facultades mentales, y sí solo intentar demostrar, como yo he hecho, que entre las del hombre y las de los animales inferiores no existe una diferencia esencial y específica, sino sólo una inmensa en grado" (ibid., p. 147), y al final de la obra: "aquel que admita el principio de la evolución debe reconocer que las facultades mentales de los animales superiores, que en naturaleza son lo mismo que las humanas, aunque en grado diferente …" (ibid., pp. 512-3).

[5] Cf. una interesante historia de las leyendas y arquetipos literarios, desde los autómatas de la *Ilíada* de Homero hasta los autómatas descritos en la obra de teatro de 1921 del checo Karel Capek *R.U.R. (Robots Universales Rossum)* que popularizó el término *robot* (del eslavo *robota*, "trabajo forzado"), pasando por el *golem* del rabino praguense Judah ben Loew en el siglo XVII, o el clásico *Frankenstein* de Mary Schelley, en P. McCorduck, *Máquinas que piensan,* Tecnos, Barcelona, 1991, 21 ss.. Cf. igualmente otra excelente introducción en T. de Andrés, *Homo cybersapiens. La inteligencia artificial y la humana,* Eunsa, Pamplona, 2002, pp. 19-55.

[6] Uno de los convocantes de la conferencia, J. McCarthy, logró imponer el nombre de "Inteligencia Artificial" para describir el tipo de trabajo llevado a cabo hasta entonces con dispositivos computacionales, en lugar del hasta entonces habitual (y más ajustado a la realidad), de "estudios de autómatas" (McCorduck, pp. 105 ss., p. 108).

oigamos hablar en numerosos contextos (sólo en sentido figurado, en principio), de "ascensores inteligentes", "edificios inteligentes", "electrodomésticos inteligentes", etc..

Sin embargo, la metáfora de la máquina o del homúnculo pensante ha perdido su carácter metafórico para algunos filósofos de la mente, y especialmente para un grupo de ingenieros informáticos partidarios de la denominada "Inteligencia artificial fuerte", que pretende que los dispositivos computacionales podrán replicar con el tiempo a la perfección la mente humana en todas sus facultades, entre ellas las más complejas y específicamente humanas, como la conciencia, la intencionalidad, la capacidad de sentir o experimentar emociones, etc.[7]. El fundador de esta corriente es el ingeniero informático y uno de los convocantes de la mencionada conferencia de Darmouth, M. Minsky (1927-), que ha escrito algunas obras de divulgación sobre la materia[8]. Discípulos directos o indirectos de Minsky que asumirían su filosofía computacional de la mente serían: Raymond Kurzweil, H. Moravec[9], y más recientemente K. E. Drexler[10] o R. A. Freitas[11], vinculados con el MIT de la universidad de Harvard.

---

[7] Mientras que los defensores de la IA "débil" afirman que las máquinas pueden actuar *como si* pensasen (lo que casi todos los investigadores en la materia están dispuestos a admitir), la IA "fuerte" sostiene que los ordenadores *piensan actualmente,* con todas las consecuencias que ello implicaría (cf. S. Russell, P. Norvig, *Artificial Intelligence. A modern approach*. Prentice Hall Pearson Education International, 2003², pp. 947 ss.).

[8] Minsky (1927-). Entre sus obras de "filosofía de la mente" cabe citar sobre todo *The Society of Minds,* New York, Simon & Schuster, 1986, en la que postula su teoría de la mente como "sociedad" o asociación de procesos *complejos* cerebrales funcionando en paralelo, de los que surgirían los procesos psíquicos superiores, cuya complejidad funcional de momento es incomprensible, pero no *en el futuro,* lo que con el tiempo permitirá construir máquinas "conscientes, ambiciosas, celosas, con sentido del humor" o cualquier otra "experiencia viva mental" ("mental life-experience", *op. cit.*, p. 19), y de modo semejante en *The Emotion Machine: Commomsense Thinking, Artificial Intelligence, and the Future of Human Mind,* New York, Simon & Schuster, 2006 (*La máquina de las emociones,* Debate, 2010), donde sostiene la posibilidad de crear máquinas que "piensen" y "sientan" ("to design machines that can feel and think", p. 7). Un artículo programático de Minsky sobre el "totalismo cibernético" puede considerarse "¿Serán los robots quienes hereden la Tierra?", *Investigación y ciencia*, diciembre 1994, pp. 86-92.

[9] Nacido en 1948 en Austria y emigrado a Canadá en 1953 y más tarde a Estados Unidos, realizó su tesis doctoral con uno de los convocantes de la conferencia Darmouth, John McCarthy. Aparte de sus publicaciones especializadas en robótica, sus dos libros futuristas sobre las consecuencias de la revolución robótico-informática son: *Mind Children: The Future of Robot and Human Intelligence,* Harvard University Press, Cambridge, Massachusetts, 1988, y *Robot. Mere Machine to Trascendent Mind,* Oxford University Press, 2000. En español: *El hombre mecánico. El futuro de la robótica y la inteligencia humana*, Barcelona, Salvat, 1993.

[10] Especialista en nanotecnología, autor de: *Engines of Creation,* New York, Anchor Books, 1986, y *Nanosystems: Molecular Machinery, Manufacturing and Computation,* New York, Wiley Interscience, 1992, y en colaboración con Ch. Peterson, *Unbounding the future: The Nanotechnology Revolution,* New York, William Morrow, 1991.

[11] Freitas es uno de los pioneros de la nanomedicina: *Nanomedicine,* vol. 1, *Basic Capabilities* Georgetown: Tex., Landes Bioscience, 1999 y *Nanomedicine,* vol. II, *Biocompatibility,* Georgetown: Tex.: Landes Bioscience, 2003. Previamente ya había escrito sobre el tema en R.A. Freitas, "El nacimiento del ciborg", en MINSKY (ed.), *Robótica*, Barcelona, 1986, pp. 121-151.

La atribución de inteligencia y conciencia (además de capacidades emocionales), en el sentido literal de la palabra, a las máquinas, supondría de modo inevitable la atribución de libertad a las mismas[12], y en consecuencia, de un comportamiento moral, lo que haría inevitable en última instancia también la atribución de derechos "humanos" a dichos artefactos, por la equiparación de su naturaleza en esos aspectos esenciales desde el punto de vista moral con la humana. Ello constituiría uno de los principios del llamado *transhumanismo*: el desbordamiento de las fronteras biológicas en la atribución de determinadas propiedades consideradas hasta ahora exclusivamente humanas a determinados seres no humanos, las máquinas "pensantes" en este caso. Dado que la atribución de todas las propiedades antedichas a las máquinas: conciencia, sensibilidad, afectos, es extremadamente contraintuitiva en el *mundo de la vida* (es decir, para las categorías precientíficas o prefilosóficas de las que hacemos uso habitualmente en la vida cotidiana), la mayoría de estos autores suelen aducir como confirmación indirecta de las mismas, que en realidad los propios seres humanos somos en un sentido literal máquinas o autómatas conscientes[13].

Algunos colegas han llegado a denominar a este grupo de ingenieros informáticos partidarios fervorosos de la inteligencia artificial fuerte, la "secta" de los "totalistas cibernéticos", por considerar su fe ciega e injustificada a todas luces desde un punto de vista puramente técnico[14].

Nos ocuparemos a continuación de Raymond Kurzweil, uno de los "totalistas cibernéticos" que más difusión mediática ha logrado en los últimos años.

---

[12] En la medida en que la razón es el fundamento de la conducta libre: "pro tanto necesse est quod homo sit liberi arbitrii, ex hoc ipso quod rationalis est" (Tomás de Aquino, *S. Th.*, q. 83, a. 1, *in corp.*).

[13] "nuestro cerebro no es más que una máquina. Y, como no es concebible que una máquina pueda ser libre, la humanidad llegaría así a comprender que la libertad individual no tiene sentido alguno" (L. Ruiz de Gopegui, *Cibernética de lo humano*, Fundesco/Tecnos, Madrid, 1983, p. 197); Minsky puso de moda por su parte en los años setenta del siglo pasado la expresión: "el cerebro es una máquina de carne" (J. L. Ruiz de la Peña, "Homo cyberneticus? Antropología e inteligencia artificial", en J. de Sahagún Lucas (dir.), *Nuevas antropologías del siglo XX,* Sígueme, Salamanca, 1994, p. 102).

[14] Entre ellos se encuentra J. Lanier, "La mitad de un manifiesto", en J. Brockmann (Ed.) *El nuevo humanismo y las fronteras de la ciencia,* Kairós, Barcelona, 2007, pp. 292-325, y J. Lanier, *Contra el rebaño digital. Un manifiesto,* Debate, Barcelona, 2011 (Lanier, él mismo informático, fue el primero en diseñar un casco de visión para realidades virtuales, y a él se debe la invención del término mismo "realidad virtual"). Antes que Lanier se debe citar a J. Weizenbaum (inventor en 1966 del programa ELIZA de simulación de comunicación interactiva hombre-máquina en lenguaje natural), que criticó duramente el paradigma de la IA fuerte en *Computer Power and Human Reason: From Judgment to Calculation*, 1976 (*La frontera entre el ordenador y la mente,* Madrid, Pirámide, 1978), además de B. Schneiderman (inventor de la idea de hacer *clic* sobre un enlace en la pantalla del ordenador), o D. Gelernter, investigador en computación paralela (cf. Lanier, 2011, p. 33). Una excelente introducción crítica a la cuestión de la "antropología cibernética" sigue siendo la de Ruiz de la Peña.

2. *El* singularitarismo *de Raymond Kurzweil:*

Raymond (o "Ray") Kurzweil (1948-), hijo de padres austríacos emigrados a Estados Unidos en los años treinta del siglo XX, es un ingeniero informático especialista en robótica e "inteligencia artificial" que estudió bajo la dirección de M. Minsky y se hizo famoso con la invención de un programa de lectura para ciegos presentado en 1976. Tras sus logros en robótica, por los que ha recibido numerosas distinciones, se ha convertido en autor de *best-sellers* de la literatura futurista científico-tecnológica, convirtiéndose en uno de los máximos exponentes de las utopías cibernético-tecnológicas[15].

Kurzweil ha elaborado especialmente en *La singularidad está cerca* (2005), una de las teoría-estrella de los "totalistas cibernéticos", la llamada teoría de la *singularidad ("Singularity")*, según la cual, el crecimiento exponencial de los conocimientos y la tecnología informática llegará a provocar en un momento impreciso, pero próximo de la historia de la humanidad, la eclosión de una nueva en la que las capacidades de las máquinas superarán con mucho a las humanas, hasta el punto de que ellas mismas serán capaces de autoreproducirse y perfeccionarse de modo mucho más perfecto que bajo supervisión humana, iniciándose de este modo un nuevo estadio de la historia regido por unos parámetros difícilmente previsibles y en cualquier caso inconmensurables con los conocidos hasta hoy[16].

---

[15] Entre sus obras, a medio camino entre la alta divulgación y la literatura futurista, se encuentran los siguientes (se indica por orden cronológico la fecha de la primera edición y la editorial y lugar de publicación si son conocidos, así como, entre paréntesis, las traducciones españolas conocidas): *The age of Intelligent Machines,* MIT, 1990, *The age of Spiritual Machines. When Computers Exceed Human Intelligence,* 1999 (*La era de las máquinas espirituales,* Planeta, 1999), *Are we Spiritual Machines? Ray Kurzweil vs. the critics of Strong A.I.,* Discovery Institute, 2002, *The Singularity is near. When humans transcend Biology,* Penguin, 2005 (*La singularidad está cerca. Cuando los humanos trascendamos la biología,* Lola Books, 2012), y *How to create a Mind. The Secret of Human Thought Revealed,* Viking, 2012. En colaboración con el médico Terry Grossmann, dentro de su programa de extensión indefinida de la vida humana ha publicado: *Fantastic Voyage. Live long to live forever,* New York, Plume, 2005, y *Trascend: nine steps for living well forever,* New York, Rodale, 2011. Kurzweil dispone de su propia página web personal, en la que se puede acceder a la mayoría de sus trabajos y a los de otros autores afines, en: http://www.kurzweilai.net.

[16] I. J. Good ya sugería en 1965 que: "an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion", and the intelligence of man would be left far behind" (J.I. Good, "Speculations concerning the first ultraintelligent machine", 1965, cit. en Russel and Norwig, 2003, p. 963); y el autor de ciencia-ficción y matemático V. Vigne anunciababa para el 2023 la eclosión de una nueva era de "singularidad tecnológica": "Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended" (V. Vigne, "The Coming Technological Singularity: How to Survive in the Post-human Era", VISION-21 Symposium, sponsored by the NASA Lewis Research Center and the Ohio Aerospace Institute, March 1993, cit. en Russel and Norwig, loc. cit.). El término habría sido empleado por vez primera por uno de los pioneros de la teoría

La *singularidad* (que Kurzweil anuncia para el año 2045), sería en este sentido el nombre de un hipotético nuevo y último acontecimiento cuasi escatológico de la evolución humana que contemplaría la aparición de una era caracterizada por el dominio completo de la tecnología y de la inteligencia creada sobre la humana, la construcción de "máquinas inteligentes", la fusión de la inteligencia humana y artificial y, como corolario, el logro de la inmortalidad humana individual[17].

El crecimiento exponencial de los éxitos de la tecnología cibernética estaría garantizado según Kurzweil por la llamada "ley de Moore", formulada por el ingeniero informático Gordon Moore a mediados de los años 70 del siglo XX, sobre el crecimiento exponencial regular de la memoria de los ordenadores[18]. De acuerdo con esta ley, y con la que Kurzweil denomina "ley de rendimientos acelerados"[19], la interacción de todas las tecnologías disponibles hoy (especialmente la genética, la nanotecnología y la robótica), conducirá a un crecimiento exponencial de la tecnología hasta límites insospechados[20].

Para Kurweil existirían cinco etapas claves en la evolución de la humanidad, evolución que Kurzweil entiende exclusivamente en términos de procesamiento de la información en niveles crecientes de orden y complejidad[21]. La singularidad comenzaría en la tierra en la quinta etapa, y "saltaría" de la tierra a todo el universo en una sexta y

---

de la información y computación, John Von Neumann (1903-1957), quien en la década de los 50 del siglo XX habría hablado del acercamiento de una "singularidad esencial en la historia de la humanidad" como consecuencia del progreso tecnológico: "some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue" (según el testimonio de Stanislaw Ulam, "Tribute to John von Neumann", *Bulletin of the American Mathematical Society, 64.3,* pt. 2 (Mayo 1958), pp. 1-49, cit. en Kurzweil, 2005, p. 10 y p. 498, nota 3).

[17] Kurzweil, 2005, pp. 135-6; sobre la cuestión de la inmortalidad, vid. *infra*.

[18] Cf. p. 56, y 1999, 408. Moore definió su famosa ley en un artículo de 1965 ("Cramming More Components onto Integrated Circuits", *Electronics* 38.8, April, 19, 1965), 114-17, cit. en Kurzweil, 111 ss., p. 526, nota 1), en el que predecía que, de acuerdo con los datos disponibles desde 1959, la tecnología informática permitiría reducir anualmente a la mitad el tamaño de los transistores alojados en un circuito integrado de un ordenador, doblándose por ello su potencia en el mismo período y por el mismo coste (Moore tuvo que corregir a la baja su predicción diez años después -Kurzweil, pp. 111-112-, pero la idea que ha prevalecido entre los informáticos es que el poder de computación aumenta constantemente, en grado supuestamente exponencial).

[19] "Law of accelerating returns" (p. 72, pp. 491-6 y *passim*).

[20] Kurzweil, 2005, p. 491; cf. en las pp. siguientes un intento tratamiento matemático de la cuestión a partir de las variables "Velocidad (poder) de computación medida en cálculos por segundo por unidad de coste", "Conocimiento del mundo, en cuanto perteneciente al diseño y construcción de aparatos computacionales" y "Tiempo". Kurzweil recurre de nuevo a esta ley en su último libro, *How to create a Mind* de 2012 si bien, como observan sus críticos, a la ley le falta "necesidad nomológica" estricta: es muy dudoso su carácter de verdadera "ley" científica, manteniéndose como mucho en el campo meramente descriptivo de un hecho histórico puntual (en caso de que realmente lo sea, cf. la recensión de *How to create a Mind* por el filósofo de la mente Collin McGinn, en "Homunculism", *The New York Review of Books, March 21* (2013). Recuperado el 12-07-2013 de: www.nybooks.com/articles/archives/2013/mar/21/homunculism/).

[21] "a process of creating patterns of increasing order" (Kurzweil, p. 14).

definitiva etapa caracterizada por una "saturación de información en el universo". En la primera, dominada por la "Física y química", aparecerían los diversos elementos físicos que componen el universo, en la segunda, dominada por la biología, aparecería la vida sobre la tierra transmitida a través de la información discreta contenida en las cadenas de ADN. La tercera etapa se caracterizaría por la aparición de los "cerebros" en la evolución de la filogenia, desde los más simples hasta el cerebro humano. La aparición del cerebro humano nos introduciría en la edad cuarta, la época de la "tecnología", en la que el hombre dotado de un cerebro capaz de inteligencia simbólica y de extremidades con capacidad de manipulación (manos con pulgar oponible), inicia el desarrollo de la tecnología en el mundo que culminaría en nuestros días con el la aparición de las tecnologías informáticas en un proceso de sofisticación creciente. La quinta época sería la de la "fusión" de la inteligencia humana con bases biológicas naturales con la inteligencia artificial, merced a los sofisticados avances de la nanotecnología, lo que supondría un desarrollo exponencial inimaginable de las facultades mentales, introduciendo un nuevo factor en la evolución biológica, que ya no avanzaría a un ritmo meramente natural, sino dirigida artificialmente por el hombre. Tras esta quinta etapa de aparición de la "singularidad", sus logros se extenderían a todo el universo, que alcanzaría una "saturación de la materia y la energía" por la inteligencia de origen tanto biológico como cibernético, materia y energía adquirirían de este modo un "nivel óptimo de computación"[22].

La aparición de la *singularidad* dependería de tres revoluciones tecnológicas simultáneas que se solapan mutuamente a ritmo creciente: las de la genética, la nanotecnología y la robótica (la *revolución GNR*), entre las que la nanotecnología y la robótica desempeñaría el papel director[23].

La nanotecnología habría sido fundada precisamente a mediados de los años ochenta del siglo XX por uno de los discípulos de Kurzweil y Minsky: E. Drexler, quien defendió su tesis doctoral a mediados de los años ochenta (origen de su libro *Engines of Creation*), bajo la dirección de Minsky. Drexler habría diseñado teóricamente un dispositivo que haría factible la sugerencia formulada por John von Neumann en 1966 sobre un *autómata autorreplicante*, que sería capaz de reproducir tanto su soporte físico como el programa que lo dirige, gracias a un brazo robótico que tomaría las piezas necesarias de un

[22] Cf. pp. 14-21.
[23] Cf. Kurzweil, cap. 5, pp. 205 ss.

impreciso "mar de partes" ("sea of parts") en el que se hallaría inmerso[24]. Drexler diseña su artilugio teórico, el "ensamblador molecular" ("molecular assembler"), combinando las intuiciones de Von Neumann con las del premio nobel de física R. P. Feynmann, quien en 1959 había sugerido que los problemas de la química y la biología se resolverían en el futuro gracias a la manipulación de la materia "átomo a átomo"[25]. La nanotecnología ha sido en este sentido definida por Drexler de modo bastante optimista como un: "control directo y sin costes de la estructura de la materia basado en un control molécula a molécula de los productos y derivados"[26].

En la década de los años veinte del presente siglo XXI, la nanotecnología se habría desarrollado hasta tal punto que sería posible "sustituir" la información genética contenida en los cromosomas de las células humanas por una réplica robótica creada por nanoingeniería, réplica que sería capaz de "mantener el código genético y simular las acciones del ARN, los ribosomas, y otros elementos de la computadora en el ensamblador de la biología [la célula natural]"[27], convirtiendo de este modo en realidad el sueño del *cyborg* perfecto. La "fusión" del hombre con la tecnología ya habría comenzado en cierto sentido, en la medida en que llevamos artilugios informáticos en los bolsillos de nuestra ropa de forma habitual, hacemos uso de máquinas "inteligentes", habitamos edificios "inteligentes", etc., todos ellos ejemplos de los éxitos ciertos de la Inteligencia Artificial[28]. La nanotecnología permitirá, entre otras cosas, la introducción de nanorobots en el flujo sanguíneo y la mejora consiguiente de todas las funciones biológicas relacionadas[29]. Los "nanorobots" podrán incluso coordinarse entre sí para dar reproducir en el mundo físico las "cualidades mórficas" ("morphing qualities") que deseasen (reproducir artificialmente el

---

[24] El trabajo de Von Neumann se publicó como *Theory of Self-Reproducing Automata,* A.W. Burks, ed., Urbana: University of Illinois Press, 1966 (cf. Kurzweil, p. 228, y p. 558, nota 72).

[25] "The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom", en una comunicación presentada al encuentro anual de la Sociedad Americana de Física de ese año con el título "Thre's Plenty of Room at the Bottom", disponible en http://www.zyvex.com/nanotech/feynman.html (cf. Kurzweil, 2005, p. 227 y p. 558, nota 71). [25] Cf. una descripción puramente hipotética de los subsistemas del "ensamblador" según Drexler en Kurzweil, pp. 228-9.

[26] "thorough, inexpensive control of the structrure of matter based on molecule by molecule control of products and byproducts" (E. Drexler, C. Peterson, *The Nanotechnology Revolution,* New York, W. Morrow, 1991, cit. en Kurzweil, p. 558, nota 70). Tras *Engines of Creation,* de 1986, Drexler ha publicado *Nanosystems: Molecular Machinery, Manufacturing, and Computation,* Wiley INterscience, 1992.

[27] Kurzweil, pp. 232-6. Las revistas sobre *nanotecnología* se han multiplicado por lo demás en el último decenio, cabe citar, entre otras, la mensual *Nature Nanotechnology*, del grupo *Nature,* publicada desde octubre de 2006, el trimestral *Journal of Nanotechnology in Engineering and Medicine* de la ASME (American Society of Mechanical Engineers), publicado desde 2010 o la mensual *Nano Letters,* de la American Chemical Society, publicada desde 2001.

[28] Cf. pp. 226 ss.

[29] Kurzweil, p. 252.

cuerpo de un perro, por ejemplo, o al menos su forma exterior). Las máquinas tendrían con el tiempo "capacidades emocionales" como los seres humanos, y la realidad virtual imitará tan perfectamente la "real" que pasaremos mucho tiempo sumergidos en ella. Al final la inteligencia no biológica "saturará" la tierra (es decir, "utilizará los diseños de la energía y la materia para alcanzar una computación óptima") en primer lugar, por último "el universo se verá saturado por nuestra inteligencia", y cuando "el universo se vuelva inteligente, si la velocidad de la luz admite excepciones (o "rodeos"), "la vasta inteligencia de nuestra civilización será capaz de explotar" (!)[30]. Como consecuencia, las máquinas del futuro "serán humanas", aunque carezcan de sustrato biológico[31], lo que no impedirá que la civilización siga siendo en lo esencial "humana", entendido el término "más allá de sus orígenes biológicos".

Kurzweil predice que este estado de cosas (ciertamente *singular*), se alcanzará hacia el 2045 (ni antes ni después, si hay que fiarse de sus genuinos poderes predictivos –ignoramos si computacionales o no-)[32].

Última estación de la "singularidad": la inmortalidad

La última consecuencia del advenimiento de la *Singularidad* sería la prolongación indefinida, y a la postre, ilimitada, de la vida humana (la inmortalidad en este mundo, en una palabra), gracias por una parte a la prolongación constante de la esperanza de vida debida a los avances de la medicina convencional, y por otra, gracias a la transducción de toda la información personal contenida en el cerebro desde el mismo cerebro a otro soporte vital o *hardware* más duradero, de tipo biológico o físico[33].

---

[30] "the vast intelligence of our civilization at this future time will be able to exploit" (Kurzweil, 29, cf. la definición cumulativa que Kurzweil proporciona en pp. 25-29).

[31] "future machines will be human even if they are not biological" (ibid., p. 30).

[32] "I set the date for the Singularity –representing a profound and disruptive transformation in human capability- as 2045. The nonbiological intelligence created in that year will be one billion times more powerfuk than all human intelligence today" (ibid., p. 136).

[33] Kurzweil, pp. 323-6. La idea de la "grabación" de las facultades mentales de los hombres en un disquete implantable en cualquier tipo de *hardware* biológico o físico es muy querida para los totalistas cibernéticos. Para Minsky: "en el futuro, se conservará en disquete la personalidad de los que mueran; eso equivale a la inmortalidad, y con el añadido de que se podrán hacer tantas copias como se quiera" (entrevista en el periódico *El País* de 4 de diciembre de 1991, cit. en Ruiz de la Peña, p. 105, nota 76). Kurzweil lo

Kurzweil se ha tomado tan en serio la posibilidad de una inmortalidad lograda por los avances de la medicina convencional, que ha llegado a redactar en colaboración con un médico norteamericano un par de obras de auto-ayuda médica, con una serie de prontuarios dietético-sanitarios que serían capaces de prolongar la propia vida un lapso de tiempo lo suficientemente grande como para poder experimentar la, según él, "inmortalidad que viene"[34]. Al final de un diálogo imaginado entre el lector, los dos autores del libro y los dos "avatares" de ambos autores, que participan en la conversación desde un hipotético 2023 ("Ray2023" y "Terry2023"), los dos autores (hablando desde el presente, 2009) pretenden proporcionar al lector un prontuario de consejos que permitirán al lector "vivir para siempre"[35].

Otro de los miembros de este club de defensores de la "inmortalidad tecnoinducida" es el nanotecnólogo Robert A. Freitas, Jr., que sugiere además que se podría especificar el estadio de la vida en que el sujeto podría detener su desarrollo (en la juventud, en la adolescencia o en la edad adulta, por ejemplo)[36]. Para Kurzweil se trata a la postre tan sólo de una cuestión de tiempo[37].

---

considera factible para finales de la década de 2030 (Kurzweil, pp. 198 ss., "Uploading the Human Brain"), lo que permitiría la existencia de "humanos biológicos" y "humanos no-biológicos" (ibid., p. 199).

[34] Cf. la introducción a su libro en colaboración con el médico T. Grossmann, *Trascend. Nine steps to living well forever,* Rodale, New York, 2011 (1ª ed. de 2009), donde expone con toda seriedad los pasos necesarios para hacer realidad su lema de "vida saludable": "live well enough to live for ever". Los autores hablan de "tres puentes" que conducirían a una vida humana virtualmente eviterna en este mundo ("three steps towards radical life extension", p. xx). El "primer puente" (*Bridge one*) consistiría en el desarrollo exponencial en los últimos años del conocimiento biológico, lo que remite a la vida humana hacia una frontera indefinida e inestable por principio, una "frontera moviente" (p. xx). El "puente segundo", sostenido por el anterior, conduciría al "florecimiento de la revolución biotecnológica", florecimiento que tendría lugar "en veinte años" (es decir, hacia 2029, de acuerdo con la fecha de publicación del libro), y que de acuerdo con Kurzweil, permitirá "perfeccionar nuestra propia biología al ser capaces de reprogramar completamente sus procesos de información", lo que nos hará posible "cambiar nuestros genes en orden a vivir bien por más décadas de lo que ahora consideramos una vida larga" (p. xx); puente biotecnológico que nos permitiría alcanzar el tercero y definitivo con el "florecimiento completo de la revolución nanotecnológica", gracias a la cual "podemos ir más allá de las limitaciones de la biología y vivir indefinidamente" ("Bridge Three –the full flowering of the nanotechnology revolution- where we can beyond the limitations of biology and live indefinitely" (p. xx). El libro resume uno anterior corredactado por los mismos Kurzweil y Grossman: *Fantastic Voyage: Live Long Enough to Live Forever.* New York, Plume, 2005.

[35] "We'll tell you what you need to do so that you can live enough to live forever!" (p. xxiii), "If you stay on the cutting edge of our rapidly expanding knowledge, you can indeed *live long enough to live forever*" (p. xxv). La idea que subyace expresamente es que, dado que la esperanza de vida se prolonga unos diez años cada diez años, basta con llegar a la próxima "meta volante" para prolongar otra decena de años la propia vida la vida y así sucesivamente..

[36] "Death is an Outrage!", comunicación presentada la "Fifth Alcor Conference on Extreme Life Extension" de noviembre de 2002 (el texto puede leerse en http://rfreitas.com/Nano/DeathIsAnOutrage.htm, cit. en Kurzweil, 2005, 568. Se podría "limpiar" el cuerpo anualmente de sus partes estropeadas y mantenerlo en un estadio óptimo de funcionamiento indefinido (cf. Kurzweil, p. 253).

[37] Cf. Kurzweil, pp. 212 ss..

De este modo los avances convergentes de la bio-gene-nanotecnología permitirían llevar a la práctica el sueño declarado del futurista italiano Marinetti (1878-1944), un siglo más o menos después de su proclamación literaria:

"Nosotros venceremos la hostilidad aparentemente irreductible que separa nuestra carne del metal de los motores. Nosotros prepararemos la creación del hombre mecánico de partes cambiables. Nosotros lo liberaremos de la idea de muerte, de la misma muerte, suprema definición de la inteligencia lógica"[38].

3. *Consideraciones antropológicas sobre el proyecto de inmortalidad tecnoinducida de R. Kurzweil y los "totalistas cibernéticos":*

El panorama escatólogico que Minsky, Kurzweil, Moravec, Drexler *et al.* delinean en sus obras, ha conseguido cierta difusión en determinados ambientes[39]. Por lo demás sus propuestas son demasiado extremas y "totalistas" como para no dejar indiferentes a un lector reflexivo, pues no deja de resultar *prima facie* inquietante o desazonador para buena parte de los mortales, que se había hecho felizmente a la idea de no sobrepasar con mucho el límite de unos ochenta años bien llevados en condiciones físicas y mentales más o menos razonables.

La ideología (o "religión", si se prefiere), pues es evidente que no se trata de una teoría científica, del llamado *totalismo cibernético*, ha sido objeto de diversas críticas y objeciones antropológicas que pasamos a resumir continuación[40]. Se trata de una serie de

---

[38] Marinetti, *Manifiesto técnico de la literatura futurista*, 1912 (cit. en J. Ballesteros Llompart, "Más allá de la eugenesia: el posthumanismo como negación del *homo patiens*", *Cuadernos de Bioética* XXIII, 2012/1ª, 16 (cf. del mismo autor su ensayo crítico introductorio al posthumanismo: "Biotecnología, biopolítica y posthumanismo", en J. Ballesteros, E. Fernández (eds.), *Biotecnología y posthumanismo,* Thomson-Aranzadi, Pamplona, 2007, pp. 21-46, cf. en nota 26 de las pp. 31-32 otras *boutades* maquinistas de Marinetti: en *El hombre multiplicado y el reino de la máquina,* 1915, proponía la sustitución del amor sexual por el amor a la máquina, y en su novela *Mafarka el futurista,* de 1910, narraba la historia de un rey africano que engendra un hijo "sin necesidad de ningún sucio útero femenino").

[39] A lo que ha contribuido además la edición de una revista especializada en robótica y que se ha convertido en el órgano de difusión escrito de este universo cibernético en Estados Unidos, la revista *Wired,* fundada en 1993 (existe una edición digital en www.wired.com, la revista dispone en este momento de edición impresa inglesa e italiana).

[40] En realidad: "El campo de la Inteligencia Artificial está lleno de intelectuales optimistas que aman las abstracciones potentes y se esfuerzan por desarrollar formalismos que lo abarquen todo" (R.C. Shank y R. P. Abelson, *Guiones, planes, metas y entendimiento,* Paidós, Barcelona, 1987, p. 17 -orig. de 1977-, cit. en De Andrés, p. 41).

objeciones de tipo teórico en primer lugar, y de tipo moral, en segundo, que abordamos en este orden.

A) *Problemas técnicos y matemáticos*

En primer lugar, y desde un punto de vista puramente técnico, no está claro que la tecnología informática avance a la velocidad que Kurzweil y otros "totalistas" suponen. Expertos en IA del más alto nivel cuestionan la fe en un progreso exponencial de las capacidades informáticas. Jaron Lanier, por ejemplo, insiste en que la "ley de Moore" es aplicable al crecimiento de la potencia del hardware, no al crecimiento del software, que parece encontrar serias dificultades intrínsecas para mejorar sus rendimientos, como el hecho de que duplicar la extensión de un problema informático supone que "se necesite mucho más que el doble de tiempo para resolverlo"[41]. Por otra parte, los *softwares* diseñados para fines heterogéneos son muy difícilmente compatibles entre sí, y representa un esfuerzo sobrehumano para el programador llegar a hacerlos compatibles[42]. Por otra parte, al ser diseñados los programas de *software* para un largo plazo (por las enormes inversiones que requieren), es difícil modificarlos una vez que se han generalizado, fenómeno que Larion denomina de "anclaje" o "*lock in*"[43] .

Otra dificultad, en este caso de principio, que impedirá por sistema que un computador pueda reproducir las facultades de un cerebro humano, es que la conducta humana no es en principio, dada su complejidad y la multitud de factores que influyen en ella, formalizable en un modelo matemático cuyas proposiciones puedan hallarse de un modo recursivo por métodos algorítmicos (el único método de que son capaces las

---

[41] Lanier, 2007, p. 313.

[42] "Cuanto mayor es el tamaño de un programa de software, más probabilidades tiene de estar bajo el dominio de alguna forma de código heredado y más atroz es el tiempo de procesamiento requerido para resolver los innumerables casos de incompatibilidad que … surgen entre bloques de software originariamente creados en diferentes contextos" (Larion, 2007, pp. 314-5, 2011, pp. 19-20).

[43] "conseguir que un ordenador ejecute determinadas tareas de considerable complejidad de un modo fiable pero que admita modificaciones, sin que se cuelgue y sin que peligre su seguridad, es esencialmente imposible" (Lanier, 2007, 322). De la misma opinión es Bill Joly (creador del programa *Java* para Internet) en un artículo de *Wired*: "Having struggled my entire career to build reliable software systems, it seems to me more than likely that this future will not work out as well as some people may imagine. My personal experience suggests we tend to overestimate our design abilities" ("Why the future doesn´t need us", *Wired,* Abril, 2000, en http://www.wired.com/wired/archive/8.04/joy.html , p. 4). Cf. igualmente Russell y Norvig, p. 963.

máquinas)[44]. La dificultad se hace *esencialmente* insuperable si tenemos en cuenta además el teorema de incompletitud de K. Gödel, formulado en 1931, y que echó por tierra definitivamente las aspiraciones de una matemática totalmente axiomatizada tal y como la había propuesto, entre otros, el matemático D. Hilbert (es decir, una matemática en la que todos sus enunciados pudiesen deducirse a partir de unos pocos axiomas tomados por verdaderos y unas reglas de deducción conocidas aplicadas a los axiomas o a los teoremas previamente deducidos de los axiomas). Gödel demostró con su teorema de incompletitud que un sistema axiomático para la aritmética elemental no puede ser a la vez consistente, axiomatizable y completo, es decir: que para determinados enunciados de dicha aritmética no se podía decidir si eran verdaderos o no dentro del sistema, habiendo de recurrir para averiguarlo a criterios externos al mismo o *metamatemáticos*. Ahora bien, la aritmética elemental está incluida en casi todas las teorías matemáticas importantes, como la teoría de conjuntos, el análisis matemático o el cálculo vectorial, por lo que la mayor parte de las matemáticas serían en el mismo sentido incompletas[45]. Pero, si en el terreno de las ciencias consideradas "exactas" por excelencia, es imposible un proceso algorítmico o computacional que permita por un proceso mecánico la obtención de algunos de sus enunciados, con mucha más razón será imposible intentar encontrar modelos algorítmicos que predigan el comportamiento humano y puedan reproducirlo artificialmente, teniendo en cuenta, la imposibilidad intrínseca en primer lugar de que acabamos de hablar, y en segundo, el hecho de que el comportamiento humano, dada la infinidad de variables de las que depende (constitución biológica única, experiencias personales, estéticas, morales o religiosas únicas, hechos biográficos únicos, incluyendo las circunstancias históricas variables, etc.), lo hace por principio inmune a cualquier tipo de modelización matemática. Ahora bien, cualquier máquina o computadora funciona, digámoslo una vez más, de acuerdo con algoritmos y sólo con ellos, por lo que le resulta imposible reproducir comportamientos no definibles algorítmicamente.

---

[44] Recordemos que un algoritmo es un "procedimiento de computación o manipulación de signos sometidos a reglas automáticas, explícitas y precisas" (cf. voz "Algoritmo", en J. Mosterín y Roberto Torretti, *Diccionario de lógica y filosofía de la ciencia*, Alianza Editorial, Madrid, 2010², pp. 33-34).

[45] Cf. Mosterín y Torretti, pp. 571-573. Agradezco a la profesora Elena Postigo haberme recordado en la discusión de la presente ponencia el punto recién expuesto.

B) *El problema del dualismo: la vida y el pensamiento*

Otra seria dificultad que presentan las teorías de la I.A. en sentido fuerte, es que recaen en un dualismo radical (mucho peor que el denostado *dualismo cartesiano*, que Descartes al menos rechazaba teóricamente), dualismo para el que el que la interacción entre el cuerpo humano, el cerebro (él mismo un órgano biológico y parte del cuerpo), y las facultades mentales es absolutamente contingente, y por tanto prescindible, hasta el punto de que algunos de ellos llegan a considerarlo una carga molesta[46]. Con razón se les ha comparado con el viejo gnosticismo de los primeros siglos del cristianismo[47].

Sin embargo, dicho dualismo es una pura quimera e imposible de acuerdo con los datos de la biología. El neurólogo de origen portugués A. Damasio ha señalado reiteradamente que la inteligencia especulativa o teórica no funciona correctamente si la inteligencia emocional está dañada, pues no sabe captar adecuadamente su relación con el mundo ni establecer los fines en que debe emplearse: para que el sujeto se sirva correctamente de su inteligencia especulativa, debe disponer de una sana inteligencia emocional, encargada de "encender" la conducta en una u otra dirección relevante para el sujeto[48]. Damasio relata en este sentido el caso de dos pacientes con sus facultades cognitivas aparentemente inalteradas, de las que no hacían un uso adecuado por tener deterioradas sus capacidades emocionales, y que exhibían, llamativamente, comportamientos maquinales:

"Rígidos y perseverantes en su enfoque de la vida, ambos [dos pacientes norteamericanos de los años 40 del siglo XX que habían sufrido lesiones en sus lóbulos frontales] eran incapaces de

---

[46] Para el profesor inglés de cibernética de la universidad de Reading Kevin Warwick (el "capitán Cyber"), entre cuyas proezas se encuentra la implantación de un chip en la muñeca con el que manejó desde Nueva York un brazo robótico en Inglaterra: "El cuerpo humano ya no es necesario. Si pudiéramos deshacernos de él viviríamos mucho más" (p. 48 de la edición de *ABC* del 17-04-2013). Se trata de una actitud común entre los totalistas cibernéticos: "Es fácil imaginar al pensamiento humano liberado del cautiverio de un cuerpo mortal", decía H. Moravec en 1988 en *Mind Children: The Future of Robot and Human Intelligence* (cit. en David F. Noble, *La religión de la tecnología,* Paidós, Barcelona, 1999, p. 198); y Minsky se refería al cuerpo como "aquel maldito revoltijo de materia orgánica" y como un "teleoperador del cerebro" (cit. en Noble, p. 192). Cf. igualmente R. Penrose, *La nueva mente del emperador,* Grijalbo-Mondadori, Barcelona, 1995, pp. 45-6, y Ruiz de la Peña.

[47] Parece que fue precisamente contra el gnosticismo contra el que se introdujo en el Credo romano antiguo la expresión "creo en la Resurrección de la carne" (cf. J.A. Sayés. *Más allá de la muerte,* San Pablo Ediciones, Madrid, 1996 p. 128, que remite a J.D.N. Kelly, *Early Christian Creeds,* Londres, 1952, pp.163-5). Agradezco al profesor J. Pérez Adán la sugerencia de la similitud entre el *totalismo cibernético* y el gnosticismo antiguo.

[48] Cf. A. Damasio, *El error de Descartes. La emoción, la razón y el cerebro humano.* Crítica, Barcelona, 2006 cap. 4, pp. 74 ss. y F. Mora Teruel, *Cómo funciona el cerebro,* Alianza Editorial, Madrid, 2009, pp. 108-110.

organizar su actividad futura …; carecían de originalidad y creatividad … ; mostraban maneras generalmente correctas, pero estereotipadas; eran menos capaces de reaccionar al dolor; sus impulsos sexuales y exploradores se hallaban disminuidos …"[49].

Parece que (cabe mencionar aquí también el caso de los niños autistas), a menor vida emocional, mayor estereotipia y rigidez en el comportamiento. Pero precisamente las máquinas no tienen ningún tipo de emoción (por mucho que pudieran llegar a imitarlas con perfección): para llegar a sentir una emoción (de dolor, de hambre, de sueño) se necesita, además de cerebro, un cuerpo que experimente necesidades biológicas y a través de complejos mecanismos neurofisiológicos transmita sus estados carenciales al cerebro. Pero los computadores no tienen cuerpos biológicos, por eso actúan *maquinalmente* … .

Por otra parte, los dispositivos del cerebro implicados en el procesamiento de la vida emocional (el llamado *sistema límbico:* tálamo, hipotálamo, amígdalas …), pertenecen al llamado "cerebro antiguo", es decir, al más antiguo filogenéticamente y que compartimos con nuestros antecesores animales, y en este sentido al cerebro más "biológico" y menos racional, si cabe hablar así[50].

Una máquina necesitaría ser un organismo vivo al menos tan complejo como el humano para llegar a experimentar una vida emocional tan compleja como la humana. Pero en ese caso ya no sería una "máquina", sino un ser vivo.[51]

C) *La cuestión de la conciencia y los* qualia*:*

La tercera gran objeción, directamente relacionada con la anterior, que se ha hecho a los partidarios del *totalismo cibernético*, es que están lejos de haber demostrado que las máquinas puedan llegar a tener conciencia y, como consecuencia, conducta libre y responsable.

---

[49] Ibid., p. 80, sin embargo: "mostraban una carencia de defectos motores, sensoriales o de comunicación y una inteligencia global dentro de las expectativas, dado su entorno sociocultural" (loc. cit.).

[50] Cf. a modo de introducción F. Mora, 2009, pp. 94 ss.

[51] Cf. en Mora, pp. 32-34, otros argumentos contra el modelo computacional del cerebro, entre ellos: su plasticidad permanente (varía incluso día a día), su carácter único en cada caso, de acuerdo con su sustrato biológico (no hay "dos cerebros iguales" como no hay "dos personas iguales"), y en consonancia con las experiencias biográficas únicas de cada sujeto, a diferencia del mecanismo físico idéntico y *en serie* de cualquier tipo de ordenadores, o el hecho de que el cerebro no interprete la información de una forma unilateral y definida como un ordenador, sino de forma muchas veces ambigua y confusa.

La religión cibernética de Kurzweil reposa toda ella efectivamente en la presuposición de que las máquinas "llegarán" realmente a ser conscientes. Kurzweil resuelve el carácter francamente contraintuitivo de semejante pretensión recurriendo al famoso "test de Turing", tal y como el famoso matemático inglés y padre de la informática lo propuso en su conocido artículo de 1950. El "test" (también llamado "equivalencia de Turing"), consistía, como es sabido, en lo siguiente: si las capacidades o habilidades de una máquina son indiscernibles de las de un ser humano, habrá que concluir que dicha máquina tiene conciencia. Por tanto, si las máquinas del futuro son capaces de imitar a la perfección una conducta humana inteligente y "subjetiva" (y a Kurzweil no le cabe ninguna duda de ello), ello querrá decir que son conscientes[52].

Sin embargo no es en absoluto claro que una máquina que replicase a la perfección la conducta humana hubiese por ello mismo de tener conciencia y ser "humana" en ese sentido. Kurzweil debe enfrentarse aquí al conocido argumento de la "caja china" del filósofo norteamericano del lenguaje y de la mente John Searle (con quien reconoce haber discutido personalmente la cuestión)[53]. Searle ha señalado en diversas ocasiones la falta de pertinencia del "test de Turing"[54] con el experimento mental de la "caja" o "habitación china": imaginemos a un hombre encerrado en una habitación al que se le entregan por un orificio de una de las paredes símbolos o frases escritas en chino (lengua que desconoce), y se le dan las complejísimas órdenes necesarias para que asocie con sentido esos signos con otros signos chinos, que extrae por un orificio de la pared opuesta, de modo que un chino-parlante en el exterior pueda comprenderlos. Si el hombre encerrado en la habitación fuese capaz de hacerlo a la velocidad suficiente, el chino-parlante del exterior creería que dentro de la habitación estaría encerrado otro chino-parlante, un hombre que realmente entiende chino, pues no podría diferenciarlo de uno real. Y sin embargo no es el caso. Exactamente eso es lo que sucede con cualquier programa computacional, que "no sabe" lo que hace, por mucho que lo haga tan bien como un hombre (incluso mejor en determinadas tareas:

---

[52] Kurzweil, 385, 458 y *passim*. El famoso artículo de Turing "¿Puede pensar una máquina?" (original inglés: "Computing Machinery and Intelligence", *Mind,* vol. 59, 1950, n. 236), en A.M. Turing, H. Putnam, D. Davidson, *Mentes y máquinas,* Tecnos, Madrid, 1985, 17-60. En el apartado 6 del artículo ("Opiniones contrarias a la cuestión principal"), Turing ya se enfrentó objeción de la falta de conciencia en la máquina ("Argumento de la conciencia", pp. 37-40), recurriendo al argumento falaz del *solipsismo*: del mismo modo que no podemos estar ciertos de que otros hombres tengan conciencia, a no ser que fuésemos ellos mismos, tampoco podemos excluir por principio la posibilidad de que las máquinas tengan conciencia si su comportamiento no se pudiese distinguir en nada del de un hombre.

[53] Cf. pp. 458 ss..

[54] Primero en *Mentes, cerebros y ciencia*, Cátedra, 1985 (orig. inglés de 1980), y más recientemente en *El redescubrimiento de la mente*, Barcelona, Crítica, 1996.

por ejemplo la famosa victoria al ajedrez del ordenador *Deep Blue* sobre Kasparov en 1997). Dicho en palabras de Searle:

"usted se comporta exactamente como si entendiese chino, pero a pesar de todo usted no entiende ni palabra de chino. Pero si pasar un programa de computador apropiado para entender chino no es suficiente para proporcionarle a *usted* comprensión del chino, entonces no es suficiente para proporcionar a *cualquier otro computador digital* comprensión del chino. … . Todo lo que el computador tiene, como usted tiene también, es un programa formal para manipular símbolos chinos no interpretados. Para repetirlo: un computador tiene una sintaxis, pero no una semántica. … . Comprender un lenguaje, o ciertamente, tener estados mentales, incluye algo más que tener un puñado de símbolos formales. Incluye tener una interpretación o un significado agregado a esos símbolos"[55].

Por lo demás, Kurzweil declara la cuestión de la intencionalidad de la mente (o de otros seres vivos) resuelta de suyo suponiendo, de acuerdo con su "informática prometedora", que se llegará a encontrar la solución, si no ahora, al menos en el futuro. El caso que nos ocupa sería una instancia de esa fe incondicional: si la conciencia es una propiedad "emergente" de un sistema biológico complejo como es el cerebro, ¿por qué no podría ser una propiedad "emergente" de un sistema cibernético que fuese tan complejo como el mismo cerebro?[56] Pero esto no es naturalmente todavía ninguna prueba sino una mera suposición, fundada en otra suposición, a saber: que en el futuro se podrán diseñar sistemas "no biológicos" tan complejos en sus prestaciones como el cerebro, o bien *cyborgs* compuestos de una mezcla de elementos biológicos y nanorobóticos que igualarían el poder del cerebro. Pero el hecho es que hasta la fecha ello es imposible, y el acto de fe y la profecía de Kurzweil (una de las muchas en que abunda su libro), no es todavía ni ciencia ni conocimiento, sino una muestra más de su, parafraseando a Popper, "cibernética prometedora", de que hacen gala los "totalistas cibernéticos" de modo habitual.

En realidad, el problema de la conciencia es uno de los problemas más complejos existentes para la filosofía de la mente, incluso "irresoluble" para algunos de sus

[55] Searle, 1985, pp. 38-39, cf. su propia exposición del experimento a partir de la p. 37. Un ejemplo parecido sería el de una ginecóloga experta que nunca ha estado ella misma embarazada: podrá describir a nivel teórico de manera magistral todo el proceso del embarazo, pero de ningún modo explicar exactamente *qué se siente* al estar embarazada (cf. W. Bechtel, *Filosofía de la mente. Una panorámica para la ciencia cognitiva,* Tecnos, Madrid, 1991, pp. 169-176).

[56] Cf. Kurzweil, p. 468.

cultivadores[57], y los diversos intentos materialistas de darle solución no han dado hasta la fecha un resultado[58], por lo que no deja de resultar bastante absurdo pretender que en el futuro podrán crearse máquinas que puedan ser conscientes mientras no sepamos cómo la conciencia surge a partir del cerebro (y de un cerebro biológico de un ser *vivo*).

Además, es imposible una conciencia del mundo (de un no-yo), sin algún tipo de "sí mismo" o de "yo" que sale de sí mismo e interactúa con el mundo en la búsqueda de medios vitales para seguir vivo, que es justamente el rasgo característico de todos los seres vivientes naturales[59]. Es en el choque de las necesidades del viviente con lo que no es él, el mundo, de donde surge la conciencia de sí.[60]

De hecho, desde el punto de vista evolutivo, hay que pensar que el cerebro humano surgió precisamente sólo como un medio más sofisticado para satisfacer esas necesidades biológicas.

Por ello no se entiende cómo un ser que no está vivo (y por tanto sin ningún tipo de necesidad biológica que satisfacer) podría llegar a tener alguna conciencia de sí. Más inconcebible todavía sería pretender *insuflar* vida a un ser inerte, máxime teniendo en cuenta que a día de hoy no sabe aún *qué* es la vida o *cómo pudo surgir* de la materia inerte, teniéndose que limitar los biólogos en este caso también a una "biología prometedora"[61].

---

[57] Según C. McGinn: "nuestras facultades cognoscitivas no están capacitadas para entender el problema mente-cuerpo, el problema de cómo la conciencia, la interioridad puede hallarse en la materia" (cit. Carlos J. Moya, *Filosofía de la mente*, Universitat de València, 2004, p. 232).

[58] Cf. J. Arana, "¿Y qué es una máquina? Consideraciones críticas sobre las teorías materialistas de la conciencia", en J. Rodríguez Valls, C. Diosdado, J. Arana (eds.) *Asalto a lo mental. Neurociencias, conciencia y libertad,* Biblitoeca Nueva, 2011, pp. 13-44. Cf. un intento de crítica a Searle (a nuestro entender fallido, pues no enfrenta el núcleo de la cuestión) en D.C. Dennett, *La actitud intencional*, Gedisa, Barcelona, 1991, pp. 286-297.

[59] Tal y como decía Max Scheler: "presentan los *seres* vivos otro carácter, para ellos esencial y que es el hecho de que no sólo son objetos para observadores externos, sino que poseen además un *ser para sí*, un *ser íntimo*, en el cual *se hacen* íntimos consigo mismos" (M. Scheler, *El puesto del hombre en el cosmos,* Losada, Buenos Aires, 1968[7], p. 27).

[60] "La subjetividad sólo puede hacerse cargo de sí propia en tanto que se hace cargo de otra realidad distinta de ella" (A. Millán-Puelles, *La estructura de la subjetividad,* Rialp, Madrid, 1967, p. 173, cit. en De Andrés, p. 251).

[61] Por ejemplo la conocida bióloga y genetista Lynn Margulis (1938-2011), se limitaba (como otros muchos) a confiar al paso del tiempo la solución del surgimiento de la vida: "Aún hay un eslabón perdido entre las más complejos invenciones de los científicos y la más sencilla célula viable, tanto en teoría como en la práctica de laboratorio. La distancia entre las substancias químicas de pequeño tamaño molecular, como aminoácidos y nucleótidos, y la de mayor tamaño y origen biológico, como el RNA y las proteínas [indispensables para la química de los organismos vivos], es enorme. Pero unos cientos de millones de años de actividad molecular suponen una extensión de tiempo muy larga. Los científicos sólo han trabajado durante unas décadas para lograr en el laboratorio las condiciones que conducen a la vida pero han llegado muy lejos. Es posible que antes de que acabe el siglo XX una célula viva sea generada espontáneamente en el tubo de ensayo. Si se consideran millones de años, las probabilidades de formación espontánea de hiperciclos

D) *Una nueva autocomprensión del hombre como consecuencia de la prolongación indefinida de la vida humana*

Dejando de lado toda la carga de ciencia-ficción y de *cibernética prometedora* que contienen las obras de Kurzweil y los *totalistas cibernéticos*, no deja de ser cierto que la esperanza de vida de los seres humanos crece de forma indefinida. Según algunos biólogos podría extenderse en no mucho tiempo en un tercio (hasta 140 años), teniendo en cuenta los recientes descubrimientos del funcionamiento de la telomerasa en los ratones, que ha logrado prolongar la vida de estos roedores en un tercio respecto a la media anterior.

Una prolongación indefinida de la esperanza de vida humana, más aún una vida humana potencialmente eviterna plantearía (está planteando ya de hecho), una serie de desafíos sociológicos, sanitarios, demográficos, antropológicos en un sentido amplio de muy de diversa índole que exigen y exigirán en el futuro una reflexión detenida, que tenga en cuenta las necesidades de las generaciones presentes y futuras (problemas que por cierto *posthumanistas* y *totalistas cibernéticos* tienden a ignorar de forma sistemática).

Salta a la vista en primer lugar el problema de sobrepoblación o superpoblación que semejantes pronósticos anuncian de forma inevitable, problema que cibernéticos y posthumanistas consideran soluble de suyo gracias a su fe inquebrantable (y ciega) en el progreso tecno-científico. Sofisticados artefactos *nanotecnológicos* con un gasto mínimo de materia prima lograrían fuentes de energía "sin coste" ("*inexpensive energy*") supuestamente inagotables (el cómo no se explica)[62]. Un *posthumanista* como Nick Bostrom, haciendo gala de una ingenuidad verdaderamente irresponsable, confía sin base alguna en el descubrimiento hipotético (y no justificado) de soluciones[63].

---

fueron inmensamente mayores que las que tienen los investigadores, los cuales han de substituir la ciega perseverancia del tiempo por la planificación, si quieren llegar a recrear la vida" (L. Margulis y D. Sagan *Microcosmos. Cuatro mil millones de años de evolución desde nuestros ancestros microbianos.* Barcelona, Tusquets, 1995, p. 73, en el cap. 2: "La animación de la materia"; por cierto que llama la atención que en el contexto de una explicación que se quiere científica, donde se supone un tipo de causalidad rigurosamente determinista, se tenga que recurrir a la "generación espontánea" –la causalidad del *porque sí,* diríamos- para intentar explicar un hecho natural).

[62] Kurzweil, 2005, p. 457 (también habrá "células solares construidas por nanotecnología", ingenios voladores de uso individual para transporte de mensajes o personas, etc. cf. ibid., pp. 457-8).

[63] "Otro problema que *podríamos* enfrentar es la sobrepoblación. *Tal vez* la gente deberá aprender a tener menos hijos, y a tenerlos a una edad más avanzada. *Tal vez* se descubran formas de sostener una mayor población mediante el uso de tecnologías más eficientes. *Tal vez* se desarrollen naves espaciales y comience la colonización del cosmos" (en su ingeniosa –e ingenua- "Fábula del dragón-tirano", en

Se han señalado asimismo los potenciales conflictos sociales que las expectativas de una prolongación indefinida de a vida podrían acarrear caso de no estar distribuidas equitativamente entre todos los sectores de la población. Según Lanier, si la hipotética "inmortalidad" individual llegase a ser de algún modo posible, o bien sería muy cara y estaría restringida a las pocas personas que pudieran sufragarla, o bien sería muy barata y todo el mundo tendría acceso a ella, lo que supondría que las técnicas de control y manipulación biológica serían fácilmente accesibles para todo el mundo, también para los que estuviesen dispuestos a hacer un uso indeseado de ellas, (terrorista, pongamos por caso), lo que tampoco resulta especialmente tranquilizador[64].

Fukuyama ha señalado otras serias objeciones al proyecto de inmortalidad tecno-inducida[65]. Al hecho del evidente envejecimiento de la población, con la inversión completa de la pirámide de población, que supondría el aumento abrumador de la "vejez dependiente", habría que añadir la ruptura de las relaciones intergeneracionales tradicionales, pues las generaciones jóvenes emergentes verían vedado su acceso a los niveles de toma de decisiones por mucho tiempo, niveles que serían ocupados indefinidamente por las personas de mayor edad y experiencia, etc.[66].

Desde el punto de vista antropológico, para terminar, la perspectiva de una vida potencialmente inmortal (a no ser por accidente), afectaría sobre todo a la autocomprensión de la vida humana, es decir: a la interpretación de la vida humana (el "mapa" de la propia vida), sin la que el hombre, a diferencia de cualquier otro ser natural, no puede llevar a cabo su existencia[67]. A diferencia del animal, cuya dotación biológica instintiva le resulta suficiente para sobrevivir, la carencia de verdaderos instintos en el hombre le exige en cambio mayor o menor medida una dirección consciente de su

---

http://www.transhumanism.org/index.php/WTA/more/842/, traducción del original "The Fable of the Dragon-Tyrant", en *Journal of Medical Ethics,* 2005, 31, 5 273-7, cursivas nuestras).

[64] Cf. Lanier, 2007, p. 323. El mismo Kurzweil dedica un capítulo a esta cuestión ("The deeply intertwined Promise and Peril of GNR", pp. 391 ss.), y reconoce abiertamente que los avances tecnológicos que pronostica son una "espada de doble filo" (p. 424) que deberían regularse severamente. Algunos especialistas en IA han llegado sencillamente a proponer el abandono de estas investigaciones, al menos de las potencialmente más peligrosas (como Bill Joly en el artículo mencionado de *Wired*).

[65] Cf. el cap. 4 de F. Fukuyama, *El fin del hombre. Consecuencias de la revolución biotecnológica*, Ediciones B, Barcelona, 2002 pp. 101 ss..

[66] Cf. *ibid.,* pp. 112 ss..

[67] Frente a la "indiferencia abrumadora" (N. Hartmann) para su vida efectiva del resto de los entes respecto a la idea que se pueda tener de ellos, el hombre es una "creación de sí por sí mismo" (Bergson), que necesita por ello mismo de un modelo que le muestre el camino que ha de ir recorriendo para "llegar a ser el que es" (cf. M. Landmann, *Antropología filosófica*, Uteha, México, 1961, p. 5).

existencia[68]. Por ello la vida humana es en buena medida la imagen o el modelo que nos hagamos de ella y de acuerdo con la cual la modelamos día a día (y ciertamente no es lo mismo ver reflejada la propia imagen en un Dios personal, amante y misericordioso, que verla reflejada en un animal … o en una máquina).

Pero la autocomprensión de la vida depende en muy buena medida de la comprensión de la propia muerte y de su proximidad o lejanía relativas, como vio Dilthey y Heidegger elaboró detenidamente en *Ser y Tiempo*[69].

Los moralistas clásicos eran conscientes de que la certeza de una muerte cercana, y en cualquier caso cierta, le daba a la vida el espesor y la "seriedad" necesarias para no dilapidarla inútilmente. Así Séneca, en su conocida carta CI a Lucilio *Sobre la muerte de Seneción,* podía recomendar a su amigo:

> "que dispongamos de nuestra alma como si hubiésemos llegado a nuestro fin. No aplacemos nada: saldemos cada día nuestras cuentas con la vida. El mayor mal de ésta es que siempre es incompleta, que siempre nos reservamos una parte para el futuro. Aquel que cada día sabe dar a su vida el último golpe de mano, no necesita tiempo"[70].

El *memento mori et non peccabis* ("recuerda que vas a morir y no pecarás") no era sino la versión cristiana de la misma recomendación.

Sin embargo, una vida concebida como potencialmente inmortal, introduciría por el contrario una visión banal e irresponsable de la vida, una infantilización de la existencia, una inautenticidad permanente si se quiere, por decirlo en términos heideggerianos, pues el

---

[68] Como ya vio el poeta F. Schiller: "En los animales y en las plantas la naturaleza no da solamente el destino, sino que ella sola lo realiza también. Pero al hombre le da sólo su destino, y le deja que lo realice él mismo" (F. Schiller, *Über Anmut und Würde, Sobre la gracia y la dignidad*, 1793, cit. en A. Gehlen*, El hombre. Su naturaleza y su posición en el mundo*, Sígueme, Salamanca, 1980, -orig. alemán de 1940, 1966-, p. 36).

[69] "la relación, en fin, que determina de un modo más profundo y más general el sentimiento que tenemos de nuestra existencia, es la relación de la vida a la muerte; pues la limitación de nuestra existencia por la muerte es siempre decisiva para nuestra comprensión y nuestra estimación de la vida" (W. Dilthey, *Das Erlebnis und die Dichtung*, 5ª ed., p. 230, cit. en M. Heidegger, *Ser y Tiempo*, traducc. de J. Gaos, FCE, México, 1962, p. 272).

[70] L.A. Séneca, *Cartas morales a Lucilio,* Traducc. de J. Bofill y Ferro, Orbis, Barcelona, 1985, vol. II, p. 124. El siguiente pasaje sobre el carácter ajeno o "impropio" de la vivencia de la muerte, que de modo habitual nunca se experimenta como una posibilidad propia, sino ajena, preludia las posteriores reflexiones de Heidegger sobre la vivencia "impropia" de la muerte: "y, mientras llevamos la muerte a nuestro lado, en la cual sólo pensamos en la persona de los demás, de donde proviene que los ejemplos de mortalidad, que tan a menudo podemos observar, no nos causen más que una primera sorpresa. Y ¿qué cosa más necia que admirarse de algo que cada día puede suceder?" (ibid., p. 124).

hombre dejaría de ser el *ser-para-la-muerte* de que hablaba Heidegger, el ser que tiene que enfrentarse de forma consciente, de modo auténtico o inauténtico, con la posibilidad de dejar de existir (dejar de *ser-ahí* en el mundo), para convertirse en el ser-sin-la-muerte, el que tiene todo el tiempo del mundo y por tanto se lo toma, el que podría dejar para mañana siempre lo que pudiese hacer hoy, porque nunca faltaría el mañana, el infante-adolescente-joven perpetuo, que nunca maduraría porque no tendría tiempo para madurar porque le faltaría la ocasión para ello.

No está claro *a priori* que una vida de este tenor (al menos en lo que hace a su duración) fuese más deseable que la que un hombre occidental contemporáneo con una esperanza de vida de unos ochenta años de vida y un nivel medio de vida aceptable.

Conclusión: Una ética para el poder de la biotecnología

El poder creciente y omnímodo de la técnica sobre los diversos aspectos de la vida humana, incluyendo su propia naturaleza biológica esencial (en el mero sentido biológico, no metafísico, de la expresión, en la medida en que la existencia de *especies* biológicas, cuyos miembros son mutuamente fecundos entre sí, es un dato biológico), que se creía hasta hace muy poco tiempo un dato biológico indisponible, sólo modificable por los lentos procesos de la evolución natural, condiciona el futuro de la humanidad en un grado tal que se hace necesario más que nunca poner en práctica una ética de la responsabilidad que tenga en cuenta todas las múltiples consecuencias a largo plazo de cada una de las decisiones que se tomen en el campo de la biotecnología o del *biopoder*. Realmente: "Las exigencias a la responsabilidad crecen proporcionalmente a los actos del *poder*"[71].

Pero el poder de la biotecnología crece, si no de modo exponencial, sí de un modo continuo, y los límites a la investigación deberán imponerse desde fuera de ella, pues la ciencia o la técnica no sabrían qué decir en cuanto tales sobre cuestiones morales. Habrá pues que: "poner el galope tecnológico bajo control extratecnológico"[72].

---

[71] H. Jonas, *Técnica, medicina y ética. La práctica del principio de responsabilidad,* Paidós, Barcelona, 1997 (orig. alemán de 1985), p. 35.

[72] Ibid., p. 39.

Algunos han sugerido establecer como límite infranqueable al *biopoder* sobre el hombre un imperativo moral *ecológico* de conservación del ser humano como especie biológica única (del mismo modo que se considera deseable y casi moralmente exigible la lucha por la conservación de las especies en peligro de extinción), para evitar su posible destrucción a través de la manipulación biotecnológica[73].

Quizás fuese exigible también, como sucede en el caso de los enfermos llamados terminales, garantizar una proporcionalidad entre los medios empleados para prolongar la vida de toda la humanidad mediante la genética, la nanotecnología, la biotecnología en general, y las consecuencias y riesgos indeseables probables, o incluso meramente posibles, que para la vida de todos los hombres (también de las futuras generaciones) implicarían dichas intervenciones.

Por otra parte, ¿cómo compaginar, desde el punto de vista de una moral meramente natural, o cristiana, el imperativo moral de conservar la propia vida tanto como sea posible –una vida cada vez más longeva-, con el imperativo de no agotar los recursos naturales disponibles, limitados en principio, para garantizar su disfrute a las futuras generaciones?

Los problemas son complejos, y probablemente no lo sean menos las soluciones, que deberán resolver complejas ecuaciones en que biólogos, médicos, economistas, sociólogos, ingenieros y moralistas (posiblemente también el ciudadano de a pie), tendrán una palabra que decir. En cualquier caso convendrá no perder nunca de vista, frente a cualquier posible partidario de una *inmortalidad tecnoinducida*, las palabras con las que Séneca terminaba la carta recién citada y que la sabiduría popular se encarga de refrendar a su modo: "Lo que importa no es que vivas mucho, sino que vivas bien; y a menudo vivir bien consiste en no vivir mucho"[74].

---

[73] Cf. G. J. Annas, L. B. Andrews y R. M. Isasi, "Protecting the endanreged human: toward an international treaty prohibiting cloning and inheritable alterations", *American Journal of Law and Medicine,* (2002), vol. 26, pp. 151-178 (cit. en R. Andorno, *Bioética y dignidad de la persona,* Tecnos, Madrid, 2012, p. 154 y nota 18).

[74] Séneca, p. 125.