La teoría de la mente: de la inteligencia artificial a la inteligencia híbrida

[Publicado en Concepción Diosdado, Francisco Rodríguez Valls y Juan Arana, Eds. *Neurofilosofía. Perspectivas contemporáneas,* Thémata / Plaza y Valdés, Sevilla / Madrid, 2010, pp. 153-176)

Resumen: La filosofía de la mente ha dejado de ser, hace ya tiempo, si es que lo fue alguna vez, un campo exclusivo para los filósofos. En el pensamiento contemporáneo se ha producido un creciente interés por estas cuestiones desde el punto de vista de la tecnología. El texto somete a crítica las ideas de Ray Kurzweil y hace un somero repaso de algunas tendencias recientes en el abordaje de estas cuestiones.

Abstract: Philosophy of Mind no longer is -if it ever was at all- an exclusive field for philosophers. Contemporary thinking has known an increasing interest over these topics, from a technology point of view. This text reviews Ray Kurzweil ideas, providing at the same time a brief summary of the latter trends in approaching Philosophy of Mind matters.

> Una de las señas de identidad de la naturaleza humana es que sitúa a los hombres ante problemas que son demasiado difíciles para ellos, sin que les quede la opción de dejarlos sin abordar en razón de esa dificultad.
> **Peter Sloterdijk,** (2000, 85)

Entre las metáforas más habituales para hablar de la mente se encuentra, sin duda alguna, la del computador, o, más generalmente, la de la máquina. Eso nos indica, inmediatamente, que, al menos en el ámbito de la psicología popular, damos por hecho que pensar es una acción y que, como todas las acciones, el pensamiento tiene un ejecutor que dispone de un instrumento o de una herramienta; incidentalmente habrá que advertir que si consideramos el pensamiento como una acción, ello se debe a que, en cierto modo, el pensar transcurre en el tiempo, esto es a que somos conscientes de que el tiempo pasa y nosotros con él. Aunque ahora no insistiré en esta observación, que me parece esencial, más adelante habrá que intentar sacarle algún provecho.

La imagen de la máquina permite una serie de analogías que han podido parecer útiles y estimulantes para entender lo que generalmente se llama la mente, especialmente una vez que, a partir de la modernidad, se abandonó la idea aristotélica de alma como forma del cuerpo. Aunque en sus orígenes cartesianos la mente fuese primordialmente una mente consciente enteramente ajena a la materia, muy pronto se abrió paso la idea de que la sustancia pensante debiera ser un producto más de la máquina del cuerpo, y, también relativamente pronto, se empezaron a sugerir analogías presuntamente iluminadoras de su función. Haya sido de ello lo que fuere, lo que me importa subrayar es que, en la filosofía de la mente que ha predominado desde mediados del pasado siglo, se ha supuesto, muy frecuentemente, que la mente es un determinado tipo de aparato, un aparato que media extrañamente entre una realidad objetiva pero, de algún modo, inasible, y una realidad subjetiva, pero que se acaba objetivando de una serie de formas o modos tales como el lenguaje, la lógica o la ciencia. Así, de una manera un tanto subrepticia, el *problema de la conciencia*, bastante intratable, se pudo traducir, de algún modo, en una especie de *problema de la inteligencia*, para asumir a renglón seguido que, si

pudiéramos hacer una máquina capaz de pensar como nosotros parecemos hacerlo, no habría ninguna objeción de fondo a que dijésemos, no solo que esa máquina piensa, sino que posee tanta conciencia como se pueda necesitar para andar por este mundo. Es decir, para sortear la evidente elusividad de la conciencia se acudió, como era lógico, a una elusión de su estudio. Se pudo hablar entonces de la inteligencia artificial y del problema mente-máquina; la gran ventaja de este planteamiento fue que los científicos obtuviesen un programa de trabajo suficientemente amplio como poder olvidar las razones que pudieran impulsarle a *perder el tiempo* con enigmas *metafísicos*. Si, al fin y al cabo, se podría decir, la cosmología escolástica pereció definitivamente a manos del telescopio, lo razonable era esperar que el *espiritualismo*, por llamarlo de algún modo, acabase pereciendo a manos de la informática. La filosofía del lenguaje se ocuparía, por otra parte, de limpiar los restos mortales del *fantasma en la máquina*, más o menos lo que ha pretendido hacer el llamado *materialismo eliminativo* y, en general, todas aquellas doctrinas que han pensado que un tratamiento verbal intensivo[1] podría apartar del horizonte lo que se ha llamado el *problema difícil*[2], además de otros que tienden a parecer más fáciles de lo que son[3].

De cualquier manera, en este, digamos, programa, hubo un olvido que casi se podría considerar freudiano; se pasó de forma inadvertida sobre un aspecto de la cuestión realmente peculiar, una carencia que ha subrayado recientemente Roger C. Shank[4], a saber, que difícilmente podríamos imitar aquello que nosotros hacemos, si no sabemos cómo lo hacemos. Porque, en efecto, es mucho olvidar que no sepamos, como no sabemos en absoluto, qué y cómo es lo que hacemos, cuando pensamos, ignorancia que no es demasiado específica porque tampoco sabemos en absoluto, aunque muchos puedan creer lo contrario, que es lo qué hacemos cuando hablamos, y cómo lo hacemos, o qué hacemos y cómo lo hacemos al mover el corazón, o al respirar, aunque de esto último, justo es reconocerlo, sepamos, al parecer, algo más. Olvidando este aspecto, sin duda decisivo, de la cuestión, lo que el programa mente máquina hacía, por decirlo de

---

[1] Algo parecido cabría decir de quienes se empeñan en encontrar modelos distintos de aproximación, como si el problema se pudiese reducir a un defecto en la lógica del planteamiento. Puede pertenecer a este género la sugerencia del filósofo californiano Alva Noë, (http://socrates.berkeley.edu/~noe/) en, por otra parte, una muy interesante entrevista en Edge, (que puede leerse en http://www.edge.org/3rd_culture/noe08/noe08_index.html) de que la mente no es algo que ocurra en el cerebro, puesto que no está en ningún *interior*, sino que es algo que hacemos.

[2] Me parece que el primero en emplear esta terminología ha sido David Chalmers (1995) y (1996).

[3] Putnam (2001, 16), advierte, por ejemplo, de que "muchos filósofos quieren descartar los problemas tradicionales en la filosofía de la percepción como si ya se hubera empleado demasiado tiempo en ellos y como si ahora ya estuvieran superados"

[4] "We can speak properly without knowing how we do it. We don't know how we comprehend. We just do. All this poses a problem for AI. How can we imitate what humans are doing when humans don't know what they are doing when they do it? This conundrum led to a major failure in AI, expert systems, that relied upon rules that were supposed to characterize expert knowledge. But, the major characteristic of experts is that they get faster when they know more, while more rules made systems slower". http://www.edge.org/q2008/q08_7.html

alguna manera, era imitar *el producto* de nuestra acción, no *la manera de producirlo*.

No hay buenas razones que aconsejen arrumbar lo que pudiéramos llamar un *dualismo primario* o, si se prefiere, *ingenuo*; no sirve de mucho gastar saliva para cambiar los nombres de las cosas, a ver si el problema se disimula o desaparece, ni hay razones suficientemente poderosas que impulsen a aceptar una u otra forma de fisicalismo, físico, biológico o digital, como algo inevitable. Sin embargo, nunca he tenido dificultad para admitir que el dualismo tampoco resulta ser una solución satisfactoria: es sólo la mejor manera de denominar una cuestión, o una familia de cuestiones, que están en la frontera de nuestra imagen del mundo, el problema que Schopenhauer bautizó, brillantemente, como *nudo del mundo*, pero nunca menos.

1. Inteligencia artificial: la esperanza en la tecnología como factor *cultural*, con un breve *ex cursus* orteguiano

Cuando todo hacía entender que el fracaso de las profecías de los primeros gurús de la IA, había apartado de la agenda la cuestión sobre el pensamiento y las máquinas, uno de los grandes creadores de software del momento ha venido a resucitarla con sus propuestas, que no solo renuevan las promesas de lo que se llamó IA en sentido fuerte, sino que las complementa, las concreta y las dota de un halo de misticismo bastante peculiar. Me refiero a Ray Kurzweil[5] quien, además de ser uno de los gurús más conocidos de Sylicon Valley (fue uno de los creadores de lo que se llaman programas OCR y de los programas capaces de leer textos en voz alta), es, además, un convencido de la posibilidad inmediata de lograr una cuasi-inmortalidad, personal y corporal (especialmente, a base de medicación); pues bien, nuestro hombre ha agitado el panorama con sus publicaciones, conferencias e, incluso, con una película documental, sobre lo que él ha llamado *máquinas espirituales* y sobre la inmediatez con la que va a producirse la síntesis perfecta entre mentes y máquinas.

Kurzweil (1999:123) se inspira explícitamente en el proyecto genoma humano, iniciado en 1991, para sugerir que, de la misma manera que, al menos en teoría, se ha podido *mapear* enteramente el genoma del hombre, se podrá *mapear* igualmente, sinapsis a sinapsis, el cerebro humano. De este modo tendríamos a nuestra disposición muy diversas posibilidades de gran interés, por ejemplo, podríamos guardar nuestra memoria personal en archivos más sólidos, e inmunes al fenómeno bien conocido de modificación del recuerdo, podríamos *copiar* nuestras mentes, agregarlas, borrarlas, se supone que de modo parcial, es decir hacer todo aquello que podemos llevar a cabo de manera casi completamente rutinaria con nuestros archivos informáticos.

[5] Puede verse su muy completa página web en http://www.kurzweiltech.com/aboutray.html

Las connotaciones del trabajo de Kurzweil nos aconsejan colocar la obra de Kurzweil en un panorama más amplio que el de la teoría de la mente. Si nos salimos del ámbito, un tanto estrecho, de la filosofía de la mente, las propuestas de Kurzweil tal vez debieran verse como una contribución a la tendencia que ha adquirido cierta notoriedad en el ámbito de la medicina, en concreto, lo que se ha llamado "medicina para el perfeccionamiento" (*enhancement medicine*, o *enhancement technology*) que Juengst (1998:29) ha definido como el conjunto de "intervenciones destinadas a mejorar la forma o el funcionamiento humanos más allá de lo que es preciso para restablecer o mantener la buena salud". El entusiasmo tecnológico no es un fenómeno nuevo, pero los avances en los dominios llamados NBIC (*nanotechnology, biotechnology, information technology, cognitive sciences*) le han dado fuelle a esta clase de propuestas, de modo que la medicina para el perfeccionamiento podría muy bien verse superada por lo que ya se ha llamado *medicina transhumana* o *medicina para el perfeccionamiento transhumano* (Wolbring, 2005). Para los defensores de esta nueva medicina, las cualidades físicas y mentales del ser humano son indefinidamente perfectibles. El baremo de la *normalidad* ya no debiera buscarse en la media empírica, sino en el estado "transhumano" alcanzado merced al empleo de tecnologías específicas[6]. Desde este punto de vista, las ideas de Ray Kurzweil, conforme a las cuales no está lejano el momento en que la humanidad trascienda sus limitaciones biológicas y llegue a una simbiosis con máquinas que, a su vez, podrían considerarse *espirituales*, supondrían la abolición de cualquier diferencia relevante entre naturaleza, humanidad y tecnología.

Lo que de esta manera se pone en juego es una cuestión muy difusa y compleja, a saber, dónde están y cómo se fijan los límites de lo humano, pregunta que permite ligar de una manera muy inmediata los campos propios de la filosofía de la tecnología y de la filosofía de la mente. Más en general, creo que Putnam (2001:85) acertó plenamente al escribir que "no tiene verdadero sentido la asignación precisa de los problemas filosóficos a diferentes "campos" filosóficos. Suponer que la filosofía se divide en compartimentos estancos etiquetados como "filosofía de la mente", "filosofía del lenguaje", "epistemología", "teoría de los valores" o "metafísica" es una manera segura de perder todo el sentido sobre cómo están conectados los problemas, lo que significa perder toda comprensión de las fuentes de nuestras dificultades".

En cualquier caso, no quisiera dejar de destacar que ha sido precisamente Ortega y Gasset uno de los autores que primero se han ocupado de la importancia de la técnica para el sentido de la vida humana y, a mi modo de ver, de manera muy brillante. Ortega supo ver en la técnica, al menos, dos dimensiones realmente profundas e interesantes que abordó con clarividencia, aunque con la dispersión que le fue característica. Por un lado, le parecía que la técnica podía desempeñar el papel de gendarme del espíritu, sometiendo a la imaginación y a la *literatura*, su aliada más frecuente, a un régimen de ascesis y de contención. Pero, en segundo término, supo ver que la técnica suponía una capacidad de invención esencial al hombre mismo, y no, meramente, una elección. En el análisis orteguiano, estas dos dimensiones de la técnica se

[6] Junto con José Luis Puerta, me he ocupado más específicamente del aspecto médico de estas cuestiones en González Quirós (2009 c).

encuentran en un punto de equilibrio relativamente inestable porque en ambas juega un papel esencial la fantasía: en la primera, como fuerza que hay que contener; en la segunda, como clave maestra del significado de la técnica.

Ortega (1996: 114) escribió que el hombre, lejos de vivir sobre la tierra, vive, en realidad, sobre unas creencias, sobre una filosofía, pero de tal modo que, pese a ello, el hombre vive en un estado de esencial insatisfacción, de inadaptación, tal vez porque es demasiado consciente de las limitaciones de sus creencias, de manera que el hombre empieza por construir su manera de estar en el mundo; el modo de habitar propio del hombre no puede reducirse a una especie de destino natural, sino que es un hallazgo, una fórmula: tal fue la idea que expuso en su conferencia a los arquitectos alemanes en Darmstadt, una pieza brillante de la que, sin embargo, Ortega se quejó con alguna amargura. Allí, (1996:107) dijo literalmente: "El hombre es, esencialmente, un insatisfecho, y esto –*la insatisfacción*- es lo más alto que el hombre posee; precisamente porque trata de tener cosas que no ha tenido nunca. Por eso suelo decir que esta insatisfacción es como un amor sin amada o como un dolor que siento en unos miembros que nunca he tenido". Se puede discutir esta caracterización del hombre mismo, pero no cabe muchas dudas de que lo que Ortega afirma, explica, en cierto modo, tanto la conducta del *hombre masa*, como el comportamiento del filósofo más exigente y egregio, habitualmente insatisfecho con cualquier teoría.

Ortega ya había utilizado ideas muy similares en *La rebelión de las masas*, mostrando cómo el hombre masa experimenta la ausencia de criterios y de coerción como una invitación a vivir a su modo, a imponerse (1962, 180), "Si la impresión tradicional decía: Vivir es sentirse limitado, y, por lo mismo, tener que contar con lo que nos limita, la voz novísima grita: Vivir es no encontrar limitación alguna, por lo tanto, abandonarse tranquilamente a sí mismo. Prácticamente nada es imposible, nada es peligroso y, en principio, nadie es superior a nadie". Este análisis implica la constatación de un importante cambio histórico en el que la civilización técnica ha tenido un papel predominante; en el pasado, vivir significaba para el hombre medio dificultades, peligros, privaciones de todo tipo y ello traía consigo la necesidad de someterse a la ley, de interiorizar profundamente el respeto a un amplio conjunto de normas. Pero el mundo actual se nos presenta como un espacio —al menos en apariencia— más seguro y con más abundancia de cosas, donde no estamos tan obligados a acatar las normas sociales, donde casi nadie nos discute el derecho a vivir según nuestras normas y saciar nuestros deseos, aunque seguramente ocurra que la pasión de reglar, inseparable de cualquier poder, se haya trasladado a otras regiones tratando de disimular su presencia para confundirnos mejor. Esta impresión general que todo individuo tiene sobre el mundo que le toca vivir, termina por convertirse en una invitación a desear, a perder el miedo a lo imposible.

Ortega fue, por supuesto, consciente de la trascendencia que tenía la técnica en el destino humano y de los riesgos que podía plantear. Por eso afirmó (1996: 55) que "acaso la enfermedad básica de nuestro tiempo sea una crisis de los deseos y por eso toda la fabulosa potencialidad de nuestra técnica parece como si no nos sirviese de nada". Sin embargo, Ortega supo ver con gran nitidez que la técnica no es meramente la adaptación del hombre al mundo sino, con mayor

propiedad la creación de un mundo nuevo, porque, según él, el hombre se encuentra inadaptado en la naturaleza y quiere un mundo que sea suyo.

A la muerte de Ortega aún no se podían adivinar casi ninguno de los avances que hoy consideramos decisivos a la hora de poner en cuestión las relaciones del hombre con su entorno, para valorar los efectos de la tecnología sobre la sociedad humana, pero nuestro filósofo supo ver cómo se dibujaba un nuevo y paradójico horizonte de confrontación entre la naturaleza y la técnica cuando decía (1996:108): "la victoria de la técnica quiere crear un mundo nuevo para nosotros, porque el mundo originario no nos va, porque en él hemos enfermado. El nuevo mundo de la técnica es, por tanto, como un gigantesco aparato ortopédico... y toda técnica tiene esta maravillosa y —como todo en el hombre— dramática tendencia y cualidad: la de ser una fabulosa y grande ortopedia".

Volviendo a Kurzweil, tenemos que colocar su promesa de que nos espera un acrecimiento inédito de nuestras capacidades intelectuales, si acertamos a convivir con esas máquinas espirituales de las que nos habla, en un contexto, como el de ahora mismo, en que, por ejemplo, sabemos que hay piernas ortopédicas que permiten correr más velozmente que las piernas naturales. Pues bien, a Kurzweil le parece que lo que la tecnología puede hacer con el cuerpo podrá hacerse también con la mente, un mejoramiento casi indefinido de su capacidad y su fiabilidad, una reforma de las evidentes carencias que ofrece, a la luz de las capacidades que nuestro autoconocimiento nos permite soñar.

La afirmación de Kurzweil supone que, en el campo de las mentes y las máquinas, de los cerebros biológicos y los computadores, la *ortopedia orteguiana* se va a ver ampliamente superada por una síntesis que supondrá una auténtica fusión y, con ello, el inicio de una nueva etapa evolutiva para el género humano; su posición, que, independientemente de otros juicios que pudiera merecer, no parece observar ninguna clase de dificultades ajenas a la tecnología, nos obliga a preguntarnos si está fundada en algo más que en una promesa deliberadamente vaga, en una extrapolación dudosamente legítima.

A día de hoy, sabemos bien que el proyecto genoma humano no ha cumplido ni siquiera un mínimo porcentaje de las esperanzas que algunos de sus promotores pretendían suscitar, y, aunque no sea buena política argumentar con el fracaso de lo que puede dejar de serlo[7], hay que preguntarse si existe algún fundamento sólido en el que puedan apoyarse las promesas kurzweilianas. Kurzweil sostiene que el poder de las ideas para transformar la realidad se está acelerando y propone una teoría a la que llama "ley de los retornos acelerados" para explicar de qué manera, según él, la tecnología y el proceso evolutivo que ha de suponer

[7] Es muy frecuente reírse de las profecías del pasado, cuando no se han cumplido, pero no sé si es muy inteligente; es interesante anotar que eso sucede tanto cuando no ocurren cosas que se supone debieran haber ocurrido, como cuando pasan cosas muy importantes que nadie supo predecir. Para el primer caso, puede ser interesante repasar el listado de profecías establecido por el recientemente desaparecido Arthur C. Clarke (1999: 536 y ss.). Sin que corresponda exactamente con el segundo caso, creo que es muy interesante echar un vistazo a la relación de cambios (aunque referidos solo al sector de las industrias de la cultura) acaecidos en los últimos cuarenta años que ha hecho el editor Mike Shaztin (http://www.idealog.com/stay-ahead-of-the-shift-what-publishers-can-do-to-flourish-in-a-community-centric-web-world). Que algo no hay sucedido hasta ahora no siempre es razón suficiente para que nunca vaya a suceder.

se comporta como una función exponencial (2005:3). Al poder comprender nuestra propia manera de comprender, al poder convertir en objeto a nuestra inteligencia, obteniendo su *source code*, podremos revisarla y expandirla de manera completamente nueva. La vida humana se verá transformada de forma irreversible (2005:7), podremos librarnos de muchas fatalidades, tendremos la inmortalidad muy a la mano, y, al final del siglo XXI (2205:30), la porción no biológica de nuestra inteligencia será, dice literalmente, trillones de veces superior a la mera inteligencia humana sin ayudas externas.

Quien quiera encontrar razones precisas de afirmaciones tan estupefacientes no deberá perder el tiempo buscándolas en las obras de Kurzweil, al menos yo no he sabido encontrarlas. Nuestro autor se apoya, sobre todo en afirmaciones tales como que el progreso tecnológico siempre ha sido exponencial, o que la escasa (¿¡!?) diferencia genómica entre el chimpancé y el ser humano, que según Kurzweil (2005:5) es de unos pocos centenares de *bytes*, no nos ha impedido la creación de maravillas tecnológicas. Y, por supuesto, asume, sin atisbo de duda, la convicción de que la idea computacional de la mente es enteramente correcta[8].

Aunque se hayan escrito ríos de tinta[9] sobre lo que se quiere decir exactamente con teoría computacional de la mente, lo esencial es suponer que existe una analogía básica e iluminadora entre las relaciones del *hardware* y el *software* y las relaciones entre cerebro (o *wetware*, como a veces se dice) y mente. En el fondo, el modelo computacional de la mente, como Carver (2007:101) ha señalado, encarna y refuerza el análisis funcionalista, suponiendo que la caja negra del modelo funcionalista, es un computador.

A su vez, podemos ver el funcionalismo, en cierto modo, como una *consecuencia* del análisis conductista de lo mental, un expediente bastante heroico para evitar las dificultades y paradojas de la idea intuitiva de mente consciente. No es nada que algunos filósofos audaces, como Dennett, por ejemplo, no se hayan atrevido a hacer con tal de poder librarse de la amenaza del dualismo, así, Dennett (1995: 441) escribió. "si lo que usted es, es el programa que corre en el ordenador de su cerebro [...] en principio, usted podría sobrevivir a la muerte de su cuerpo tan intacto como un programa que puede sobrevivir a la destrucción del ordenador en el que fue creado por primera vez"[10]. Kurzweil va supuestamente más allá de

---

[8] Lo que no siempre se comparte por los investigadores del área. Aduciré aquí el testimonio reciente de Noel Sharkley: "Roboticist Hans Moravec says that computer processing speed will eventually overtake that of the human brain and make them our superiors. The inventor Ray Kurzweil says humans will merge with machines and live forever by 2045. To me these are just fairy tales. I don't see any sign of it happening. These ideas are based on the assumption that intelligence is computational. It might be, and equally it might not be. My work is on immediate problems in AI, and there is no evidence that machines will ever overtake us or gain sentience" Puede consultarse en http://www.newscientist.com/article/mg20327231.100-why-ai-is-a-dangerous-dream.html?full=true.

[9] La discusión reciente sobre el particular entre Ray Tallis e Igor Aleksander (2008) puede verse en http://www.palgrave-journals.com/jit/journal/v23/n1/full/2000128a.html.

[10] No sé si Dennett habrá caído en la cuenta de que esta formulación podría tomarse, aunque solo en cierto modo, como una puesta al día de la idea de alma como forma del cuerpo, y prestarse admirablemente a la idea, religiosa en este caso, de que el alma bien podría sobrevivir

esta afirmación denettiana, aunque, como se ve, no sea el primero que lo ha hecho, al suponer que tenemos a nuestro alcance algo más que una teoría iluminadora porque poseeremos de manera inmediata, según predice su lectura de la conocida ley de Moore, la tecnología suficientemente poderosa para hacer efectivo el poder pasar del hardware al cerebro y del software a las mentes, del mismo modo que ahora podemos, biológicamente, pasar de la mente al cerebro y viceversa, lo que supone olvidar todas las dificultades que los filósofos han aducido a propósito de esa conversión. Resulta sorprendente que se pretenda arrojar por la ventana el conjunto de problemas que los filósofos han analizado al respecto sin caer en la cuenta de lo peligrosas que resultan ciertas metáforas cuando se las quiere hacer pasar por ciencia. A propósito del funcionalismo escribió Crick (1994, 94) que resulta tan estrafalario que muchos científicos se asombran al saber que de verdad existe como teoría seria. Si suponemos que, a imagen de lo que, al parecer, sucede con las neuronas, haremos surgir la conciencia en el momento en que demos con la clave organizativa capaz de hacer pensar al silicio, estamos diciendo algo que difícilmente debería tomarse en serio. A este respecto, una de las críticas más feroces se debe a uno de los padres fundadores de la idea, a Hilary Putnam. En una de sus últimas aportaciones a esta embrollada cuestión, Putnam (2001:104) mantiene que, en la media en que no se le ha dado un significado determinado a la propiedad computacional, el funcionalismo es ciencia ficción apoyada en un equívoco, porque nadie ha conseguido, hasta la fecha, resolver un problema mal planteado, de manera que el reproche que dirige a los reduccionistas puede valer igualmente para los ilusos que, como Kurzweil, creen ver claramente en el futuro aquello que no consiguen entender ahora: decir que "algún día la ciencia podrá encontrar la manera de reducir la conciencia (o la referencia o lo que sea) a la física", aquí y ahora, es lo mismo que decir que algún día la ciencia "puede que haga no-sabemos qué de manera que no-sabemos-cómo" (2001:204).

Putnam añade algo más, que me parece especialmente interesante: el rechazo de los planteamientos reduccionistas no solamente no entraña el abandono de la investigación científica seria, sino que son esos planteamientos los que con frecuencia llevan a que los investigadores conciban mal los problemas empíricos. Putnam (2001:208) defiende la filosofía, que ha desaparecido casi por completo de los escritos de Kurzweil, quien practica un eliminativismo efectivo, y no parece caer en la cuenta de que no basta con no ser filósofo para evitar los problemas, porque, de nuevo en palabras de Putnam (2001:206), "la confusión filosófica se extiende más allá de los límites de quienes estudian filosofía, ya sea profesionalmente o como simples aficionados", y de que, aunque muchas cosas susciten nuestra admiración, "la formulación de una pregunta inteligible exige más que admiración".

---

al cuerpo, ser eterna. Tampoco veo mucha dificultad, si se me permite seguir con el juego, en que un Dios capaz de manejar toda la información (apenas un poco más informado que el genio laplaceano) pudiese darnos de nuevo nuestra mejor forma corporal para no dejarnos, por así decir, con el alma en pena.

2. **Sobre lo que seguramente olvida Kurzweil**

Al margen de las abundantes y repetidas críticas que hayan podido hacer los filósofos al llamado programa fuerte de la IA, su fracaso consistió, básicamente, en una manifiesta incapacidad para crear algo parecido a una conciencia viva, pero también en el escasísimo éxito alcanzado en imitar actividades inteligentes que los primeros gurús consideraban realmente simples. Aunque pueda resultar escasamente piadoso, no me resisto a volver a citar algunos testimonios que aduje en otro libro[11] al caracterizar lo que allí llamaba *ciberfilosofía*. En sus primeros pasos, algunos de los más conspicuos representantes de la IA proclamaron que su trabajo podía considerarse, como mínimo, el tercer gran acontecimiento de la historia de la humanidad. Marvin Minsky declaró a la revista LIFE en noviembre de 1970. "Dentro de tres a ocho años tendremos una máquina con la inteligencia general de un ser humano medio. Me refiero a una máquina que podrá leer a Shakespeare, engrasar un coche, intervenir en las politiquerías de la oficina, contar un chiste, sostener una pelea. En este punto la máquina empezará a educarse con fantástica velocidad. En unos meses habrá alcanzado el nivel de genio y, transcurridos varios meses más, su poder será incalculable". Pese a la demora que el plan llevaba ya en 1984, Roger Schank, alguien habitualmente más sobrio y comedido que Minsky, afirmaba que "Algún día habrá una máquina omnisciente. En eso estamos". Seguramente fueron afirmaciones como estas las que indujeron a David Gelertner a asegurar que la ciencia informática es un campo lleno de chiflados ávidos de novedades.

Este programa supuso, desde luego, un fracaso, pero no el abandono de las intenciones de fondo, como lo muestra la aparición de Kurzweil en el mercado. Este verano el propio Kurzweil presentó un documental sobre sus ideas en Harvard; Karim Gherab estaba allí, y me ha dicho que Kurzweil no supo sino responder vaguedades a una pregunta sobre cómo pensaba arreglárselas para copiar una mente (o un cerebro, si le parece más fácil) en un computador[12].

En el libro que se editó con un debate entre Kurzweil y algunos de sus críticos, el único filósofo presente fue Searle (2002:71-72) quien insiste en sus conocidos puntos de vista sobre la cuestión, a saber la irreductibilidad de la semántica a la sintaxis y su prueba de la "habitación china" (que, a mi modesto entender, es una reelaboración del conocido argumento leibniziano del *molino*[13], y le espeta a Kurzweil tres objeciones decisivas. En primer lugar, le reprocha dar la impresión al público de que entiende lo que realmente no entiende, feo vicio, desde luego; en segundo lugar, asumir como verdades definitivamente establecidas teorías que no lo están, y, por último, el hecho de que no tengamos un conocimiento mínimamente claro de cómo el cerebro hace lo que hace.

---

[11] Pueden verse las referencias correspondientes en González Quirós (1998), pp. 110 y 111.
[12] Además de usar abundantemente el peculiar e inválido argumento de que cómo los tecno-escépticos se habían equivocado en otras muchas ocasiones, él iba a acertar ahora.
[13] Monadología, 14.

Las críticas de Michael Denton, un bioquímico cercano a las tesis del "diseño inteligente", le reprochan a Kurzweil, con toda justicia, el olvido de algunas características esenciales de los seres vivos, que, en modo alguno, parecen existir en el terreno de las máquinas. La primera de esas diferencias es la capacidad de autoreplicación que tienen los seres vivos y a muchos niveles distintos. Se trata de una propiedad de la que, evidentemente, carecen los computadores, aunque no hayan faltado los teóricos, futuristas, eso sí, que afirmen que se les podrá dotar de funciones similares a esta propiedad de la vida que parece establecer una diferencia decisiva.

Otra propiedad aducida por Denton es la capacidad de los seres vivos para crecer y transformarse cambiando su forma y su estructura. Desde luego a todo esto se le puede llamar *información*, pero estamos muy lejos de comprender y saber hacer de verdad lo que hace una simple gallina calentando un huevo: desplegar la *información* de la yema y hacer que aparezca un polluelo, aunque el ejemplo sea mío, no de Denton.

Denton acude al Kant de la *Crítica del Juicio*[14] para recordar la peculiar implicación entre causas y efectos que es típica de las formas de vida y que no parece reducible a un análisis causal ortodoxo desde el punto de vista de la ciencia. La forma orgánica es irreductible a una reducción simple. Denton (20002, 94) afirma, por ejemplo, que "Yet no artifact has ever been built, even one consisting of only 100 components (the same number of components in a simple protein), which exhibits a reciprocal self-formative relationship between the parts. This unique property [...] is the hallmark of organic design".

Es muy característico de la peculiar audacia de los funcionalistas y de los pensadores del estilo de Kurzweil esa capacidad para saltar por encima de la vida, como si la vida fuese realmente algo simple. La vida no resulta fácil de comprender, si entendemos por comprender lo que hemos podido hacer en el ámbito de la mecánica. Sería una necedad negar que en el futuro puedan entenderse cosas que ahora no comprendemos en absoluto, pero no es menos arrogante suponer que se puede prescindir de la vida si se quiere lograr algo como la conciencia. La vida es, además, una realidad estrictamente atenida al tiempo, es temporalidad y tiene ciertos caracteres, digamos, *neguentrópicos*, por emplear el término creado por Schrödinger, que tampoco se pueden hacer de menos. No se arregla nada suponiendo que todo se puede reducir a mayor complejidad a partir de los mismos elementos simples que ya conocemos, al parecer, perfectamente. No es ningún baldón suponer que hay cosas que, hoy por hoy, se nos escapan. El fondo del error que se comete al prescindir de los caracteres de la vida que parecen ser irreductibles a la mecánica y a la informática, es la tendencia a confundir lo abstracto con lo concreto. La mente decía Smullyan (1989:197), a quien la identificación de lo abstracto con lo concreto le parecía uno de los errores filosóficos más trágicos de nuestros días, "es todo lo concreta que pueda ser una entidad". El hecho es que, hasta ahora, y más allá de las fantasías computacionales o literarias, solo la podemos suponer, con buen criterio, en los seres vivos. La vida es un fenómeno muy difícil de

[14] Kant, como se recordará, afirmó que no existiría nunca el "Newton [...] de una brizna de hierba" (Kritik der Urteilskraft, § 75).

entender y del que se supone se ocupa la Biología, pero que, como repite Emilio Cervantes[15], escapa como el agua del cesto de la ciencia. Es oportuno recordar aquí la advertencia leibniziana[16] sobre la diferencia entre las obras de Dios y las de los hombres, pero mejor pasemos a otra cosa. La mente, por su parte, es temporal e intuitiva, singularísima, casi se siente la tentación de decir que es lo único que conocemos en concreto, algo seguramente muy distinto a lo que pueda ser el más complejo y creativo de los programas.

3. **La ciencia y las propiedades de lo mental**

Desde un punto de vista epistémico, se puede decir que la ventaja principal del dualismo primario frente a cualquier forma de reduccionismo, consiste en que el problema empírico de las relaciones entre el cerebro y la conciencia tiene perfecto sentido para el dualista, mientras que es casi un sinsentido para las posiciones reduccionistas, pues quien crea que hay algo distinto al cerebro que ven los neurofisiólogos[17], el que crea en la existencia de una mente distinguible del cuerpo físico que, en todo caso, es el cerebro, sabe muy bien que hay algo más que su mera creencia o que la coherencia y el atractivo de sus ideas; sabe muy bien que todo eso se produce en un endiabladamente complejo sistema de implicaciones e influencias entre su percepción consciente y temporal y los sucesos plenamente físicos que acaecen en su cerebro. Así lo vio magistralmente Schrödinger (2006:93), "el mundo es un constructo de nuestras sensaciones, percepciones, recuerdos. Es conveniente considerarlo como si existiera objetivamente por sí mismo. Pero ciertamente no se hace manifiesto por su mera existencia, sino de un modo condicionado en relación con ciertos eventos especiales que se dan en partes muy especiales de este mundo, señaladamente en ciertos eventos que acontecen en el cerebro. Este es un tipo inhabitual de implicación que nos plantea la cuestión siguiente: ¿qué propiedades particulares distinguen a estos procesos cerebrales y los hacen capaces de producir la manifestación?"

Schrödinger, además de advertir las paradojas metafísicas del caso, traza aquí todo un programa de trabajo que, de una u otra manera, se está llevando a cabo. Explorar todo ese complejísimo universo de implicaciones es asunto de la

---

[15] Es el lema de su interesante blog: http://weblogs.madrimasd.org/biologia_pensamiento/

16 "Porque una Máquina debida al artificio humano no es Máquina en cada una de sus partes. Por ejemplo, el diente de una rueda de metal contiene partes o fragmentos que nada tienen de artificial para nosotros ni que sea específico de la máquina respecto del uso al que la rueda está destinada. En cambio, las Máquinas de la Naturaleza, esto es, los cuerpos vivientes, son aún Máquinas en sus más pequeñas partes, hasta el infinito. En esto consiste la diferencia entre la Naturaleza y el Arte, es decir, entre el arte Divino y el Nuestro1." (*Monadología*, 64, traducción de Julián Velarde).

[17] Ese del que Bertrand Russell (1976:265) decía: "He sido censurado por decir que lo que un fisiólogo ve cuando examina el cerebro de otro hombre está en su propio cerebro, y no en el de otro".

ciencia, no de la filosofía, pero no es una tarea imposible o que no tenga sentido, por difícil que parezca ahora a nuestros ojos.

Por lo que ahora sabemos, un cerebro humano medio tiene aproximadamente cien mil millones de neuronas que se conectan unas con otras mediante, aproximadamente unas 5000 sinapsis de media. El cerebro puede establecer o interrumpir, aproximadamente, un millón de conexiones por segundo, y puede mantener información utilizable por muchas décadas, *etiquetándola*, utilizando su significado en diversas relaciones, cambiando su *ubicación* o modificándola cuando lo crea necesario y, al tiempo que está haciendo todo eso, coordina el trabajo de cientos de músculos y los procesos necesarios para el funcionamiento de nuestro cuerpo de una manera enteramente inconsciente para nosotros mismos. Puede interpretar de manera correcta miles de señales y tomar las decisiones adecuadas en milisegundos. Además de todo eso, nos permite pensar, hablar, mantener relaciones y aprender. Toda esa actividad está siendo estudiada con tecnologías cada vez más sutiles, y nos está proporcionando unas ingentes cantidades de información que hay que relacionar, valorar, interpretar, y someter a una teoría coherente. Hoy conocemos relativamente bien qué partes del cerebro intervienen en la percepción, cómo trabaja el cerebro, cómo procesa las señales que recibe, cómo se formas los recuerdos o cómo se controla el movimiento de los músculos. Sabemos qué regiones se activan con el habla, cuando miramos algo o cuando hacemos cálculos sencillos, y estamos empezando a saber qué pasa cuando se toman decisiones.

Sería raro que esa tarea que han de llevar a cabo los científicos no se viese complicada por cuestiones de carácter categorial; el hecho es que los avances en esta clase de temas son mucho más lentos y escasamente significantes de lo conveniente. Hemos celebrado años, décadas y, casi, siglos del cerebro, nadie duda de la importancia de estas cuestiones y, sin embargo, por decirlo de algún modo, no se anuncian seriamente ninguna clase de cambios de paradigma, salvo en terrenos en que, en lugar de la ciencia, prime la profecía.

Algunos podrán sentir la tentación de creer que estemos ante un campo en el que, como pudiera decir un Kant redivivo, no se pueda progresar por el seguro camino de la ciencia. No es esta mi opinión, si se me permite el ejercicio de inmodestia de pronunciarme. Prefiero creer que podamos estar a las puertas de algunos avances realmente espectaculares, que dar por hecho el fin de la ciencia. Sin embargo, no creo que el tipo de avances que se puede esperar sea capaz de resolver ningún enigma metafísico, por razones muy similares a las que me hacen dudar de que podamos ser capaces de descorrer el velo de Maya o de trasladarnos fuera del espacio y del tiempo.

¿Qué es lo que cabe esperar? Como el buen *empirista* que me gustaría ser, lo que espero es que una ciencia más precisa que la actual, pero no necesariamente muy distinta a la que ahora disponemos, nos permita comprender mejor cómo trabaja el cerebro y, por tanto, nos abra la posibilidad de colaborar conscientemente con ese trabajo para mejorar las posibilidades intelectuales que estén a nuestro alcance. Ello hará posible, por ejemplo, crear nuevos instrumentos para potenciar nuestras capacidades perceptivas e, incluso, nuestra inteligencia. Creo, en suma, que el cerebro podrá contar con *exoinstrumentos* que se conecten con él de una manera bastante simple y

efectiva para mejorar su rendimiento, espero que sea posible alguna ortopedia intelectual, alguna forma de inteligencia híbrida y que, por ahí, se abrirán nuevos caminos. Me parece, además, que esa nueva forma de *inteligencia híbrida* no vendrá únicamente por el lado del *hardware*, sino también por el lado del *software*, de la muy posible mejora del sistema de signos que usamos para pensar y calcular, y de las formas de automatizar sus relaciones a través de nuevas redes externas a nosotros o, a su manera, también híbridas. Todo esto pudiera parecer ciencia-ficción, pero tal vez no estemos demasiado lejos de ello.

La ciencia está empezando a liberarse de marcos conceptuales y de imágenes que la mantenían, en cierto modo, cautiva y a poseer métodos que le permitan avances modestos, pero sólidos y continuados[18]. Para explicar mejor lo que pretendo decir recurriré a una comparación. Voy a referirme a la crítica que hace del darwinismo el bioquímico Michael J. Behe porque creo que es una metáfora adecuada de lo que intento decir. La crítica de Behe a Darwin, o, mejor, al darwinismo contemporáneo, se mueve en diversos frentes y, como es obvio, no niega que haya existido, o pueda existir, algo como la evolución biológica en diferentes aspectos del amplísimo fenómeno de la vida. Lo que pone severamente en tela de juicio es su capacidad explicativa. Para Behe, en el momento en que se ha podido abrir la *caja negra* de la biología molecular, el tipo de explicaciones a nivel de los organismos que son típicas de los argumentos darwinianos, carecen de cualquier fuerza. Como cualquier cita del modo de argumentar de Behe suficientemente expresiva sería excesivamente larga, en nota al pie[19] reproduzco un texto de Behe (1999: 37 y ss.) respecto a la

---

[18] Como reconoce Terrence Sejnowski, que se dedica a la neurobiología computacional en el *Salk Institute, "*The way that neuroscientists perform experiments is biased by their theoretical views", http://www.edge.org/q2008/q08_8.html.

[19] Para Darwin, la visión era una caja negra, pero después del esforzado trabajo acumulativo de muchos bioquímicos, ahora nos aproximamos a ciertas respuestas. Los siguientes cinco párrafos nos brindan un bosquejo bioquímico de la operación del ojo. Los extraños nombres de los componentes no deberían desalentar al lector. Son sólo etiquetas, no más esotéricas que las palabras *carburador* o *diferencial* para quien lee por primera vez un manual del automóvil. [...] Cuando la luz llega a la retina, un fotón interactúa con una molécula llamada 11-*cis*-retinal, que en picosegundos se reconfigura para ser *trans*-retinal. (Un picosegundo es el tiempo que la luz tarda en viajar a lo ancho de un cabello humano). El cambio de forma de la molécula retinal impone un cambio a la forma de la proteína, la rodopsina, a la cual el retinal está estrechamente enlazado. La metamorfosis de la proteína altera su conducta. Ahora llamada metarrodopsina II, la proteína se adhiere a otra proteína llamada transducina. Antes de chocar con la metarrodopsina II, la transducina se había enlazado con una pequeña molécula llamada GDP. Pero cuando la transducina interactúa con la metarrodopsina II, el GDP se desprende y una molécula llamada GTP se enlaza con la transducina. (La GTP está muy emparentada con la GDP, pero exhibe diferencias críticas). [...] La explicación anterior es sólo un bosquejo elemental de la bioquímica de la visión. Pero, en última instancia, es éste el nivel de explicación al cual debe apuntar la ciencia biológica. Para comprender una función, debemos comprender detalladamente cada paso importante del proceso. Los pasos importantes de los procesos biológicos se producen a nivel molecular, de modo que una explicación satisfactoria de un fenómeno biológico –tal como la vista, la digestión o la inmunidad- debe incluir su explicación molecular. Ahora que hemos abierto la caja negra de la visión, ya no basta con que una explicación evolucionista de esa facultad tenga en cuenta la estructura *anatómica* del ojo, como hizo Darwin en el siglo diecinueve (y como hacen hoy los divulgadores de la evolución). Cada uno de los pasos y estructuras anatómicos que Darwin consideraba tan simples implican procesos bioquímicos abrumadoramente complejos que no se pueden eludir con retórica. Los metafóricos saltos darwinianos de elevación en elevación ahora se revelan, en muchos casos, como saltos enormes entre máquinas cuidadosamente diseñadas, distancias que necesitarían un helicóptero para recorrerlas en un viaje.

bioquímica de la visión que muestra lo lejos que se encuentra nuestro entendimiento actual de esa función del análisis organicista. Darwin, de cualquier manera, no habría podido razonar de ningún otro modo, dado que, en la segunda mitad del XIX, la Bioquímica no existía en absoluto.

Los materialistas postcartesianos han sido materialistas no por conocer qué es lo que hace el cerebro, sino a pesar de no tener ni idea de qué y cómo lo hace. Han sido materialistas *a priori*, creyentes en la curiosa idea metafísica, si se me permite decirlo así, de que el uno es más creíble que el dos[20], de que era razonable que en el universo hubiese una única sustancia y escándalosa la pretensión de que pudiera haber dos. Es decir, a los efectos del materialismo clásico, el cerebro muy bien hubiera podido ser de madera maciza, su materialismo se originaba en algo enteramente independiente de la peculiar naturaleza biológica del cerebro. Aquí cobra sentido la metáfora de la crítica behiana al evolucionismo. No tenemos todavía todas las llaves que nos permitan conocer el funcionamiento del cerebro, pero tenemos cada día más y sería penoso que se interrumpiesen o estropeasen estupendas investigaciones por mor de una metafísica que nos hiciese creer que ya sabemos lo que son los estados mentales, a saber, productos del cerebro teórico, de un cerebro, digamos, de madera.

Soy perfectamente consciente de que la comparación que acabo de hacer, como todas, tiene graves defectos y de que el materialismo es ligeramente más sutil que la afirmación de que el cerebro bien pudiera ser un buen trozo de madera. Sin embargo, sí ha sucedido que las teorías *cajanegristas* del cerebro han dado píe a que cobrase una importancia absolutamente infundada la concepción computacional de la mente, que, como se ha dicho, se desarrolló sobre la base de las críticas conductistas y funcionalistas del dualismo cartesiano, absolutamente incomprendido, por otra parte, y que se comportaron, pudiera decirse, conforme al criterio del viejo dicho castellano de a *moro muerto, gran lanzada*.

No podemos seguir confundiendo, para terminar, el análisis lógico de los productos de la mente, o del cerebro, si se prefiere, con el verdadero

---

La bioquímica presenta pues a Darwin un reto liliputiense. La anatomía es simplemente irrelevante cuando nos preguntamos si la evolución podría ocurrir a nivel molecular. También lo es el registro fósil. Ya no importa si en el registro fósil hay enormes lagunas o si su registro posee la continuidad del registro de los presidentes de los Estados Unidos. Y si hay lagunas, tampoco importa que se puedan explicar de modo plausible. El registro fósil no puede decirnos si las interacciones entre el 11-*cis*-retinal con la rodopsina, la transducina y la fosfodiesterasa se pudieron desarrollar paso a paso. Tampoco importan los patrones biogeográficos, ni los de biología de poblaciones, ni las explicaciones tradicionales de la teoría evolucionista para los órganos rudimentarios o la abundancia de ciertas especies. Ello no significa que la mutación aleatoria sea un mito, ni que el darwinismo no logre explicar nada (explica muy bien la microevolución), ni que los fenómenos de gran escala como la genética de poblaciones no importen. Importan. Sin embargo, hasta hace poco los biólogos de la evolución podían pasar por alto los detalles moleculares de la vida porque se sabía muy poco sobre ellos. Ahora se ha abierto la caja negra de la célula, y es preciso explicar el mundo infinitesimal que nos revela.

[20] Esta es una de las objeciones que hace Sherrington (1984) al materialismo de la primera mitad del siglo XX.

conocimiento de cuáles son y cómo actúan las estructuras del cerebro que permiten nuestra actividad consciente.

La confusión de la mente con un sistema computacional ha durado demasiado tiempo, y no está extinguida, como para que no sea interesante tratar de entender los motivos que han llevado a que se haya podido mantener en píe una hipótesis tan arriscada. Creo que una de las razones es el hecho, tan olvidado como decisivo, de que, como le gustaba decir a Schrödinger, la mente siempre se ha *experimentado* en singular, lo que sugiere fuertemente una cierta unicidad de la mente que queda salvada en el modelo computacional que supone que las mentes son indistinguibles como singulares. Me parece que hay también otra razón que explica la confusión, si es que lo es, entre mente y cerebro desde esta perspectiva. Me refiero a que los fantásticos progresos tecnológicos de la era digital han supuesto una síntesis entre tecnologías reduccionistas (o inspiradas en saberes que lo son metodológicamente) y las tecnologías propiamente digitales, que *no son* reduccionistas o fisicalistas, sino que se apoyan en la capacidad de manejar propiedades semánticas, y que esa síntesis ha favorecido la confusión del significado metafísico de unas y otras. Es necesario, sin embargo, distinguir entre ambas fuentes de tecnología, y eso puede ayudar a disipar la ilusión en que se apoya el modelo computacional de la mente. El reduccionismo en el ámbito físico busca determinar los elementos que componen una realidad dada y cuya articulación explica los fenómenos; la digitalización, por el contrario, se construye a partir de un significado que está previamente dado, que no se descifra sino que se cifra digitalmente. El hecho de que ese cifrado se preste maravillosamente a una manipulación electrónica infinitamente rápida no debería confundirnos acerca de esta diferencia esencial.

Enteramente al margen de cualquier metáfora computacional, la ciencia del cerebro está en condiciones de obtener un conocimiento cada vez más preciso del funcionamiento de la biología neurológica, y cabe esperar que de este modo, se encuentren formas de implementar mejoras tecnológicas de nuestro rendimiento intelectual, eso a lo que podríamos, sin mucha duda, llamar inteligencia híbrida. No me cabe ninguna duda de que esos avances exigirán, en su momento, que afinemos el paradigma teórico que pueda dar cabida a todas esas nuevas formas de saber, pero creo que la metafísica más apropiada al caso seguirá siendo, de uno u otro modo, *dualista,* y apostaría a que ningún descubrimiento empírico derribará nunca una categoría metafísica bien fundada, como lo es la categoría de lo *mental*.

Me parece, para concluir con este análisis, que esta clase de categorías están suficientemente protegidas de cualquier amenaza reduccionista por diversas razones de principio, tales como las que vinculan la conciencia con la percepción del tiempo, una realidad que no está tan plenamente a nuestra disposición como el espacio, o las que se refieren a la peculiar libertad de que ha de gozar nuestro entendimiento, siempre que pretendamos reconocernos una capacidad de construir formas de saber que no sean absurdamente contradictorias; respecto a esta última consideración, una de sus mejores, y más breves, formulaciones es la advertencia de Epicuro, que, a mi juicio, no desmerecería en comparación con la demostración euclidiana de que el número de los números primos es infinito; me refiero a la siguiente sentencia: "El que dice que todo acontece por necesidad

nada puede objetar al que niega que todo acontece por necesidad, pues esto mismo afirma que acontece por necesidad[21]".

Así pues, estoy completamente de acuerdo con un texto muy reciente de Juan Arana (2009:302-303) "Dado que una leve tosecilla puede interrumpir mi libertad y la propia conciencia de mi mismo, ¿qué problema hay en admitir que su ejercicio este ligado a la actividad coordinada o desacompasada de estas o aquellas neuronas, o a la emisión y recepción de tal o cual neurotransmisor? En buena hora llegar el momento de que tales incógnitas sean despejadas, si es que su desvelamiento redunda en beneficio de la humanidad. Sin embargo, y dado que lo que se busca en el horizonte de las neurociencias es establecer una correlación entre la actividad neuronal y el despertar de la conciencia o el ejercicio de la voluntad, la eventual determinación, aunque sea completa, de tales correlaciones en nada equivaldría a la completa naturalización de la conciencia ni de la libertad. Lo único que quedaría desacreditado con tales avances seria un dualismo descarnado, de un tipo que ni siquiera Descartes llegó a defender. La unidad del hombre implica que las dimensiones corpóreas y anímicas que descubrimos en él están entrelazadas de un modo tan íntimo, que no sabríamos acabar de separar unas de otras ni física ni conceptualmente. Tradicionalmente se asume sin problemas que esto afecta a las nociones anímicas, psíquicas o espirituales, pero lo cierto es que afecta del mismo modo a las corpóreas o materiales. o dicho en otras palabras: tiene tan poco sentido hablar de meras neuronas, moléculas o átomos, como de mera conciencia, voluntad o libertad. Se trata en ambos casos de abstracciones, aspectos más o menos definidos de lo real que hemos separado conceptualmente, cortando artificialmente sus lazos con el todo del que forman parte, y asumiendo que en una primera aproximación podemos valernos de ellas para acercarnos al conocimiento de la verdad que buscamos."

4. **La inteligencia híbrida y la nueva mente colectiva**

Cuando se ha tratado de imitar el funcionamiento del cerebro a base de construir redes neurales, se ha incurrido en el error de suponer que el cerebro real utilizaba el mismo tipo de arquitectura, física y lógica, relativamente elemental que usa el software corriente; la investigación de las propiedades y el funcionamiento de las sinapsis, al parecer, se basa, por el contrario, en una disposición harto más compleja. El trabajo de los matemáticos y de los neurobiólogos, amén de muchos otros especialistas, está empezando a poner de manifiesto que la circuitería cerebral utiliza propiedades composicionales y físicas que, quizá sin necesidad de recurrir a la mecánica cuántica, nos van a permitir conocer mejor el funcionamiento del cerebro y, al tiempo, inspirar la forma de construir arquitecturas de computación más rápidas y mucho más potentes

[21] Exhortaciones de Epicuro (Gnomologio Vatican0, 40), según García Gual & Acosta (1974: 125)

En lo que respecta a los posibles avances en el conocimiento de la peculiar física del cerebro para soportar nuestras actividades mentales, me referiré a la historia del memristor, siguiendo las sugerencias de un *paper* muy reciente de Justin Mullins (2009). En 1971, Leon Chua[22] un joven ingeniero electrónico californiano, tenía la sensación de que la teoría electrónica vigente en torno a la naturaleza de los circuitos eléctricos era matemáticamente deficiente. De manera semejante a como Mendeléyev fue capaz de sugerir que había huecos que se tendrían que llenar en su tabla de elementos químicos, Chua pensaba que, además que los tres elementos (o dispositivos) conocidos en un circuito, el condensador, la resistencia y el inductor, tendría que existir un cuarto. Sus razones para pensarlo eran de índole estrictamente matemática, a saber que entre cuatro entidades (en este caso, la carga eléctrica, su comportamiento en el tiempo o corriente, el campo magnético que crea, y la intensidad o voltaje de ese campo) que guardan relaciones binarias deben existir seis formas básicas de relación, y en el caso de la teoría de circuitos había sólo cinco[23] lo que le parecía a Chua un defecto insoportable. Esa sexta relación debiera servir para hacer algo que no se pudiese hacer mediante combinaciones del resto y Chua, en particular, supuso que debería comportarse como una resistencia pero, además, debiera poder *recordar* qué corriente había pasado a su través, de manera que lo llamó memristor (*memory resistor*). Durante casi cuarenta años, el memristor fue una criatura estrictamente teórica, porque no existía ningún dispositivo físico capaz de efectuar esa función precisa. A comienzos del nuevo siglo se empezó a hablar de materiales y diseños que podían hacer esa función; se trataba de sistemas que funcionaban a nivel de la nanotecnología y que eran inobservables en las escalas milimétricas. Se trataba de un descubrimiento que podía ser útil para la construcción de memorias de flash o sólidas, que requieren una capacidad de escritura y reescritura mucho más rápida. Pero, para lo que a nosotros interesa, la verdadera novedad llegó cuando se conoció el modo de comportamiento realmente sorprendente de una criatura unicelular, el *Physarum polycephalum*, un hongo mucilaginoso que, al parecer, podía resolver ciertos puzles elementales y, sobre todo, parecía capaz de anticiparse a sucesos que se produjesen de una manera periódica, de manera que sus estudiosos dieron en suponer que, pese a carecer de neuronas, debiera poseer alguna especie de memoria. Max di Ventra, un físico de San Diego que conocía las ideas de Chua, tuvo conocimiento del caso y pudo compararlo con el comportamiento de un memristor y, posteriormente, en colaboración con Yuriy Pershin (2008) y otros, ha construido un modelo de memristor que, al parecer, se comporta como lo hace realmente una sinapsis.

He aducido este ejemplo porque muestra, a mi modo de ver, una de las múltiples maneras en que se podrá ir avanzando en el conocimiento de la peculiar física del cerebro. Creo que la clave estará en que nuestras tecnologías aprendan a hacer lo que hacen las células vivas, en general, las neuronas, más en particular, y cómo funcionan las sinapsis, por ejemplo, sin pretender, por el contrario, que nuestras máquinas de tratamiento de la información tengan algo

[22] El texto de Chua, muy técnico, puede verse en la siguiente dirección: http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1083337.

[23] Carga y corriente, campo y voltaje nos dan dos de esas posibles relaciones, y las otras tres, hasta llegar a cinco, son los tres elementos o dispositivos que tradicionalmente ha considerado la teoría d circuitos, a saber, condensador, resistencia e inductor.

que enseñarles. En la medida que eso pueda hacerse, y nos enfrentamos a una dificultad de un orden casi extravagante, tal vez fuere posible pensar en una convergencia entre mentes y máquinas que está hoy muy lejos de nuestra agenda inmediata. La superioridad de la vida sobre el diseño formal es de tal naturaleza que desmiente claramente la presunción de Kurzweil (2005:478) de que "The patterns are more important than the materials that embody them". No pretendo negar la necesidad de una teoría general del funcionamiento del cerebro, de algo a lo que podamos llamar un *modelo*, pero sí creo, en primer lugar, que el conocimiento de base para llegar a algo como eso es todavía notoriamente insuficiente, y, en segundo lugar, que los modelos inspirados en el análisis formal del conocimiento, los modelos puramente lógicos y/o funcionales, se basan en un error de principio. Una de las últimas modas en neurobiología ha sido suponer que el cerebro funciona de una manera bayesiana, mediante mecanismos de anticipación, sobre todo a partir de las investigaciones de Friston (Huang, 2008), y de libros de divulgación como el de Hawkins (2004)[24], pero, incluso suponiendo que la teoría sea fértil, lo más interesante será siempre saber qué clase de propiedades tienen las neuronas y los sistemas que integran para permitir ese supuesto funcionamiento bayesiano.

No me parece que la posibilidad, todavía remota, de que por esta clase de vías, digamos, empíricas, podamos disponer de tecnologías que complementen y/o potencien nuestras funciones mentales deba plantear ninguna revolución conceptual, aunque reconozca sin dificultad que algunas de las casi infinitas revoluciones sugeridas por los historiadores se han basado en bastante menos. Al fin y al cabo, algunos podrían acabar reconociendo que, como dice Nicholas Humphrey[25], el "teatro cartesiano de la conciencia" sobre el que los filósofos modernos han sido generalmente tan escépticos, sea de hecho una realidad biológica, que nuestros cerebros son cómo son, precisamente, para sostenerlo. Incluso desde una perspectiva muy neutral y hasta puramente pragmatista, lo que debiera interesarnos no es tanto resolver un problema de carácter metafísico, como poder mejorar el conocimiento de nuestra mente y nuestro cerebro, de modo que podamos obtener mejoras de nuestro rendimiento y ayudas en un gran número de dificultades y problemas.

Un pensador tan escasamente mitómano como Freeman Dyson[26] se ha referido a la posibilidad, por ejemplo, de instrumentar alguna forma de radiotelepatía, una comunicación directa de sentimientos y pensamientos entre cerebro y cerebro. No es muy distinto a eso lo que hacemos cuando hablamos por teléfono, por ejemplo. Para hacer posible la radiotelepatía, según Dyson, sería necesario saber convertir señales neurales en señales de radio y construir radioreceptores microscópicos, además de saber reconocer el significado mental de las señales neurales, cosa que Dyson no menciona. Se trata de saberes que no poseemos, pero que no son inconcebibles. Dyson especula también con la posibilidad de sentir lo que un pájaro siente al volar, o lo que un ciervo siente al

[24] Puede verse una recensión de su libro e informaciones para seguir las ideas de Hawkins en http://www.uoc.edu/uocpapers/3/dt/esp/climent.pdf
[25] http://www.edge.org/q2008/q08_12.html
[26] http://www.edge.org/q2009/q09_3.html

ser abatido, posibilidad que imagino le podrá parecer tan quimérica a Thomas Nagel como me lo parece a mí.

Hay un segundo sentido en que podríamos hablar de inteligencia híbrida, si pensamos en los cambios que se acabarán dando en torno a las posibilidades de realización de lo que se pudiera llamar la "nueva mente colectiva", es decir integrando en el funcionamiento de nuestra inteligencia las nuevas posibilidades de información, y de conocimiento, que nos acabe brindando el crecimiento de la red, su especialización y la habilidad de cada cual para integrar todo eso en una imagen coherente de la realidad.

Parece claro que hay dos fuentes de innovación a las que se puede suponer una gran capacidad de modificar nuestros hábitos intelectuales, y, por tanto, las funciones y dimensiones de nuestra efectiva inteligencia de las cosas, a saber: la digitalización y el funcionamiento de la red que conocemos como Internet. La primera novedad consiste en que la información disponible pasa de estar soportada físicamente, de estar inscrita en documentos materiales, a estar soportada digitalmente, a estar en documentos intangibles, pero muy manejables, lo que permite mejoras espectaculares de accesibilidad, trasparencia, economía y participación. Una segunda gran novedad es que la lectura pueda dejar de ser una actividad enteramente pasiva, de modo que el lector podrá dejar su huella en lo que lee, lo que incrementará de manera notable la cantidad de información disponible, aunque lógicamente suponga también ciertos riesgos evidentes. En tercer lugar, la escritura podrá ejercerse en condiciones muy distintas a las históricas, sin apenas escasez de apoyos documentales, y con posibilidades de que se obtengan casi de manera inmediata respuestas a lo escrito, de manera que el debate, las conversaciones y las cartas entre los eruditos, que es la forma en que comenzó a despegar la ciencia moderna podrá desarrollarse en condiciones absolutamente ideales[27]. Por último no cabe olvidar que el aumento espectacular del número de los que tienen capacidad para intervenir en un debate cualquiera hace que podamos considerar, como dice Tim O'Reilly, que las ideas están siendo por sí mismas una forma especialmente relevante de software social. No hay que olvidar que el mayor negocio existente en la red, Google, se basa precisamente en una explotación muy inteligente de lo que la gente hace cuando usa la red, es una herramienta que explota el *software social* como fuente primaria.

Hemos de pensar que, a no mucho tardar, vamos a estar en condiciones de manejar con cierta facilidad masivas colecciones de datos relevantes para cualquier cosa y que esas colecciones estarán siendo permanentemente renovadas porque, por decirlo así, siempre podremos encontrar junto a los hechos contemplados, las mejores razones que se hayan podido averiguar para tenerlos por válidos. Esto supondrá una continua reedición de todo, una especie de historia interminable e instantánea que algunos podrán considerar como una espantosa amenaza si son esclavos de la imagen de que las verdades más sólidas deben estar escritas en algún lugar inaccesible y en letras de bronce. Los que pensemos que la verdad está en las proposiciones y que estas son, sobre todo, inmateriales, tan inmateriales como nuestra conciencia, estamos de

[27] Puede verse, al respecto González Quiros (2009 a) y González Quirós & Gherab Martín (2009 b).

enhorabuena, porque podremos disponer de algo que casi dan ganas de comparar con una fuente inagotable de sabiduría, claro es que para el que sepa beber y guste de hacerlo.

La decisiva influencia que han tenido en el progreso humano la escritura o la imprenta es un tópico de la historia cultural que nadie discute, y es también muy claro que la era digital está creando unas posibilidades considerablemente más poderosas y eficaces que las que trajeron consigo esas tecnologías históricas. Sin embargo, para cualquiera mínimamente familiarizado con el uso de la red, es inmediatamente evidente que las cosas no son todavía así, que las rémoras sociales, culturales e institucionales suponen un freno gravísimo de posibilidades tecnológicas enteramente ciertas. Al fin y al cabo, tenía razón Bacon y, puesto que la información es poder, no basta un poder meramente tecnológico para alcanzar un poder de verdad, pero esta es, sin duda, otra cuestión.

Referencias

Arana, Juan (2009): "¿Puede la libertad ser suplantada por elementos sucedáneos?", Anuario filosófico, XLII/2, pp. 273-303.

Behe, Michael J. (1999), *La Caja negra de Darwin*, Andrés Bello. Barcelona.

Carver, Mat (2007): *Minds and Computers. An Introduction to the Philosophy of Artificial Intelligence*, Edinburgh University Press, Trowbridge, Wilts.

Chalmers, David (1995): "Facing up to the problem of consciousness", Journal of Consciousness Studies, 2 (3), 1995, pp. 200-219, accesible en http://www.imprint.co.uk/jcs.html, y http://www.imprint.co.uk/chalmers.html.

Chalmers, D. J. (1996): *The Conscious Mind, In Search of a Fundamental Theory*, Oxford University Press, New York.

Clarke, Arthur C. (1999): *Greetings, Carbon-Based Bipeds!*, Ian T. Macauley, New York.

Davidson, Donald (2006). "Mental Events" en Lepore & Ludwig, Eds. pp. 105-121.

Dennett, Daniel (1995): *La conciencia explicada. Una teoría interdisciplinar*, Paidós, Barcelona.

Denton, Michael (2002): "Organism and Machine: The Flawed Analogy", en Richards, Jay W., Ed., p. 78-97.

García Gual, Carlos & y Acosta, Eduardo (1974): *Ética de Epicuro. La génesis de una moral utilitaria,* texto bilingüe, Barral, Barcelona.

González Quirós, José Luis (1998): *El porvenir de la razón en la era digital*, Síntesis, Madrid.

González Quirós, José Luis (2009 a): "El trabajo intelectual en el entorno digital: nuevas formas de escritura y erudición", Arbor, CLXXXV, 737, V-VI-2009, pp. 541-550.

González Quirós, José Luis & Gherab Martín, Karim (2009 b): Arguments for an Open Model of e-Science", en Bill Cope & Philip Angus, Eds., *The Future of the Academic Journal*, Chandos Publishing, London, 2009, ISBN 978-18-433-4416-2, pp. 63-83.

González Quirós, José Luis & Puerta, José Luis (2009 c): Tecnología, demanda social y *medicina del deseo*", en prensa, Medicina Clínica, Barcelona [MedClin(Barc).2009. doi:10.1016/j.medcli.2009.07.002]

Hawkins, Jeff & Blakeslee, Sandra (2004): *On intelligence*. Times Books, Nueva York. Existe una traducción española de 2005, *Sobre la inteligencia,* Espasa Calpe, Madrid.

Huang, Gregory T. (2008): "Essence of thought", New Scientist, 02624079, 5/31/2008, Vol. 198, Issue 2658

Juengst, Erik (1998): "What does enhancement mean?", en Parens, E. Ed. *Enhancing Human Traits: Ethical and Social Implications,* Georgetown University Press, Washington, D.C.

Kurzweil, Ray (1999): *The Age of Spiritual Machines*, Penguin, New York.

Kurzweil, Ray (2005): *The Singularity is Near*, Penguin, New York.

Lepore, Ernie & Ludwig, Kirk Eds. (2006): *The Essential Davidson*, Clarendon Pres, Oxford.

Mullins, Justin (2009): "Memristor minds: The future of artificial intelligence", The New Scientist, 08 July 2009, Magazine issue 2715, tal como puede verse en http://www.newscientist.com/article/mg20327151.600-memristor-minds-the-future-of-artificial-intelligence.html?full=true

Ortega y Gasset, José (1962): *La rebelión de las masas*, Obras completas, IV, Revista de Occidente, Madrid.

Ortega y Gasset, José (1996): *Meditación de la técnica y otros ensayos sobre ciencia y filosofía*, Revista de Occidente en Alianza Editorial, Madrid.

Pershin, Yuriy V. & La Fontaine, Steven & Di Ventra, Massimiliano (2008): "Memristive model of amoeba's learning", en http://arxiv.org/abs/0810.4179

Putnam, Hilary (1997): "Acerca de un mal uso del teorema de Gödel en la especulación sobre la mente" Revista de libros, nº 3, III-1997, pp, 31-32.

Putnam, Hilary (2001): *La trenza de tres cabos. La mente, el cuerpo y el mundo*, Siglo XXI de España Editores, Madrid.

Richards, Jay W., Ed. (2002): *Are We Spiritual Machines? Ray Kurzweil vs, the Critics of Strong A. I.,* Seatle, Discovery

Russell, Bertrand (1976): *La evolución de mi pensamiento filosófico*. Alianza, Madrid.

Schrödinger, Edwin (2006*) What is Life*, & *Mind and Matter, & Autobiographical Sketches*, Cambridge University Press, 15ª edición, Cambridge.

Sherrington, C.S (1984): *Hombre versus naturaleza,* Tusquets, Barcelona.

Searle, John (2002): "I Married a Computer", en Richards, Jay W., Ed., p. 56-78.

Sloterdijk, Peter (2000): *Normas para el parque humano. Una respuesta a "Carta sobre el humanismo" de Heidegger*, Siruela, Madrid.

Smullyan, Raymond (1989): *5000 años a. de C. y otras fantasías filosóficas*, Cátedra, Madrid.

Tallis, Ray & Aleksander, Igor (2008): 'Computer models of the mind are invalid', (texto de Tallis y discussion posterior de ambos), Journal of Information Technology 23, 55–62., doi:10.1057/palgrave.jit.2000128

Wolbring G. (2005): "The Triangle of Enhancement Medicine, Disabled People, and the Concept of Health: A New Challenge for HTA, Health Research, and Health Policy". Health Technology Assessment Unit, Alberta Heritage Foundation for Medical Research, HTA Initiative 23, December 2005 (disponible en: www.ihe.ca/documents/HTA-FR23.pdf).