

EL IMPACTO DE LAS CIENCIAS COGNITIVAS EN LA FILOSOFIA DEL CONOCIMIENTO

Pascual F. Martínez-Freire
Grupo de Investigación en Ciencias Cognitivas
Universidad de Málaga

RESUMEN

Después de señalar algunos enfoques erróneos en la filosofía del conocimiento, se defiende la necesidad de que esta disciplina se desarrolle teniendo en cuenta los datos científicos sobre el fenómeno del conocimiento. En particular, y ante el actual auge de las ciencias cognitivas, que son presentadas de forma sucinta, se defiende que la filosofía del conocimiento se desarrolle teniendo en cuenta algunas tesis básicas de las ciencias cognitivas, tales como la teoría representacional de la mente, la computación y algoritmos en el conocimiento, la tesis funcionalista del conocimiento y la naturalización moderada de la filosofía del conocimiento.

ABSTRACT

After pointing out some mistaken approaches in the philosophy of knowledge, it is put forward the need for this subject matter of being developed taking into account the scientific data about knowledge. Specially, and because of the progressive importance of cognitive sciences, which are presented briefly, it is put forward that the philosophy of knowledge must be developed taking into account some main theses of cognitive sciences, such as the representational theory of the mind, computation and algorithms in knowledge, the functionalist thesis of knowledge, and the moderate naturalism in the philosophy of knowledge.

1.- Introducción

En mi trabajo «Delimitación de las Ciencias Cognitivas» (1992) sostuve que la filosofía de la mente, tal como debe concebirse en los tiempos actuales, tiene tres caracteres básicos. En primer lugar, tiene el carácter de filosofía de la ciencia cognitiva, y en este sentido la filosofía de la mente (o mejor la *nueva* filosofía de la mente) aborda los problemas generales de las ciencias cognitivas, así como la comparación y evaluación de sus distintas teorías. En segundo lugar, tiene el carácter de constituir la nueva y más eficaz teoría del conocimiento, desplazando así a la crítica (metafísica) o vieja teoría del conocimiento. Y en tercer lugar la filosofía de la mente tiene también el carácter de proporcionar fundamentos a las ciencias cognitivas a través de la semántica filosófica, a la que preferimos denominar «teoría de la representación».

Del primer tema, es decir, de la filosofía de la mente en cuanto filosofía de las ciencias cognitivas, me he ocupado en mi libro *La Nueva Filosofía de la Mente* (1995). Sobre el tercer tema, esto es, de la filosofía de la mente como teoría de la representación, he anticipado algunas tesis en mi trabajo «Representación e interpretación»(1993), aunque un desarrollo completo del mismo ocupa buena parte de mis actuales estudios. Y respecto del segundo tema, es decir, acerca de la filosofía de la mente como la nueva y más eficaz teoría del conocimiento, pretendo establecer las tesis y argumentos principales en el presente texto.

2.- Enfoques erróneos en la filosofía del conocimiento

La teoría o filosofía del conocimiento es una disciplina filosófica que se remonta a los comienzos de la actividad intelectual en la Grecia clásica. No tiene nada de sorprendente que así haya ocurrido puesto que las cuestiones acerca de la naturaleza, fuentes y alcance (límites) del conocimiento preocupan al ser humano justamente en cuanto sus conocimientos pretenden ser objetivos y fundamentados. Es decir, en la misma medida en que los seres humanos alcanzan conocimientos ciertos surgen las cuestiones acerca de la índole del conocimiento.

Sin embargo, el estudio de los problemas del conocimiento, tal como deben abordarse en los tiempos actuales, debe evitar dos enfoques que podemos calificar de erróneos, a saber, el enfoque erudito-histórico y el enfoque meramente especulativo. Según lo primero, poco avance real cabe esperar si al estudiar los problemas del conocimiento nos limitamos a señalar las opiniones y puntos de vista de filósofos anteriores, oponiéndolas entre sí y comparándolas, en un ejercicio de erudición que puede resultar muy interesante para la historia de la filosofía pero que es de escaso valor para la filosofía. A este respecto, conviene dejar bien claro que la historia de la filosofía no es filosofía, de la misma manera que la historia de la física no es física ni la historia de la química es química, aunque deba reconocerse que, en general, la noticia de la historia o desarrollo de una disciplina puede tener utilidad para la construcción de la disciplina misma. Pero en todo caso constituye una aberración en la formación filosófica exagerar, tal como suele hacerse, el estudio de la historia de la filosofía, ya que al proceder así se

sustituye el estudio de los problemas y soluciones filosóficas tal como se plantean e importan al ser humano actual por el recorrido de intereses y planteamientos pasados que frecuentemente son ajenos a la altura de nuestros tiempos.

Por otra parte, según el enfoque meramente especulativo, las cuestiones acerca del conocimiento son tratadas con un vocabulario y un método que no tienen nada que ver con el conocimiento real que se da en los seres humanos (y, como veremos, en máquinas y en animales). Una jerga filosófica, debidamente oscura y confusa, pretende ocultar el hecho de que las tesis ofrecidas no han sido sometidas a ninguna contrastación empírica ni tampoco comparadas con los datos científicos. Es posible así escribir gruesos libros sobre el conocimiento sin referencia alguna a la psicología o a la biología, ya que las doctrinas sustentadas son asunto meramente especulativo. A este respecto, conviene señalar que la relación entre filosofía y ciencia puede tener dos estilos bien diferenciados. Por un lado, la filosofía puede pretender, fuera de tiempo y de contexto, hablar acerca de problemas que encuentran su lugar adecuado (y en función del momento histórico) en alguna disciplina científica. Este estilo en la relación entre filosofía y ciencia resulta inaceptable, ya que nadie (salvo algunos trasnochados) puede defender con fundamento que la filosofía es un saber absoluto que está por encima de los descubrimientos científicos. Por otro lado, la filosofía puede emprender la tarea de adelantar conjeturas sobre temas para los cuales aún no existe una sólida investigación científica, ofreciendo así sugerencias que pueden ser útiles u orientadoras a los científicos. Este estilo en la relación entre filosofía y ciencia es perfectamente legítimo, puesto que el filósofo cumple entonces el papel de proporcionar hipótesis de trabajo a la ciencia.

Pues bien, la filosofía del conocimiento habitual incurre en ambos enfoques erróneos, ya que muy frecuentemente se limita a un estudio erudito-histórico o bien a un desarrollo meramente especulativo al ocuparse de los problemas del conocimiento.

3.- El enfoque adecuado en la filosofía del conocimiento

En realidad, el enfoque adecuado en la filosofía del conocimiento exige tener en cuenta los descubrimientos científicos respecto de los mecanismos de nuestro conocer, y en particular los hallazgos de la psicología, de la neurociencia y de la biología. No pretendo con ello defender que la filosofía del conocimiento sea sustituida por estas disciplinas científicas, puesto que entiendo que siempre subsistirán cuestiones y problemas que rebasan las soluciones científicas del momento, bien porque la investigación científica aún no ha progresado lo suficiente (y en este caso la especulación filosófica establece hipótesis orientadoras para la ciencia), bien porque la investigación científica es naturalmente ajena a determinados problemas radicales o de fondo. En una y en otra línea se encuentran los problemas filosóficos, aunque creo que podemos decir que es en la segunda línea donde tiene su lugar más natural la filosofía.

Muy frecuentemente, cuando hablamos con los científicos (y el diálogo con ellos es muy importante para el filósofo, como es importante el diálogo con los filósofos para el científico) nos sorprende que ciertos problemas, que ocupan buena parte de la tarea

filosófica, no les interesan en absoluto, o incluso manifiestan su desdén hacia el hecho de que nos ocupemos de ellos, o más aún creen que nuestra ocupación no es sino palabrería inútil. Y sin embargo, también de modo frecuente, al proseguir de manera sincera el diálogo, terminan mostrando cierto interés, menos desdén y más respeto hacia las tesis filosóficas; y normalmente este cambio de actitud se produce en la medida en que abandonan su mentalidad de especialistas para adoptar una mentalidad más abierta.

4.- Filosofía del conocimiento y ciencias cognitivas

Si en términos generales la filosofía del conocimiento debe estar atenta a los descubrimientos científicos acerca de los mecanismos reales de los procesos de conocer, a la altura de nuestro tiempo la filosofía del conocimiento debe permanecer atenta a las teorías y hallazgos de las ciencias cognitivas contemporáneas. La tesis central que defendemos señala que el impacto de las ciencias cognitivas sobre la filosofía del conocimiento supone una reformulación profunda de sus doctrinas, alcance e intereses, justamente en la medida en que las ciencias cognitivas han determinado la constitución de una *nueva* filosofía de la mente, que se erige en el núcleo de la filosofía del conocimiento.

Pero antes debemos hablar, aunque sea de modo sucinto, de las propias ciencias cognitivas, en cuanto a su surgimiento y naturaleza.

Es un punto de vista generalmente admitido la idea de Howard Gardner, expuesta en *The Mind's New Science. A History of the Cognitive Revolution* (1985), de que las ciencias cognitivas se abren camino en 1948 en el Simposio celebrado en Pasadena, en la sede del Instituto de Tecnología de California y bajo el patrocinio de la Fundación Hixon. En esta notable reunión científica sucedieron varias cosas al mismo tiempo novedosas e importantes. El gran físico y matemático John Von Neumann habló acerca de la analogía entre el computador y el cerebro humano. Al respecto debe observarse que el título del Simposio era «Los Mecanismos Cerebrales en la Conducta» y que en 1946, dos años antes, se había completado el primer computador electrónico americano, el célebre ENIAC, en la Universidad de Pensilvania.

Por otra parte, también en el Simposio de Pasadena, Warren McCulloch, destacado neurocientífico, disertó sobre el modo en que el cerebro humano procesa información. Y finalmente, en este escogido muestrario, el psicólogo Karl Lashley, que había sido discípulo de Watson (el creador de la psicología conductista), atacó sin piedad al conductismo dominante por entonces, estableciendo de manera solemne que la organización de la conducta no es impuesta desde el exterior sino que dimana del interior del organismo. Si reunimos los datos anteriores, advertiremos algunas características propias de las ciencias cognitivas. En primer lugar el carácter interdisciplinar del campo de investigación, también la aparición de la analogía entre computador y cerebro, asimismo el interés por los procesos de información y, finalmente, la postura anticonductista tan típica de la psicología cognitiva.

Otro momento importante en el surgimiento de las ciencias cognitivas, también

señalado por Gardner, fué el Simposio sobre Teoría de la Información que tuvo lugar en 1956 en el célebre Instituto de Tecnología de Massachusetts. Entre las aportaciones más relevantes de esta reunión científica cabe citar las siguientes. En primer lugar, Allen Newell y Herbert Simon, dos notables informáticos, presentaron la primera demostración completa de un teorema realizable por un computador, en la línea de la naciente inteligencia artificial. En segundo lugar, el conocido lingüista Noam Chomsky puso de relieve su actitud mentalista, frente al conductismo, al insistir en que la organización del lenguaje es innata y universal. Finalmente, el psicólogo George Miller presentó un estudio sobre los límites humanos en la capacidad de procesar información. También en esta reunión científica se puso de manifiesto el carácter interdisciplinar del campo de investigación en marcha, así como el interés por el estudio de los mecanismos y procesos de información.

Sin embargo, las ciencias cognitivas se consolidan de manera definitiva, tal como nos enseña la moderna sociología de la ciencia, cuando a finales de los años setenta adquieren un carácter institucional al crearse una revista y una sociedad científica de investigación cognitiva. Como es habitual en el mundo académico norteamericano, tal consolidación se debió en buena medida al apoyo financiero de una fundación, la Fundación Sloan de Nueva York. Y en efecto, en 1977 se creó la revista *Cognitive Science* y, a su vez, en 1979 se reunió por primera vez en La Jolla (California) la Cognitive Science Society.

Conviene ahora, aunque sea de modo breve, referirnos a la naturaleza general de las ciencias cognitivas, así como a las ciencias concretas integrantes de este nuevo campo de investigación.

5.- Caracterización de las ciencias cognitivas

En términos generales podemos decir que las ciencias cognitivas constituyen un campo de investigación interdisciplinar cuyo tema central es el estudio del fenómeno de la cognición tanto en seres humanos como en máquinas y en animales. Y en este punto, claro está, resulta obligada una caracterización de la noción de cognición.

Al respecto, y tal como señalé en mi trabajo «Procesos mentales y cognitivismo» (1992), debemos distinguir dos sentidos de «cognición», uno inmediato y aproximado y otro más elaborado y adecuado. En primera instancia, cognición puede entenderse como sinónimo de conocimiento, en cuanto tomar cuenta de cualquier realidad dada o, dicho de un modo más apropiado a la terminología científica actual, en cuanto recepción de información. No debe olvidarse que un principio general de la biología actual es que todo ser vivo se esfuerza en obtener no sólo energía sino también información; y además es esencial a nuestra sociedad actual la construcción de máquinas para (al menos) almacenar información.

En segunda instancia, cognición quiere decir uso y manipulación de conocimiento, esto es, uso y manejo de información. Tanto en el primero como en el segundo

sentido de «cognición» nos encontramos con procesos portadores de información, con lo que la distinción entre simple conocimiento y cognición propia no es tajante, pero resulta claro que no pertenecen al mismo nivel cognitivo una sensación de calor asfixiante y la creencia de que una ducha fría aliviará la sensación de calor; en el primer caso recogemos datos siguiendo una línea básicamente pasiva, mientras que en el segundo caso manipulamos conocimientos diversos siguiendo una línea básicamente activa.

Dicho de manera más precisa, la cognición propia se caracteriza por ser un fenómeno fundamentalmente activo que procesa información. Además constituye un fenómeno interno (al igual que la cognición simple), en cuanto es causa interna de una conducta observable e incluso puede no manifestarse en conducta. Finalmente, y éste es un punto de vista novedoso aportado por las ciencias cognitivas, la cognición es algo compartido por seres humanos, animales y máquinas. En realidad, el hecho de haber acuñado un nuevo término («cognición») para referirnos a procesos que hubieran podido denominarse «conocimiento» está justificado en mi opinión por la novedad subyacente, proporcionada por las ciencias cognitivas, de entender que los procesos de conocimiento se explican como procesamiento de información y que tal procesamiento de información es común a humanos, animales y máquinas.

En efecto, para las ciencias cognitivas los sujetos de conocimiento son sistemas de procesamiento de información, concepción que sólo pudo desarrollarse tras la aparición de la teoría de la información y, más concretamente, tras la aparición de la teoría de la computación con Alan Turing. Por otro lado, también para las ciencias cognitivas, un ser humano, un gato o un computador son igualmente sistemas de procesamiento de información. Allen Newell y Herbert Simon, al comparar los seres humanos con los computadores, dicen: «El punto de vista básico presente en nuestra obra ha sido que el computador programado y el solucionador de problemas humano son ambas especies que pertenecen al género sistema de procesamiento de información. Difieren, por supuesto, en el nivel específico y en los modos en que nuestros conceptos describen: en organización de memoria, procesos elementales y organización de programa» (Newell y Simon, 1972, p.870).

En cuanto a las ciencias integrantes del campo de investigación cognitiva, debemos distinguir aquellas disciplinas que constituyen el núcleo de la investigación y aquellas otras disciplinas implicadas de manera instrumental. Las ciencias cognitivas básicas son la psicología cognitiva y la ciencia de la inteligencia artificial.

La psicología cognitiva puede considerarse constituída en 1960, cuando los psicólogos George Miller y Jerome Bruner fundan el Harvard Center for Cognitive Studies en la Universidad de Harvard. Justamente en ese mismo año aparece el libro *Plans and the Structure of Behavior*, escrito por Miller, Galanter y Pribram, que constituye un auténtico manifiesto del cognitivismo en psicología. A su vez, el primer libro de texto importante de psicología cognitiva se publica en 1967, su autor es Ulric Neisser y su título *Cognitive Psychology*. Para una adecuada psicología cognitiva, los procesos mentales (en humanos y en animales) son procesos cognitivos en el sentido ya advertido de

manipuladores de información; con ello, los estados internos portadores de información son representaciones. La psicología cognitiva supone la recuperación del mentalismo (o vida mental interna) enriquecida con una pluralidad de métodos: observación y experimentación de la conducta (pues se descarta el conductismo, pero no el análisis de la conducta), métodos computacionales e investigación neurobiológica, además del recurso a la introspección y al testimonio personal.

Por otro lado, la ciencia de la inteligencia artificial o, brevemente, la inteligencia artificial puede caracterizarse de modo amplio como el estudio del diseño y construcción de máquinas inteligentes, entendiendo por tales mecanismos capaces de imitar la actividad inteligente humana. Por ejemplo, los computadores dotados de programas para jugar al ajedrez son máquinas inteligentes. En inteligencia artificial importa básicamente el diseño de programas (es decir, el software de las máquinas), aunque el estudio de los soportes físicos (esto es, el hardware de las máquinas) también interesa si tiene relevancia para la programación. El principal animador del campo de la inteligencia artificial, y también quien acuñó tal expresión, fué el matemático americano John McCarthy. En 1957 fundó el laboratorio de inteligencia artificial del Instituto Tecnológico de Massachusetts y, más tarde, en 1963 el laboratorio de inteligencia artificial de la Universidad de Stanford. Otras grandes figuras clásicas de esta ciencia son Allen Newell, Herbert Simon y Marvin Minsky.

Pero no son la psicología cognitiva y la inteligencia artificial las únicas ciencias implicadas en el campo de las ciencias cognitivas. También debe contarse con otras disciplinas que, aunque no agotan sus intereses en el fenómeno de la cognición, constituyen importantes recursos o instrumentos para el estudio de la cognición.

Tal ocurre con la informática, que se ocupa en general de la teoría, diseño, producción y empleo de los ordenadores. La informática es relevante para las ciencias cognitivas en cuanto proporciona los conocimientos y detalles técnicos precisos para comprender y utilizar los sujetos cognitivos (humanos, animales o máquinas) como sistemas procesadores de información. La informática en particular aporta a las ciencias cognitivas un arsenal de conceptos y técnicas que han revolucionado, a su vez, la investigación sobre los fundamentos y procesos de conocimiento.

A su vez, la lógica también interesa a las ciencias cognitivas en varios sentidos. En primer lugar constituye un poderoso instrumento para la representación del conocimiento, esto es, para la ordenación y presentación del conocimiento, ya sea mediante la lógica clásica de primer orden, ya sea mediante el empleo de lógicas no-clásicas como la lógica modal o la lógica epistémica, o incluso mediante el recurso a lógicas no-monótonas que estudian el razonamiento revisable. Además la lógica sirve a la informática para la elaboración y evaluación de programas. Las técnicas formales de la lógica y sus desarrollos del control del razonamiento son esenciales en las ciencias cognitivas.

La neurociencia es asimismo un poderoso instrumento en el campo de las ciencias cognitivas. Por un lado, y respecto de la psicología, permite identificar los estados

cerebrales responsables de la actividad mental en humanos y en animales. Por otro lado, y respecto de la inteligencia artificial, proporciona modelos para el diseño de redes neuronales artificiales, que constituye una línea de investigación de creciente interés dentro de la inteligencia artificial.

La lingüística, por su parte, tiene gran importancia para la psicología cognitiva en particular y para las ciencias cognitivas en general. En efecto, la lingüística ha desempeñado un gran papel en el abandono del conductismo (y con ello en la instauración de la psicología cognitiva) con su insistencia en el carácter innato de la competencia lingüística; además, puesto que los estudios acerca del lenguaje ocupan un lugar destacado en psicología, la colaboración entre lingüistas y psicólogos se ha hecho obligada. Por otra parte, la noción de lenguaje y el estudio de sus diferentes tipos (del pensamiento, naturales, formales, de computación) tienen un alcance general dentro de las ciencias cognitivas.

Finalmente, la filosofía de la mente también se incluye en el campo de investigación cognitiva en cuanto la metaciencia cognitiva, es decir, el examen de los problemas generales de las ciencias cognitivas, así como la comparación y evaluación de sus teorías.

6.— La teoría representacional de la mente

La nueva filosofía de la mente, sensible al desarrollo de las ciencias cognitivas, proporciona a la filosofía del conocimiento, de la cual constituye su núcleo, nuevas claves para reformular sus doctrinas, alcance e intereses. Aunque la lista de estas claves podría hacerse extensa, creo que podemos reducir la lista a cuatro items básicos, a saber, la teoría representacional de la mente, computación y algoritmos en el conocimiento, la tesis funcionalista del conocimiento y, por último, la naturalización moderada de la filosofía del conocimiento. Vamos a ver separada y sucesivamente tales claves.

Dentro del campo de investigación de las ciencias cognitivas constituye una asunción básica la tesis de que el fenómeno de la cognición no sólo está presente en los seres humanos, sino también en los animales y asimismo en algunas máquinas. Por ello, y frente a Descartes y a toda tradición cartesiana, la mente no es un privilegio humano sino que es legítimo hablar de mentes animales e incluso de mentes mecánicas. Como es sabido, para Descartes, la evidencia de que algo tiene mente es su posesión de un lenguaje, pero, como (para el filósofo francés) ni los animales ni las máquinas poseen lenguaje, se concluye que ni los animales ni máquina alguna poseen mente. En este punto, las ciencias cognitivas señalan a la filosofía del conocimiento un ámbito de estudio mucho más amplio que el habitualmente considerado.

Por otra parte, otra asunción básica de las ciencias cognitivas establece que la mente es un sistema de estados internos que representa (con diferentes grados de convencionalidad y eficacia) el mundo externo. Ahora bien, cuando hablamos de un sistema de estados internos nos situamos claramente dentro del punto de vista mentalista y fuera del punto de vista conductista. Para el conductismo, en efecto, no

existen procesos mentales internos ni por ende propiamente una «vida mental», sino que los fenómenos psicológicos se reducen a la conducta, observable y registrable pública y externamente. En cambio, para el mentalismo existen procesos mentales internos, es decir, procesos que pueden no manifestarse en conducta o bien son causa interna de la conducta, y por ello no son reducibles a la conducta.

En cuanto *sistema* de estados internos, la mente constituye un conjunto organizado, una estructura, en la que deben distinguirse aspectos preformados (e incluso innatos en el caso de humanos y animales) así como aspectos desarrollados, inducidos o aprendidos. Aunque sea posible aislar módulos o «facultades» de la mente, que están cerradas salvo en sus salidas de información, en su conjunto la mente es una entidad de totalización, una arquitectura.

Pero tiene mayor interés destacar el papel y la importancia de las representaciones en cuanto elementos constitutivos de cualquier mente. Desde el punto de vista cognitivo, propiamente no existe «la mente», en la medida en que pudiera ser entendida como una substancia ajena a cualquier investigación empírica, sino que lo que realmente existe son procesos mentales (clases de distinta naturaleza y funciones), que son representaciones del mundo externo (externo al organismo o, al menos, a la «vida mental» misma). Por ello cuando hablamos de la naturaleza representacional de la mente queremos decir que la mente es un conjunto de diversas clases de representaciones. También en este punto las ciencias cognitivas señalan a la filosofía del conocimiento que las investigaciones acerca de los fenómenos cognoscitivos deben reducirse al estudio de las representaciones reales.

Frecuentemente se ha sostenido que la intencionalidad es un rasgo característico de los procesos mentales, es decir, que los procesos mentales se caracterizan por referirse a algo distinto de ellos mismos, versar sobre algo, dirigirse a algo, etc.; por ejemplo, se percibe algo, se recuerda algo, se cree algo o se desea algo. Sin embargo, el tema de la intencionalidad, que tanto alboroto y tantos ríos de tinta ha provocado, no encierra ningún secreto misterio si se considera a la luz de las ciencias cognitivas. En efecto, la intencionalidad es la propiedad de representar y, en este sentido, toda representación es una intención (aunque no, claro está, en el sentido de propósito). Las representaciones o intenciones, dentro del marco de las ciencias cognitivas, son estados internos, ya sea de un cerebro natural o de un cerebro artificial (computador). Con ello, la intencionalidad o propiedad de representar no es exclusiva de los seres humanos.

Por otra parte, en toda representación podemos distinguir su instanciación física y su valor semántico. Efectivamente, aunque la descripción de las representaciones puede hacerse a un nivel general, de tal manera que puedan cubrirse sus distintos géneros y modalidades, también es posible tener en cuenta la realización física de las distintas muestras de representaciones. Desde este nivel, las representaciones pueden instanciarse en circuitos neuronales, es decir, en mecanismos cerebrales vivos, o bien pueden instanciarse en circuitos electrónicos, esto es, en mecanismos artificiales.

El carácter representacional de las mentes humanas y animales es un punto de

vista que no suscita apenas oposición una vez planteado. Quizás existan todavía algunas resistencias a aceptar las mentes animales como comparables a las mentes humanas, pero Donald Griffin, en su libro *Animal Minds* (1992), proporciona datos de investigación empírica y argumentos más que suficientes para aceptar tal punto de vista. En cambio, en el caso de las mentes mecánicas cabe abrigar algunas dudas acerca del carácter completo de sus representaciones. En efecto, por un lado, no existen dudas razonables sobre la existencia de estados internos en los computadores, ya que todo computador en su descripción general como máquina de Turing posee estados internos que deben entenderse como la suma total de información almacenada en la máquina en un momento dado. Pero, por otro lado, se ha suscitado de modo reiterado la duda acerca del valor semántico de tales estados internos, es decir, se ha puesto en duda que las representaciones de un computador posean significados por sí mismos.

Sobre este asunto podemos ser (en este momento) moderados y remitirnos a la célebre hipótesis de representación del conocimiento, propuesta por Brian Smith en su trabajo «Reflections and Semantics in a Procedural Language» (1982). Según tal hipótesis, cualquier proceso inteligente realizado mediante un mecanismo estará compuesto de ingredientes estructurales que, en primer lugar, nosotros consideramos de manera natural (como observadores externos) que representan una explicación proposicional del conocimiento que el proceso global exhibe, y que, en segundo lugar, independientemente de tal atribución semántica externa, juegan un papel formal pero causal y esencial en la producción de la conducta que manifiesta ese conocimiento. Es decir, según esta hipótesis de representación del conocimiento, las máquinas inteligentes manifiestan un conocimiento y, al mismo tiempo, poseen ingredientes a los que atribuimos de modo natural la capacidad de representar ese conocimiento, además de producirlo.

En la línea de esta hipótesis, el valor semántico, esto es, los significados son una atribución externa del programador y del usuario de la máquina. Aunque creo que la hipótesis puede hacerse más radical (y apuntar que la propia máquina también interpreta), es importante para la filosofía del conocimiento tomar nota de que, al menos, los computadores exhiben conocimiento y poseen representaciones.

7.- Computación y algoritmos en el conocimiento

Una segunda clave para reformular la filosofía del conocimiento viene proporcionada por la idea de que la computación y sus algoritmos correspondientes están presentes en los procesos cognoscitivos.

John Haugeland señala (*Artificial Intelligence. The Very Idea*, 1985, p. 23) que fué el filósofo Thomas Hobbes (1588-1679) quien proclamó que por razonamiento entendía computación. Con ello se abría la posibilidad de concebir e incluso diseñar los procesos del conocimiento como procesos de computación. Justamente las representaciones, tanto en las mentes humanas y animales como en las mentes mecánicas, constituyen, desde el punto de vista de las ciencias cognitivas, el medio y el instrumento de computación.

El paralelismo entre la computación cerebral y la computación artificial es, en ciencias cognitivas, una tesis aceptada, aunque las discusiones sobre el alcance exacto de tal paralelismo sean numerosas y sus resultados diversos. Por ejemplo, John Searle, en su trabajo «Minds, Brains, and Programs» (1980), hace la distinción entre lo que denomina «inteligencia artificial fuerte» y lo que estima «inteligencia artificial débil o prudente». La segunda se caracteriza por considerar los computadores como poderosos instrumentos para el estudio de la mente (reservada, según Searle, a los cerebros naturales), permitiéndonos formular y comprobar hipótesis de un modo riguroso y preciso. En cambio, la inteligencia artificial fuerte considera los computadores como más que un instrumento para el estudio de la mente, ya que sostiene que un computador programado de manera adecuada es realmente una mente al poder literalmente pensar y tener otros estados cognitivos. Searle, como es sabido, arremete contra la inteligencia artificial fuerte (y con ello contra las ciencias cognitivas), aunque no pone objeciones a la inteligencia artificial débil, esto es, al empleo de los computadores como un recurso instrumental y conceptual para la comprensión de la mente.

La postura habitual en las ciencias cognitivas es solidaria de la inteligencia artificial fuerte. La tesis generalmente asumida defiende que hay realmente mentes mecánicas (además de las mentes humanas y animales). Sin embargo, desde mi punto de vista ya manifestado con anterioridad, las mentes cerebrales naturales y las mentes artificiales, aunque ambas pertenecen al mismo género de sistema de procesamiento de información, disponen de mecanismos y cualidades distintas. Más aún, en el caso de las mentes humanas se dan procesos «espirituales» no compartidos con otras mentes. En efecto, los procesos mentales de autoconciencia (o conciencia de sí mismo como sujeto de procesos mentales), los procesos de formación de un proyecto vital personal, y las voliciones libres o indeterministas parecen ser características exclusivas de los seres humanos, las cuales de modo muy plausible suponen una índole no-física (espiritual).

En todo caso, los procesos mentales se entienden dentro de las ciencias cognitivas como procesos de computación. La inteligencia artificial puede eludir hablar de procesos mentales pero evidentemente no puede dejar de hablar de procesos de computación que generan conocimientos, mientras que la psicología cognitiva no puede dejar de hablar de procesos mentales y se denomina frecuentemente «psicología computacional».

Ahora bien, dentro del campo de las ciencias cognitivas, tanto en la psicología cognitiva como en la inteligencia artificial, se distinguen dos tipos generales de computación, que son la computación clásica y la computación conexionista. Por una parte, la computación clásica consiste en manipulación de símbolos (representaciones) definida en términos de reglas formales (más o menos fijas) y realizada en un procesamiento serial o secuencial. Este tipo de computación se denomina «tipo von Neumann», en atención al famoso matemático e informático John von Neumann. Por otra parte, la computación conexionista se diseña como manipulación de unidades simples con cierto grado de activación y que están masivamente interconectadas, de tal manera que se producen entre las unidades excitaciones e inhibiciones, y además tal computación sigue un procesamiento paralelo, no secuencial.

Dicho de otro modo, la computación de corte clásico consiste en la aplicación serial o sucesiva de reglas formales, almacenadas de manera explícita, sobre representaciones simbólicas explícitas y localizables de manera definida. En cambio, la computación de corte conexionista consiste en la aplicación en paralelo de reglas no predeterminadas de modo fijo (sino flexibles) sobre unidades de activación conexas, las cuales representan al nivel de la red constituida por las unidades y no al nivel de cada unidad localizable.

En el seno de la psicología cognitiva se ha producido un duro enfrentamiento entre psicólogos clásicos y conexionistas, siendo Jerry Fodor y Zenon Pylyshyn los defensores de la computación clásica mientras que Paul Smolensky, entre otros, defiende la computación conexionista. Mi punto de vista en tal enfrentamiento es que ambos tipos de computación constituyen recursos y metodologías diferentes dentro de un mismo paradigma general, a saber, el paradigma centrado en la idea de que los procesos mentales son estados internos que manipulan información. Por otra parte, entiendo que es competencia directa de la inteligencia artificial evaluar ambas metodologías en cuanto técnicas y, a su vez, compete a la psicología cognitiva decidir la aplicación de una u otra según criterios de oportunidad y éxito.

Resulta interesante observar cómo se han originado ambas metodologías. Desde los años sesenta los psicólogos cognitivos asumieron, con menor o mayor decisión, que los procesos mentales eran procesos computacionales y, en consecuencia, recurrieron a los conceptos e instrumentos de la informática y, en particular, de la inteligencia artificial. Como la metodología dominante por entonces era la de la computación clásica, los psicólogos cognitivos asumieron tal metodología. Sin embargo, en la segunda mitad de los años ochenta la computación conexionista consiguió ser reconocida como una alternativa seria y viable, lo cual provocó que algunos psicólogos cognitivos asumieran entonces esta nueva metodología.

Pero conviene poner de relieve que la computación conexionista está, en cierto sentido, naturalmente más próxima a la psicología que la computación clásica. En efecto, la metodología conexionista realiza el diseño de redes neuronales artificiales que imitan la organización y el funcionamiento del cerebro vivo, el cual consta de redes de neuronas o células nerviosas interconectadas de manera masiva, con uniones o sinapsis excitatorias o inhibitorias, y con un tipo de procesamiento en paralelo. Dado que los procesos mentales de humanos y animales se instancian físicamente en redes neuronales resulta más natural, en este sentido, el recurso a la computación conexionista.

Por otra parte, cualquier computación requiere un algoritmo, que es, en primera instancia, un conjunto de instrucciones para operar. En términos ideales, un algoritmo es un procedimiento mecánico, sometido a reglas fijas y definidas, que permite resolver un problema o ejecutar una tarea en un número finito de pasos. Un ejemplo es el algoritmo de Euclides que establece si dos enteros positivos tienen o no algún común divisor diferente de 1. En cuanto procedimiento o método mecánico, un algoritmo no precisa para su aplicación de talento inventivo o ingenio alguno; por ejemplo, el

algoritmo de tablas veritativas para decidir si una fórmula de lógica de enunciados es o no una ley lógica no requiere ningún especial talento. Además, las instrucciones o reglas que definen un algoritmo deben hacerse explícitas, así como las ocasiones en que se emplea cada una de ellas. También es imprescindible que un algoritmo quede cumplido en un número finito de ejecuciones, ya que en caso contrario no se produciría una respuesta al quedar continuamente aplazada. Finalmente, un algoritmo tiene un propósito general, es decir, se establece para resolver una clase de problemas y no un solo problema concreto.

Aunque estas son las condiciones ideales de un algoritmo, existen algunas modificaciones de interés, e incluso podemos distinguir, en cierto sentido, entre algoritmos fuertes y métodos heurísticos o algoritmos débiles. En efecto, por un lado, frecuentemente un algoritmo puede poseer cierta ambigüedad, con lo cual, y según los casos de aplicación, necesitará ciertas reglas o instrucciones suplementarias para poder ser desarrollado. Por otro lado, también es posible que un algoritmo diseñado para ejecutar cierta clase de tareas sea transferido a otra clase de tareas distinta, donde también tiene éxito.

Mayor importancia tiene la existencia de heurísticos o algoritmos débiles. En estos casos ya no se trata de algoritmos que producen una respuesta de manera totalmente mecánica y siguiendo un conjunto de reglas fijas y explícitas, aunque se sigue exigiendo que la posible respuesta se produzca en un tiempo finito. En el caso de los heurísticos tenemos un conjunto de instrucciones que desempeñan el papel de estrategias o trucos que facilitan la resolución de una clase de problemas o de un tipo de tareas. La función básica de un heurístico consiste en simplificar la búsqueda de soluciones en espacios amplios de problemas. Pero también es una característica suya que no garantizan el encuentro de una solución y tampoco garantizan que la respuesta (o respuestas) dada sea la mejor.

El empleo de heurísticos o algoritmos débiles es habitual en el diseño de sistemas expertos en inteligencia artificial, esto es, en la confección de programas que simulan el conocimiento de un experto humano (médico, químico, geólogo, etc.). En estos casos se trata de acotar aquellos parámetros o aspectos más relevantes en la búsqueda de soluciones, dejando de lado aspectos o caminos de búsqueda que, de acuerdo con el conocimiento del experto humano, son irrelevantes o de muy difícil manipulación.

Sin duda, algunos procesos mentales pueden desarrollarse y diseñarse siguiendo un algoritmo en sentido fuerte, tales como ciertas operaciones aritméticas o demostraciones en lógica y en matemáticas, pero muy frecuentemente los procesos mentales se ordenan siguiendo un programa complejo que es variable y flexible, es decir, de acuerdo con un algoritmo en sentido débil. En estas condiciones, el algoritmo de computación consiste en una ordenación racional de pasos, caracterizada por la posibilidad de improvisar o completar algún paso, así como por la índole variable del resultado dentro de un esquema genérico.

8.- La tesis funcionalista del conocimiento.

En el análisis de los fenómenos del conocimiento son posibles dos estrategias diferentes. En primer lugar, una estrategia que se cumple desde arriba hacia abajo, es decir, desde el nivel de los fenómenos de cognición descritos en términos generales, pasando por el nivel bien neurobiológico bien informático (maquinaria electrónica), y llegando hasta el nivel físico. En segundo lugar, tenemos la estrategia inversa, de abajo hacia arriba, que se desarrolla desde el nivel físico (y en su caso bioquímico), pasando por el nivel de la neurociencia o de la informática (tecnología de ordenadores), hasta llegar al nivel más abstracto de la cognición. Ambas estrategias son legítimas y necesarias en las ciencias cognitivas, aunque de hecho exista cierta preferencia por la estrategia de arriba hacia abajo, preferencia que se justifica a fin de evitar el reduccionismo de la psicología a la neurociencia o bien el reduccionismo de la psicología a la bioquímica (e incluso, aunque no es habitual tal postura, a la física), así como también a fin de evitar la reducción de la inteligencia artificial a la física (opción, por cierto, que no se presenta).

La primera estrategia, de arriba hacia abajo, es solidaria de la tesis funcionalista del conocimiento, tal como se entiende en ciencias cognitivas, mientras que la segunda estrategia, de abajo hacia arriba, va acompañada frecuentemente de una naturalización de la teoría del conocimiento. En este apartado presentaré los argumentos básicos de la tesis funcionalista del conocimiento, mientras que en el siguiente apartado defenderé una moderada naturalización de la filosofía del conocimiento.

El funcionalismo, dentro del campo de las ciencias cognitivas, sostiene dos tesis básicas. Por una parte, los procesos mentales se conciben como estados funcionales, es decir, como estados internos, que median entre entradas sensoriales y salidas conductuales al tiempo que se relacionan con otros estados mentales, y que se caracterizan por el papel o rol causal que desempeñan respecto de la conducta. Por otra parte, los procesos mentales, caracterizados según su papel causal sobre la conducta, pueden ser estudiados prescindiendo de su instanciación o soporte físico. En suma, lo que importa cuando analizamos los procesos mentales de un sujeto cognitivo es la organización funcional que tales procesos constituyen y no la materia de la que está hecho tal sujeto cognitivo.

Hilary Putnam, en su trabajo «Psychological Predicates» (1967), fué el primero en proponer el funcionalismo. En una obra posterior escribe: «Un espíritu incorpóreo podría presentar un cierto programa, un cerebro podría presentar cierto programa, una máquina podría presentar cierto programa, pero la organización funcional de los tres (el espíritu incorpóreo, el cerebro y la máquina) podría ser exactamente la misma aún cuando su materia, su sustancia, fuese completamente diferente» (*Razón, verdad e historia*, p. 86). Así pues, el comportamiento cognoscitivo de un sistema u organismo se define ante todo por su organización funcional, siendo de importancia secundaria la constitución física del sistema. Por otra parte, el funcionalismo no es incompatible con el dualismo, puesto que un sistema, si lo hubiera, compuesto de cuerpo y alma puede ser un sujeto cognitivo definido según una cierta organización funcional.:

El funcionalismo es una tesis que goza de muy amplia aceptación. En efecto, dentro de las ciencias cognitivas, es generalmente aceptado por los psicólogos cognitivos y por los científicos de inteligencia artificial. Además, para los filósofos de la mente el funcionalismo supone una especie de «tiempo muerto» en el problema mente-cerebro, ya que puede ser asumido tanto por el filósofo de orientación materialista (para quien el soporte de los procesos mentales es el cerebro) como por el filósofo de orientación espiritualista (para quien algunos o todos los procesos mentales tienen su soporte último en el espíritu).

La filosofía del conocimiento puede adoptar la tesis funcionalista. Con ello se sitúa en un nivel de análisis del conocimiento lo bastante general como para poder caracterizar los fenómenos cognoscitivos con independencia de las peculiaridades «orgánicas» del sujeto cognitivo (humanos, animales o máquinas).

9.- Naturalización moderada de la filosofía del conocimiento.

Aunque al asumir la hipótesis funcionalista abstraemos de los caracteres concretos de los soportes de los sujetos cognitivos, tal abstracción, que se justifica por su búsqueda de las tesis generales en filosofía del conocimiento, no pretende ni debe pretender renunciar a la estrategia de análisis de abajo hacia arriba que antes mencioné.

En efecto, y tal como señalé en el apartado 3, la filosofía del conocimiento debe tener en cuenta los descubrimientos científicos acerca de los mecanismos reales y concretos del conocimiento de los sujetos cognitivos, tomados ahora en su diversidad. Esta exigencia va unida a una naturalización de la filosofía del conocimiento que, con todo, hay que adoptar en sus términos adecuados.

Werner Callebaut, en su libro *Taking the Naturalistic Turn* (1993), entrevista a veinticuatro intelectuales de diferentes disciplinas, con el objetivo confesado de naturalizar la filosofía de la ciencia, y lo que en esta obra se dice de la epistemología cabe extenderlo a la teoría del conocimiento en general. Por naturalismo se entiende la doctrina de que todo lo que sucede o existe en el mundo es susceptible de una explicación sometida exclusivamente a los métodos científicos naturales, con lo que al mismo tiempo se rechaza que haya o pueda haber algo más allá del alcance de la explicación científica. A su vez, cabe añadir, la naturalización de la filosofía del conocimiento consistiría en su reducción completa a la ciencia natural.

Este programa de naturalización puede incluso cumplirse desde una ciencia de «nivel superior» como la psicología hacia una ciencia de «nivel inferior» como la neurociencia, siguiendo en este caso una vía reduccionista. Tal es la propuesta, por ejemplo, contenida en el materialismo eliminativo de Paul Churchland, quien defiende en efecto sustituir la psicología en general por la neurociencia. Precisamente el materialismo eliminativo de Churchland va unido, tal como se indica en su libro *Matter and Consciousness* (1988), al materialismo metodológico, consistente en una aproximación de abajo hacia arriba que sostiene que las actividades cognitivas son en última instancia

actividades del sistema nervioso, de tal manera que se debe comenzar por comprender el comportamiento físico, químico, eléctrico y de desarrollo de las neuronas y así podremos comprender todo lo que hay que saber acerca de los procesos mentales.

Los datos y teorías empíricos que pueden proporcionar disciplinas científicas como la biología, la bioquímica, la neurociencia, la tecnología de ordenadores, etc. a la filosofía del conocimiento son realmente de gran importancia e interés, pero esta naturalización de la filosofía del conocimiento debe ser moderada en dos sentidos. En primer lugar, el estudio de las bases «orgánicas» del conocimiento debe permanecer alejado de cualquier reduccionismo, al igual que la psicología no se reduce a neurociencia ni la inteligencia artificial a la tecnología de ordenadores. En segundo lugar, tal como señalé en el apartado 3, en la filosofía del conocimiento subsisten cuestiones y problemas que rebasan las soluciones científicas del momento, bien porque la investigación científica aún no ha progresado lo suficiente (y en este caso la especulación filosófica suministra hipótesis orientadoras para la ciencia), bien porque la investigación científica es naturalmente ajena a ciertas cuestiones radicales o de fondo.

BIBLIOGRAFÍA

- Callebaut, W., *Taking the Naturalistic Turn or How Real Philosophy of Science is Done*, Chicago, The University of Chicago Press, 1993.
- Churchland, P. M., *Matter and Consciousness* (ed. revis.), Cambridge (Mass.), MIT Press, 1988.
- Gardner, H., *La nueva ciencia de la mente* (trad. Leandro Wolfson), Barcelona, Paidós, 1988.
- Griffin, D., *Animal Minds*, Chicago, The University of Chicago Press, 1992.
- Haugeland, J., *Artificial Intelligence. The Very Idea*, Cambridge (Mass.), MIT Press, 1985.
- Martínez-Freire, P., «Procesos mentales y cognitivismo», *Revista de Filosofía*, vol. V, nº 7, 1992, pp. 143-159.
- Martínez-Freire, P., «Delimitación de las ciencias cognitivas», *Anales del Seminario de Metafísica*, nº extra, 1992, pp. 443-451.
- Martínez-Freire, P., «Representación e interpretación», *Themata*, nº 11, 1993, pp. 37-46.
- Martínez-Freire, P., *La nueva filosofía de la mente*, Barcelona, Gedisa, 1995 (en prensa).
- Miller, G., Galanter, E. y Pribram, K., *Plans and the Structure of Behavior*, New York, Holt, Rinehart and Winston, 1960.
- Neisser, U., *Cognitive Psychology*, New York, Meredith, 1967.
- Newell, A., y Simon, H., *Human Problem Solving*, Englewood Cliffs (N. J.), Prentice-Hall, 1972.
- Putnam, H., «Psychological Predicates», W. H. Capitan y D. D. Merrill (eds.), *Art, Mind and Religion*, Pittsburgh, University of Pittsburgh Press, 1967. (Reimpreso posteriormente con el título de «The Nature of Mental States»).
- Putnam, H., *Razón, verdad e historia* (trad. José Miguel Esteban), Madrid, Tecnos, 1988.
- Searle, J., «Minds, Brains and Programs», *The Behavioral and Brain Sciences*, 3, 1980, pp. 417-424.
- Smith, B. C., «Reflection and Semantics in a Procedural Language», Tech. Report MIT/LCS/TR-272, Cambridge (Mass.), MIT, 1982.