

# EL PRONÓSTICO DE DESCARTES SOBRE LOS PROBLEMAS DE LA INTELIGENCIA ARTIFICIAL

## DESCARTES' PREDICTIONS ON THE PROBLEMS OF ARTIFICIAL INTELLIGENCE

*Manuel Carabantes López<sup>a</sup>*

Fechas de recepción y aceptación: 16 de marzo de 2016, 2 de septiembre de 2016

*Resumen:* En el siglo XVII, el fundador de la filosofía moderna, René Descartes, se basó en su teoría metafísica del dualismo de sustancias para afirmar que jamás podrían recrearse mediante disposiciones de la sustancia material dos cualidades exclusivas de la sustancia pensante: la capacidad para el lenguaje natural y la flexibilidad para habérselas con problemas de todo tipo. Casi cuatro siglos después, el pronóstico de Descartes todavía se cumple, y las dos cualidades por él señaladas se resisten a los intentos de duplicación de la inteligencia artificial (IA) contemporánea. El presente artículo analiza las causas del acierto del filósofo francés.

*Palabras clave:* Descartes, inteligencia artificial, lenguaje natural, autómatas, dualismo de sustancias.

*Abstract:* In the seventeenth century, the founder of modern philosophy, René Descartes, stated from his metaphysical theory of substance dualism that by setting up the material substance it would never be possible to replicate two unique qualities of the thinking substance: the capacity for natural language and the flexibility to deal with problems of all types. Nearly four centuries later, Descartes' forecast still holds, and the two qual-

<sup>a</sup> Universidad Complutense de Madrid. Departamento de Lógica y Filosofía de la Ciencia. Correspondencia: Calle Profesor Aranguren, 5. 28040 Madrid. España.  
E-mail: manuel.carabantes@gmail.com



ities mentioned by him resist the duplication attempts of contemporary artificial intelligence (AI). This article analyzes the causes of the success of the French philosopher's forecast.

*Keywords:* Descartes, artificial intelligence, natural language, automata, substance dualism.

## §1. UN PRONÓSTICO SORPRENDENTE

En la quinta parte del *Discurso del método*, publicado por primera vez en 1637, Descartes traza una línea divisoria entre el ser humano y las máquinas que todavía hoy, cuatro siglos después y con toda la tecnología a nuestro alcance, no ha sido traspasada. En ese lugar el genio francés comienza exponiendo lo que él mismo dice que es un resumen de una obra propia sobre anatomía y fisiología del cuerpo humano escrita unos años antes, en 1633, pero no publicada hasta después de su muerte, muy probablemente por miedo a que le valiese unas represalias por herejía como las sufridas por Galileo en noviembre de ese mismo año (López-Muñoz y Álamo, 2000: 247). Tras dedicar casi toda la extensión de la quinta parte del *Discurso* a extraer la mencionada obra, que sería publicada con el título de *El tratado del hombre* (*Le traité de l'homme*) en 1662 por Florentius Schuyt, Descartes concluye que sería posible mediante ingeniería reproducir la apariencia física de cualquier ser vivo de manera perfecta. De los animales, dice, además de la apariencia, podría duplicarse toda su conducta. Y de los seres humanos, casi toda ella, a excepción de dos facultades observables que, por ser exclusivas del alma, no es moralmente posible que se puedan recrear manipulando la materia. Esas dos facultades son el lenguaje natural y la flexibilidad de nuestro intelecto para habérselas con problemas de todo tipo (Descartes, 1637: 92; AT VI: 56).

Lo sorprendente del doble criterio de Descartes para distinguir al ser humano verdadero del autómatas impostor es que, como decimos, todavía sigue vigente, en tanto que la inteligencia artificial (IA) no ha conseguido construir máquinas capaces de hablar el lenguaje natural o de hacer un intento pasable en casi cualquier cosa. Y no se trata de dos accidentes sin importancia, sino que, muy al contrario, son cualidades consideradas de forma unánime por los investigadores de la IA, desde la fundación de esta disciplina a mediados del siglo XX, como



condiciones necesarias para conceder a un sujeto el atributo de la inteligencia en términos antropocéntricos. Por poner un caso significativo que ilustra esta afirmación, Alan Turing, uno de los padres de las computadoras electrónicas, propuso en un artículo de 1950, “*Computing machinery and intelligence*”, una prueba para dirimir si un sujeto es inteligente. Dicha prueba, conocida hoy como el test de Turing, consistía en demostrar una competencia lingüística por escrito indiscernible de la de un ser humano normal (Turing, 1950: 441). Desde entonces, superar el test de Turing ha sido uno de los objetivos prioritarios de la IA.

En cuanto a la flexibilidad de nuestro intelecto para hacer un intento pasable en casi cualquier cosa, según algunos expertos en la materia es la característica distintiva del ser humano (Michie, 1974: 51), y es, sin duda, el otro gran obstáculo para la duplicación perfecta de nuestra conducta. Todos los programas de IA creados hasta la fecha padecen el llamado mal de la limitación de dominio (Crevier, 1993: 250), que quiere decir que esas máquinas son competentes solo en ciertas tareas, más allá de las cuales son incapaces de hacer nada. Un ejemplo célebre es el de Deep Blue, la IA jugadora de ajedrez creada por IBM que en 1997 derrotó al entonces campeón del mundo Gary Kaspárov. Como señala con sorna el conocido físico norteamericano Michio Kaku, Deep Blue ganó, pero al final del encuentro el único que compareció ante la prensa fue Kaspárov, porque Deep Blue no sabía hacer nada aparte de jugar al ajedrez; ni tan siquiera hablar sobre ajedrez (Kaku, 2011: 110). En definitiva, las inteligencias artificiales creadas hasta ahora son al reino de las máquinas lo que los deficientes geniales o síndrome de *savant* a la especie humana: sujetos que realizan unas pocas tareas de forma genial pero que en el resto son deficientes.

Descartes, por tanto, acertó en su pronóstico, y esto es algo que llama la atención de cualquier estudioso de la IA, pues adelantarse cuatrocientos años en la identificación de dos de los principales problemas de esta disciplina es un hecho extraordinario que debe de tener una explicación; la cual quizás podría arrojar luz sobre las razones por las cuales la IA lleva cincuenta años atascada en las mismas grandes cuestiones. La búsqueda de tal explicación es el objetivo del presente artículo. Nuestro trabajo consistirá en reconstruir la quinta parte del *Discurso del método*, la cual citábamos al principio, por ser el texto en el que Descartes señala el doble criterio para distinguir al ser humano del autómatas antropoide. Para profundizar en su comprensión cruzaremos ese fragmento con referencias a otros lugares en los que el filósofo francés describe con detalle su revolucionario



modelo del hombre-máquina, siendo los principales el también antes mencionado *Tratado del hombre* y la primera parte de *Las pasiones del alma*. El resto de la bibliografía del filósofo francés nos será de ayuda igualmente, aunque con menor relevancia. Asimismo, recurriremos a lo que otros han escrito sobre él. De entre ellos nuestro referente principal será André Robinet, filósofo francés experto en cibernética y autor de un ensayo formidable titulado *Le défi cybernétique* sobre el tema del autómeta antropoide.

## §2. FÍSICA Y METAFÍSICA CARTESIANAS

Como es sabido, Descartes divide al ser humano en dos sustancias: la *res extensa*, o sustancia material, y la *res cogitans*, o sustancia pensante. Así lo afirma en la sexta meditación metafísica:

Y aun cuando, acaso, o más bien, ciertamente, como luego diré, tengo yo un cuerpo al que estoy estrechamente unido, sin embargo, puesto que por una parte tengo una idea clara y distinta de mí mismo, según la cual soy algo que piensa y no extenso y, por otra parte, tengo una idea distinta del cuerpo, según la cual este es una cosa extensa, que no piensa, resulta cierto que yo, es decir, mi alma, por la cual soy lo que soy, es entera y verdaderamente distinta de mi cuerpo, pudiendo ser y existir sin el cuerpo (Descartes, 1641: 187; AT VII: 78; AT IX: 62).

En este texto se aprecia con claridad que la unidad que cuerpo y alma conforman es, como señala Ricardo Parellada, una unidad de composición, y no unidad de naturaleza, lo cual quiere decir que sus partes pueden separarse (Parellada, 2000: 237).

El alma puede ser y existir sin el cuerpo, dice Descartes. Y, a la inversa, el cuerpo puede ser y existir sin el alma. Tal es el caso, en el ámbito de los seres vivos, de los animales, que son cuerpos sin alma (Descartes, 1637: 84; AT VI: 46). E incluso en los seres humanos el cuerpo tiene un alto grado de independencia respecto del alma unida a él, tal y como se desprende de la afirmación de que la muerte, según Descartes, no ocurre por la ausencia del alma, sino por una disfunción catastrófica de la que luego se sigue, eso sí, la separación del alma como efecto, pero no como causa (Descartes, 1649: §6; AT XI: 330-331). Los estudios del filósofo francés sobre anatomía y fisiología son una parte muy importante de



su pensamiento, en tanto que la obra principal en la que están plasmados, el *Tratado del hombre*, se considera como una de las más influyentes en la concepción de la psicofisiología humana durante todo el siglo XVII (López-Muñoz y Álamo, 2000: 247). El modelo del hombre-máquina allí descrito supuso una ruptura radical con la teoría aristotélica de la naturaleza humana imperante en la escolástica contemporánea, para proponer en su lugar una visión mecanicista basada en la doctrina neumática de Galeno, de ascendencia estoica.

Según el testimonio del editor coetáneo de Descartes Claude Clerselier, el *Tratado del hombre* fue concebido originalmente como el capítulo XVIII de una obra más amplia titulada *El mundo*. Sin embargo, como ya hemos comentado, Descartes decidió no publicarlo al enterarse, tal y como consta en una carta a su amigo Marin Mersenne, de la condena de la Inquisición a Galileo por herejía. El mundo se divide en dos partes (Descartes, 1633a: 17): Tratado de la luz y Tratado del hombre. En la primera, Descartes admite el movimiento de la Tierra, una afirmación temeraria en aquel contexto político y religioso. En cuanto al *Tratado del hombre*, el filósofo francés postula a modo de hipótesis, sea o no por precaución ante posibles represalias de la Iglesia (Descartes, 1633a: 15), la existencia de unos hombres semejantes a nosotros. Dice de ellos:

No me detendré describiendo los huesos, los nervios, los músculos, las arterias, el estómago, el hígado, el bazo, el corazón, el cerebro ni todas las otras piezas que componen esta máquina. Supongo que son iguales en todo a aquellas partes de nuestro cuerpo que poseen el mismo nombre (Descartes, 1633b: 23; AT XI: 120).

Ese cuerpo en todo semejante al nuestro, que en realidad es el nuestro, tiene un funcionamiento explicable según los dos elementos fundamentales de la teoría física cartesiana, expuestos en los primeros capítulos de *El mundo*, a saber: la materia y el movimiento (Benítez, 2013b: 235). En palabras de Manuel García Morente, en su prólogo a la edición de Austral del *Discurso del método* y las *Meditaciones Metafísicas*:

La física de Descartes es, como todo el mundo sabe, mecanicista; Descartes no quiere más elementos, para explicar los fenómenos y sus relaciones, que la materia y el movimiento. Todo en el mundo es mecanismo, y en la mecánica misma, todo es geométrico (Descartes, 1641: 26).



Por otra parte, Antonio Rodríguez Huéscar, en otro prólogo al *Discurso*, explica por qué el filósofo francés no incluye la fuerza en su física: «Descartes excluye (...) la idea de fuerza porque no es una noción clara y distinta» (Descartes, 1637: 30). Y la claridad y la distinción son, recordemos, los criterios de la verdad según el método cartesiano (Íbid.: 73; AT VI: 33).

### §3. EL HOMBRE-MÁQUINA HIDRÁULICO

Con estos dos elementos, la materia y el movimiento, Descartes comienza el *Tratado del hombre* describiendo a ese hipotético ser antropomorfo como sigue:

Supongo que el cuerpo no es otra cosa que una estatua o máquina de tierra a la que Dios forma con el propósito de hacerla tan semejante a nosotros como sea posible, de modo que no solo confiere al exterior de la misma el color y la forma de todos nuestros miembros, sino que también dispone en su interior todas las piezas requeridas para lograr que se mueva, coma, respire y, en resumen, imite todas las funciones que nos son propias, así como cuantas podemos imaginar que tienen su origen en la materia y solo dependen de la disposición de los órganos (Descartes, 1633b: 22; AT XI: 120).

De las que tienen su origen en la sustancia pensante, como son el lenguaje y la flexibilidad intelectual, nos ocuparemos más adelante.

La manera en que todas las piezas del cuerpo funcionan puede ser entendida, dice Descartes, tomando como modelo el funcionamiento del corazón. Así lo afirma al comienzo de la quinta parte del *Discurso*:

Quiero traer aquí la explicación del movimiento del corazón y de las arterias porque, siendo el primero y el más general que se observa en los animales, podrá juzgarse por él fácilmente lo que se debe pensar de todos los demás (Descartes, 1637: 84; AT VI: 46-47).

Y un poco más adelante resume cómo hay que entender el movimiento de la sangre a través del corazón y las arterias, y por tanto, de todas las piezas del cuerpo: «Este movimiento que acabo de explicar se sigue de la disposición misma de los órganos (...) con la misma necesidad que el movimiento de un reloj se sigue



de la fuerza, situación y figura de sus contrapesos y sus ruedas» (Descartes, 1637: 87; AT VI: 50). La analogía del reloj es constante en la obra de Descartes para explicar el funcionamiento del cuerpo por sí solo, sin tener en cuenta el efecto del alma sobre él. En *Las pasiones del alma* insiste en la metáfora, poniendo los tres mismos ejemplos de acciones: moverse, comer y respirar:

Todos los movimientos que realizamos sin que nuestra voluntad intervenga en ello (como ocurre a menudo cuando respiramos, cuando andamos, cuando comemos y, en fin, cuando ejecutamos todos los actos que nos son comunes con los animales) no dependen más que de la conformación de nuestros miembros y del curso de los espíritus [...] de la misma manera que el movimiento de un reloj es producido únicamente por la fuerza de su resorte y la forma de sus ruedas (Descartes, 1649: §16; AT XI: 341-342).

En este punto hay que aclarar que cuando habla de espíritus Descartes no se refiere a entidades inmateriales, sino a los pneumatas de la antes mencionada doctrina galénica, que son sustancias materiales. Según la teoría de Galeno, la cual se trasluce notablemente en las investigaciones fisiológicas cartesianas (López-Muñoz y Álamo, 2000: 241), por el cuerpo humano circulan unas partículas sutilísimas llamadas pneumatas o espíritus, que son de tres tipos: naturales, vitales y animales. La división se corresponde con la del alma tripartita de Platón, de manera que los espíritus naturales son los asociados al apetito concupiscible, que tiene su sede en el hígado; los espíritus vitales son los propios del apetito irascible, con sede en el corazón y los pulmones, y, por último, los espíritus animales son los relativos al alma racional, o simplemente alma en la filosofía de Descartes, que está emplazada en el cerebro. El pensador francés define en estos términos lo que son los espíritus en fisiología: «Lo que aquí llamo espíritus no son sino cuerpos, y no tienen otra propiedad que la de ser cuerpos muy pequeños y que se mueven muy rápidamente, como las partes de la llama que sale de una antorcha» (Descartes, 1649: §10; AT XI: 335).

Este artículo de *Las pasiones del alma* continúa explicando cómo los espíritus animales salen del cerebro y circulan por el sistema nervioso periférico para mover el resto del cuerpo:

A medida que algunos de ellos entran en la cavidad del cerebro, salen también algunos otros por los poros que hay en su sustancia, los cuales conducen a los



nervios, y de aquí a los músculos, lo que les permite mover el cuerpo (Descartes, 1649: §10; AT XI: 335).

El funcionamiento del autómatas descrito por Descartes es, por tanto, de tipo hidráulico, pues la causa que actúa sobre las partes del cuerpo para moverlas son los espíritus animales, una especie de líquido que se evapora y se rarifica a lo largo de su recorrido circulatorio. Habría de ser el fisiólogo italiano Luigi Galvani un siglo y medio después, a finales del XVIII, quien descubriera la naturaleza eléctrica de la actividad nerviosa (Shepherd, 1988: 4). Hasta entonces, el modelo hidráulico de Descartes fue, como hemos apuntado, uno de los más influyentes.

Para el filósofo francés los nervios son, como señala Guillermo Quintás en una nota de su edición del *Tratado del hombre*, canales de conducción de doble sentido (Descartes, 1633b: 38). Es decir, que el propio nervio tiene funciones tanto de aferencia, o entrada de información sensorial, como de eferencia, o salida de información motora desde el cerebro hasta los músculos. Los nervios tienen tres partes: una cubierta o membrana en forma tubular, una médula o sustancia interior, y espacio para permitir la circulación de los espíritus animales (Descartes, 1649: §12; AT XI: 337). La aferencia se realiza a través de la médula, que tiene forma de hilo y que, al tensarse en un extremo, transmite la señal al otro de manera inmediata, como sucede en una cuerda. La eferencia, en cambio, implica un desplazamiento, pues se realiza por el descrito tránsito de los espíritus animales a través del espacio tubular.

El sistema nervioso, en líneas generales, funcionaría por tanto como una gran fuente muy compleja, de esas que decoraban los jardines del Renacimiento. Junto con la metáfora del reloj, esta es la otra favorita de Descartes. Así, al comienzo del *Tratado del hombre* dice:

En verdad puede establecerse una correcta comparación de los nervios de esta máquina que estoy describiendo con los tubos que forman la parte mecánica de estas fuentes; sus músculos y tendones pueden compararse con los ingenios y los resortes que sirven para moverlas; los espíritus animales con el agua que las pone en movimiento; su corazón con el manantial y, finalmente, las concavidades del cerebro con los registros del agua (Descartes, 1633b: 35; AT XI: 130-131).

Sean relojes o fuentes, lo importante es que se trata de máquinas. Máquinas que, además, tienen un alto grado de autonomía respecto del alma, pues acaba-



mos de ver que Descartes refiere numerosas funciones, como respirar, andar y comer, que el cuerpo puede ejecutar por sí solo, sin participación de la voluntad, que emana del alma. Todos los actos que nos son comunes con los animales, dice, no dependen más que de la disposición material de las partes de la máquina (Descartes, 1649: §16; AT XI: 341).

Como señala André Robinet, la lista de funciones simulables mecánicamente por un cuerpo carente del alma que Descartes redacta al término del *Tratado del hombre* es muy larga (Robinet, 1973: 85):

La digestión de los alimentos, el latido del corazón y de las arterias, la alimentación y crecimiento de los miembros, la respiración, la vigilia y el sueño; la recepción de la luz, de los sonidos, de los olores, de los sabores, del calor y tantas otras cualidades, mediante los órganos de los sentidos exteriores; la impresión de sus ideas en el órgano del sentido común y de la imaginación, la retención o la huella que las mismas dejan en la memoria; los movimientos interiores de los apetitos y de las pasiones y, finalmente, los movimientos exteriores de todos los miembros (Descartes, 1633b: 109; AT XI: 201-202).

En definitiva, todo lo que observamos que hacen los animales puede ser ejecutado por una máquina, porque eso es todo lo que son los animales, e incluso nuestros propios cuerpos por sí solos sin alma: máquinas. La única diferencia entre esas máquinas y las creadas por nosotros es que aquellas son infinitamente más perfectas, por ser obra de Dios (Descartes, 1637: 92; AT VI: 56). No obstante, en tanto que máquinas, no hay objeción de principio contra la posibilidad de duplicarlas mediante la técnica. Lograrlo es cuestión de aplicar el método para descubrir su funcionamiento en términos de ecuaciones matemáticas con las que construir dobles indistinguibles. Es como falsificar un reloj de marca: solo hay que aplicar ingeniería inversa.

#### §4. EL FANTASMA EN LA MÁQUINA

Una vez expuesto el funcionamiento del cuerpo, veamos cómo el alma interacciona con él y de qué funciones es responsable. El alma, dice Descartes, se une al cuerpo a través de una glándula muy pequeña que, por la descripción que da de ella (Descartes, 1649: §31; AT XI: 351-352), no es otra que la glán-



dula pineal, situada en la porción rostral de la cisura de Silvio, que conecta los ventrículos cerebrales tercero y cuarto (López-Muñoz y Álamo, 2000: 250). Su posición es, aproximadamente, el centro geométrico del encéfalo, razón que, sumada a su carácter único frente a la duplicidad de las otras partes del cerebro, dice Descartes, lleva a colegir que se trata del lugar en el que el alma se une al cuerpo. En palabras de Gilbert Ryle, enemigo del dualismo cartesiano, es el sitio por el que el fantasma se introduce en la máquina (Gardner, 1985: 83). La glándula en cuestión está inmersa en la sustancia cerebral, que es la parte más profunda del cerebro. Más hacia afuera se encuentra la superficie interna, que corresponde a las paredes de los ventrículos. Y por último, en el exterior, se distingue la superficie externa, envuelta por la piamadre (López-Muñoz y Álamo, 2000: 248). Todo este sistema de tres partes funciona como un gran centro de distribución de espíritus animales. Estos espíritus, recordemos, fluyen desde el corazón hacia el cerebro. Allí se produce una criba: los más groseros se desvían a la superficie externa, donde sirven de alimento a su sustancia (Descartes, 1633b: 35; AT XI: 130), mientras que los más sutiles penetran en la glándula pineal, y esta los redirige hacia los nervios conectados a unos u otros músculos para producir movimiento.

El movimiento del cuerpo puede ser voluntario o involuntario. Este último se explica sin participación del alma, tan solo por la disposición y la interacción mecánica de los nervios (Descartes, 1649: §13; AT XI: 338), que son capaces de afectar a la glándula pineal en tanto que materia. Descartes pone el ejemplo del reflejo que nos impulsa involuntariamente a cerrar los ojos cuando vemos que la mano de alguien se lanza contra nuestra cara, aun sabiendo que no nos golpeará. No obstante, hay controversia sobre si el genio francés contribuyó o no a denominar y definir el concepto de reflejo. Algunos autores sí se lo conceden (Carpintero, 1996: 101), mientras que otros no (Canguilhem, 1955: 46). En cualquier caso, lo importante e indudable es la descripción mecanicista del movimiento involuntario.

En cuanto al movimiento voluntario surge del alma, sede del libre albedrío (Descartes, 1649: §17; AT XI: 342). El alma, como decimos, está unida al cuerpo a través de la glándula pineal, pero no como un capitán a su navío, según el célebre símil rechazado en la sexta meditación metafísica (Descartes, 1641: 189; AT VII: 81; AT IX: 64), sino que está verdaderamente unida a todo el cuerpo, de manera que «no se puede decir que esté en alguna de sus partes con exclusión de las demás» (Descartes, 1649: §30; AT XI: 351). Descartes describe la glándula



pineal como una especie de almendra porosa que permanece suspendida, flotando entre las cavidades de los ventrículos, debido a la fuerza de empuje vertical de los espíritus animales entrantes. La analogía empleada para explicarlo es, nuevamente, la de un fenómeno físico, en este caso el de un cuerpo suspendido sobre la corriente de aire caliente de un horno (Descartes, 1633b: 90; AT XI: 180). En función de su posición, que puede oscilar levemente, la glándula contactará con unos u otros nervios de la superficie interior del cerebro que la rodea, dando lugar a que los espíritus animales que han entrado en ella fluyan hacia unos u otros músculos, lo cual resulta en la ejecución de diferentes movimientos. El alma actúa sobre el cuerpo a través de la glándula pineal, dice Descartes, como lo haría un fontanero que, situado en el núcleo del complejo sistema de tuberías que mueve la gran fuente que es el cuerpo, puede «provocar, impedir o modificar en cierto modo los movimientos de la fuente» (Descartes, 1633b: 36; AT XI: 131-132).

Es importante reparar en el medio por el cual se comunican cuerpo y alma: el movimiento. Para actuar sobre el cuerpo, el alma mueve la glándula pineal. Así se enuncia en *Las pasiones del alma*: “Toda la acción del alma consiste en que, sólo con querer algo, hace que la pequeña glándula a la que el alma va estrechamente unida se mueva de la manera necesaria para producir el efecto que esa voluntad quiere” (Descartes, 1649: §41; AT XI: 360).

Y, a la inversa, los estímulos sensoriales captados por el cuerpo son transmitidos al alma mediante unos códigos de movimiento: “la pequeña glándula que es la sede principal del alma está de tal modo suspendida entre las cavidades que contienen estos espíritus que puede ser movida por ellos de tantas maneras diferentes como diferencias sensibles hay en los objetos” (Descartes, 1649: §34; AT XI: 354-355).

Como ya dijimos, la física cartesiana se basa en dos principios: materia y movimiento. De manera consistente, por tanto, Descartes considera el movimiento como el medio de comunicación bidireccional entre cuerpo y alma.

## §5. LOS LÍMITES DE LA DUPLICACIÓN ARTIFICIAL

Así descrito el hombre-máquina cartesiano, es momento de abordar ya las potencias exclusivas del alma que permiten distinguir al autómatas antropoide del



verdadero ser humano. Hay que tener siempre presente que Descartes es el padre de la Modernidad, una etapa que en filosofía se caracteriza por comenzar a pensar el universo partiendo del propio hombre. Por eso, dice Robinet, “el discurso filosófico cartesiano, aunque haga largos rodeos, vuelve siempre a la cuestión del “verdadero hombre”, a esta diferencia entre nuestra “naturaleza” y las dos sustancias de las que idealmente estamos compuestos” (Robinet, 1973: 91).

Si el hombre fuese solo un cuerpo, entonces la totalidad de su ser sería explicable en términos de pura física, y quedaría por tanto disuelto en la naturaleza. En tal caso no tendría sentido empezar a pensar por el hombre. El dualismo sujeto-objeto, que tantos problemas ha generado en la Historia de la filosofía, quedaría en cierto sentido resuelto. El espíritu, como señala Horkheimer, sería mero producto de la naturaleza (Horkheimer, 1947: 140). La razón dejaría de ser una capacidad independiente para pasar a convertirse en un producto de la evolución darwiniana, como las antenas o las garras.

Descartes, sin embargo, todavía es muy anterior a Darwin, y por tanto reserva para el hombre un lugar especial en la Creación de Dios. La particularidad que lo hace especial es el alma, sustancia pensante inmortal que sobrevive a la descomposición del cuerpo (Descartes, 1637: 95; AT VI: 60). En el *Discurso* se ofrecen dos pruebas de la existencia del alma. Una es la que figura en la cuarta parte tras proclamar el famoso «pienso, luego existo». Allí dice: dado que puedo pensar que existo sin cuerpo y sin que existiese lugar alguno en el espacio, es evidente que soy «una sustancia cuya completa esencia o naturaleza consiste solo en pensar, y que para existir no tiene necesidad de ningún lugar ni depende de ninguna cosa material» (Descartes, 1637: 72; AT VI: 33). Esta es una prueba a priori, pues no depende de la experiencia. La segunda prueba de la existencia del alma es, en cambio, a posteriori, y consiste en estudiar las diferencias entre el hombre y los otros seres vivos. Cuando se conoce, tras un examen empírico, cuán grandes son las diferencias entre nuestra alma y la supuesta alma de las bestias, dice, entonces «se comprenden mucho mejor las razones que prueban que la nuestra es de una naturaleza enteramente diferente del cuerpo» (Descartes, 1637: 95; AT VI: 59).

Las diferencias son las dos citadas al comienzo de este artículo: el lenguaje natural y la flexibilidad del intelecto para acometer todo tipo de tareas (Descartes, 1637: 92; AT VI: 56). En tanto que el animal no es más que un autómatas, ambas diferencias sirven también, dice Descartes, para distinguir al verdadero ser humano del autómatas antropoide. Empezando por el lenguaje, el filósofo



francés observa que hay animales y, por tanto, también puede haber máquinas fabricadas por el hombre que profieran palabras como respuesta a ciertos estímulos. Pueden expresar, por ejemplo, una sensación de dolor mediante un grito. Se concede, así pues, que los animales y las máquinas pueden expresar sus pasiones, es decir, comunicar a otros un cierto estado general interno, pero esto no debe confundirse con las palabras, dice Descartes, porque se trata de reacciones simples. El verdadero lenguaje, en cambio, se caracteriza por combinar «esas palabras de distintas maneras para responder al sentido de todo lo que se diga en su presencia» (Descartes, 1637: 93; AT VI: 57). Se trata, por tanto, de un problema de combinatoria. Ni siquiera los animales mejor dotados para el lenguaje, como sería el caso de los monos más listos o de los loros más perfectos de su especie, dice Descartes, pueden igualar la capacidad de expresión lingüística de un ser humano con el cerebro perturbado, o que haya nacido sordo o mudo, pues se observa que los hombres así discapacitados a menudo inventan por sí mismos lenguajes de signos con los que hacerse entender (Descartes, 1637: 94; AT VI: 58). La diferencia es insuperable.

Respecto a la flexibilidad del intelecto para acometer todo tipo de tareas, la causa de que los animales no puedan poseerla es la misma. Es cosa muy de notar, señala el filósofo francés, que hay muchos animales mejor dotados que nosotros para realizar tareas muy particulares. Por ejemplo, los castores levantan diques en el agua sin haber aprendido a hacerlo observando a sus semejantes, y las cigüeñas construyen nidos que aguantan el paso del tiempo mejor que algunos edificios. Otros ejemplos válidos, dado que también son máquinas, serían Deep Blue, que como ya hemos dicho jugaba al ajedrez mejor que el mejor ajedrecista humano de su tiempo, y R1, un sistema experto de principios de los años ochenta que configuraba computadoras a un ritmo imposible para un ser humano (Jackson, 1986: 137). Sin embargo, dice Descartes, no hay ningún animal que, fuera de su especialidad, demuestre tener la capacidad del hombre (Descartes, 1637: 94; AT VI: 58). La causa es, nuevamente, la combinatoria: dado que sus movimientos se producen de acuerdo a la disposición de sus órganos, de manera semejante a como funcionan los relojes, es moralmente imposible, dice, que pueda haber una máquina que tenga sus disposiciones internas «en número suficiente para permitirle obrar en todas las ocurrencias de la vida de la misma manera que nuestra razón nos lo permite» (Descartes, 1637: 93; AT VI: 57). El problema de fondo, por tanto, es el de la cantidad de las configuraciones adoptables. Descartes no ve



moralmente posible que pueda haber una máquina universal capaz de adoptar cualquier configuración interna.

El alma, a través de la glándula pineal, añade a la máquina que es el cuerpo una pieza mágica, capaz de trascender las limitaciones de la materia: la universalidad (Descartes, 1637: 93; AT VI: 57). Mientras que las configuraciones adoptables por una máquina son finitas, el alma es un instrumento universal, capaz de infinitas combinaciones. Ahora bien, detengámonos en el importantísimo adverbio que acompaña a la imposibilidad de que una máquina pueda obrar de manera universal. Ese adverbio es moralmente; en el original, en francés, *moralement* (AT VI: 57). Como señala Robinet, aquí hay una debilidad en el argumento: “Si ese “moralmente” significa “según toda verosimilitud”, el esfuerzo de investigación por vía de “modelo” conduce a una satisfacción cierta, aunque provisoria, puesto que es solamente verosímil” (Robinet, 1973: 92).

El padre del racionalismo moderno despliega en la cuarta parte del *Discurso* una batería de pruebas irrefutables, según él, de la existencia del alma y de Dios. En cambio, cuando en la quinta parte quiere trazar una línea divisoria entre el hombre y el resto de la Creación para garantizarle a aquel un lugar distinguido, no encuentra más que un argumento verosímil. El autómatas antropoide indistinguible de la obra más perfecta de Dios es, en definitiva, solo moralmente imposible. Y Robinet se pregunta: «¿No podría provenir esa universalidad más amplia de respuestas racionales de un refuerzo en la diversidad de disposiciones que animan al autómatas?» (Robinet, 1973: 92).

## §6. LA FORMALIZACIÓN DEL UNIVERSO

Descartes, obviamente, se equivocó. Hoy sabemos que el cerebro es una máquina con una capacidad combinatoria de sus disposiciones que, si no infinita, por lo menos sí es suficiente para el lenguaje y para acometer cualquier tipo de problema. El error del genio francés se justifica por dos factores contextuales. Uno, el científico, ya que el conocimiento del universo a pequeña escala era muy precario en su época. El modelo atómico del siglo XVII seguía siendo el impreciso –y más filosófico que científico– de Demócrito, situación que se mantendría hasta que John Galton publicase en 1808 *A new system of chemical philosophy*. En el ámbito de la vida microscópica habría que esperar hasta 1676, algo después



de la muerte de Descartes, a que Anton van Leeuwenhoek propusiese la teoría microbiana de la enfermedad. Hasta esa fecha hay que recordar que se creía en la teoría de la generación espontánea, según la cual la vida animal y vegetal surge sin más de la materia inorgánica, así como en la teoría miasmática de la enfermedad, que afirmaba que ciertas dolencias como la peste se debían al contacto con unos miasmas o gases emanados de sustancias impuras. En un marco científico semejante, Descartes no tenía los conocimientos que le hubieran permitido creer en la posibilidad verosímil de que en el cerebro hubiese una maquinaria de “fontanería” combinatoria tan compleja como para albergar 1.011 neuronas y 1.014 conexiones entre ellas (Kandel, Schwartz y Jessell, 1995: 183).

El otro factor contextual que explica el error de Descartes es la ascendencia que tenían sobre él sus convicciones religiosas. Educado por los jesuitas en el colegio de La Flèche, Descartes era católico, y como tal necesitaba moralmente darle al hombre un estatuto metafísico distinto del resto de las cosas de la Creación. Mientras que su parte de hombre moderno y revolucionario proponía un modelo mecánico del cuerpo contrario a la doctrina escolástica, su parte de hombre temeroso de Dios lo obligaba a dotar a esa máquina de algo especial que ninguna otra máquina pudiera tener. Ese algo es el alma, y sus facultades exclusivas, derivadas de su potencia universal, son el lenguaje y la flexibilidad del intelecto. El propio Descartes estaba así tan dividido en dos partes como el modelo del hombre que elaboró. La dualidad del hombre-máquina es reflejo de la dualidad de su creador, un hombre dividido entre la ciencia y la fe.

En la actualidad, además de saber que el cerebro es la causa única de toda conducta en virtud de su compleja configuración resultante de la interacción de factores biológicos y ambientales, tenemos también unas máquinas capaces de simular el funcionamiento del propio cerebro, así como de cualquier otra parcela del universo, entendiéndolo por tal la realidad física. Estas máquinas son las computadoras electrónicas. Una computadora electrónica consta de dos grandes partes: la material, denominada *hardware*, y la formal, denominada *software*. La más importante para lo que aquí nos interesa es esta última, pues el *hardware* es ampliamente variable, y su única función es servir para ejecutar el *software*. Tanto es así que el conocido como principio de *realizabilidad* múltiple (Copeland, 1993: 128) establece que un *software*, o programa, es ejecutable por cualquier computadora con independencia de cuál sea la naturaleza física del *hardware*: da igual que sea electrónica con circuitos de silicio como las actuales, o que sea



un computador humano como los del siglo XVIII (Guijarro y González, 2010: 198), o una enorme máquina de acero compuesta de tuercas y engranajes como la máquina analítica que no pudo terminar el ingeniero inglés Charles Babbage a finales del siglo XIX (Guijarro y González, 2010: 310), o incluso una máquina de agua construida con tuberías, a semejanza de las fuentes descritas por Descartes.

Pues bien, todo programa de computadora, cualquiera que sea su soporte material, es siempre, en esencia, un sistema formal automático (Haugeland, 1981: 5). Empezando por los dos primeros términos, un sistema formal es un constructo teórico que se compone de dos partes: lenguaje formal y mecanismo deductivo (Falguera y Martínez, 1999: 62). Un lenguaje formal es un lenguaje exento de interpretación semántica en su definición (Falguera y Martínez, 1999: 61), y se divide en un vocabulario y un conjunto de reglas que establecen cómo han de combinarse los símbolos del vocabulario para formar expresiones correctas. En cuanto al mecanismo deductivo, consiste en esencia en un conjunto de reglas que indican cómo transformar unas expresiones en otras. Por último, que un sistema formal es automático quiere decir que manipulará símbolos de forma autónoma atendiendo a las reglas del sistema, como por ejemplo hace una IA jugadora de ajedrez (Haugeland, 1981: 10).

En tanto que sistemas formales, las computadoras, entendidas siempre en el sentido de *software*, son máquinas capaces de simular el funcionamiento de cualquier fenómeno físico, una tesis que acabamos de señalar y que procedemos a demostrar. La razón de esta potencia es que las ciencias naturales, tal y como fueron formuladas en su sentido moderno por científicos como Descartes y Galileo, suponen por principio que el universo es un ente calculable. A esto se refiere Habermas cuando caracteriza las ciencias empírico-analíticas –que es como él denomina a las ciencias naturales– de la manera siguiente. La cita es un poco larga, pero resulta imprescindible para justificar nuestra tesis de que las computadoras electrónicas son capaces de simular cualquier fenómeno del universo físico. En primer lugar, Habermas define las condiciones trascendentales del conocimiento científico natural:

En la esfera funcional de la actividad instrumental, la realidad se constituye como la suma de lo que puede ser experimentado bajo el punto de vista de la manipulación técnica posible: a la realidad objetivada en condiciones trascendentales corresponde una experiencia restringida (Habermas, 1968b: 194).



A continuación, da cuenta del modo en que se explican los fenómenos que acontecen en dicha experiencia:

Bajo las mismas condiciones se configura también el lenguaje de los enunciados empírico-analíticos sobre la realidad. Las proposiciones teóricas pertenecen a un lenguaje o bien formalizado o por lo menos formalizable. Según su forma lógica, se trata de cálculos que podemos generar y reconstruir en todo momento manipulando unos signos según ciertas reglas. Bajo las condiciones de la acción instrumental, un lenguaje puro se constituye como conjunto de tales complejos simbólicos que pueden ser producidos operando según reglas. (...) Ambos, el lenguaje restringido y la experiencia restringida, vienen definidos por el hecho de que son resultados de operaciones con signos o con cuerpos móviles. Igual que la propia actividad instrumental, también el uso lingüístico integrado en ella es nomológico. Asegura a las proposiciones teóricas una cohesión sistemática regida por las reglas de la deducción (Habermas, 1968b: 195).

Así pues, dado que las ciencias naturales, que se ocupan de explicar el universo en su dimensión puramente física, proceden conforme a una noción restringida de la realidad en la cual solo se contempla aquello que puede ser manipulado en forma de signos mediante el cálculo de proposiciones formales o formalizables, y en tanto que, por otra parte, las computadoras son precisamente sistemas automáticos de cálculo de proposiciones formales, la conclusión es que no hay fenómeno natural alguno, por definición, que no pueda ser simulado computacionalmente. Por eso Heidegger pronosticó con acierto que la cibernética se convertiría en la nueva ciencia fundamental que desplazaría a la ontología.

Como estamos hablando de Descartes, es pertinente señalar su contribución al establecimiento de la ciencia moderna sobre estas bases. Recordemos que en su física el universo es sustancia extensa (Descartes, 1641: 140; AT VII: 31; AT IX: 24) en un espacio cuantificable, y por tanto lo relativo a ella puede ser descrito en ecuaciones matemáticas. El universo en sí mismo, ciertamente, no es computable, sino que lo computable es una determinada visión de este, la moderna, que se fundamenta en el prejuicio de la calculabilidad. A este respecto, comentando el análisis de Husserl de la estructura socio-histórica de la razón científica, Marcuse señala que las nociones de la exactitud y la fungibilidad matemática en las que se basa la ciencia moderna no garantizan la verdad objetiva, sino que «en-



vuelven una experiencia concreta específica de la Lebenswelt: un modo específico de “ver” el mundo» (Marcuse, 1964: 151).

Una computadora electrónica, por tanto, puede simular el funcionamiento del cerebro, y en consecuencia puede también producir una conducta indistinguible de la de un ser humano, incluyendo el lenguaje y la flexibilidad del intelecto. A esta tarea se dedica la corriente de investigación de la inteligencia artificial denominada IA subsimbólica o conexionismo. Los signos manipulados por los sistemas formales en que consisten sus programas se refieren a entidades anatómicas y procesos fisiológicos inspirados, con mayor o menor realismo, en las redes de neuronas del cerebro, tales como sinapsis, potenciales, umbrales de disparo, etc.

Cierto es que este tipo de inteligencias artificiales todavía no es capaz del lenguaje y de la flexibilidad reclamadas por Descartes como facultades exclusivas del alma y, por tanto, el pronóstico del filósofo francés es acertado, pero es preciso reparar en que su acierto respecto a la IA subsimbólica es solo temporal, pues hay que insistir en que el cerebro es la causa de toda conducta humana, y su funcionamiento, en tanto que entidad física, puede ser duplicado en principio por una computadora electrónica (Hawkins y Blakeslee, 2004: 34). El cerebro en particular, y el cuerpo humano en general, es, desde el punto de vista de la ciencia moderna, tan *simulable* como el ala de un avión: lo único que hace falta es encontrar el modelo matemático adecuado.

## §7. UNIVERSO Y MUNDO

Bien distinto es, sin embargo, el caso de la IA simbólica, la otra gran corriente de investigación de esta disciplina, pues el pronóstico de Descartes sobre ella sí que es acertado de manera definitiva. Mientras que en la IA subsimbólica, como acabamos de señalar, los programas informáticos manipulan signos que refieren entidades del universo, es decir, de la dimensión física de la realidad, los programas de IA simbólica manipulan signos que refieren entidades del mundo, esto es, de la dimensión social. Con más precisión, definiremos el mundo en los términos en que lo hace Heidegger, según una formulación de Ramón Rodríguez: el mundo es «la totalidad de relaciones de sentido, que forma parte de la propia existencia humana, y que actúa como el trasfondo a partir del cual



una cosa determinada puede ser lo que es» (Rodríguez, 1987: 214). Creer que el mundo es formalizable y calculable como si fuera una parcela más del universo es el prejuicio más profundo de la IA simbólica, y es la causa de buena parte de su plausibilidad (Dreyfus, 1992: 213). Hubert Dreyfus, quien quizás sea el más célebre detractor de la IA simbólica, lo denomina supuesto ontológico, y lo define así: «El supuesto ontológico [establece] que todo lo que es esencial para la conducta inteligente debe en principio ser comprensible en términos de un conjunto de elementos independientes determinados» (Dreyfus, 1992: 206). Así es como procede la ciencia empírico-analítica: descomponiendo el universo en elementos independientes determinados, ya sean símbolos o proposiciones.

Las computadoras no tienen problemas con el universo, pero en cambio, al ser el supuesto ontológico falso, son incapaces de habérselas con el mundo. Pongamos un ejemplo significativo: el reconocimiento de objetos. Las máquinas, dirigidas por computadoras, son capaces de detectar que están ante un objeto del universo físico. Gracias a sofisticados sistemas de radar, pueden representar-se la forma de esos cuerpos para esquivarlos y no chocar con ellos. Incluso son capaces de distinguir de qué objetos se trata; pero eso sí, solo a nivel físico. Así, por ejemplo, las sondas de exploración enviadas a otros planetas están equipadas con espectroscopios capaces de identificar si están sobre un terreno granítico, basáltico o de cualquier otro tipo. En cambio, la identificación a nivel mundano sigue siendo una prueba de fuego que se les resiste. Distinguir entre el carbono y el hidrógeno es fácil para las máquinas, pero distinguir entre una silla y una mesa suele resultarles difícil.

La razón de esta asimetría radica en que los objetos del mundo no se caracterizan habitualmente por rasgos empíricos invariables. Interpretar que un objeto es una silla, e identificarlo como tal, no depende de sus partes físicas, dice Dreyfus, sino que significa entender su relación con otros objetos y con los seres humanos:

Esto implica todo un contexto de actividad humana del cual la forma de nuestro cuerpo, la institución social de los muebles, y la inevitabilidad de la fatiga constituyen solo una pequeña parte. Y estos factores a su vez no son más aislables que la silla. Todos adquieren su significado en el contexto de la actividad humana de la que forman parte (Dreyfus, 1992: 210).

Mientras que en el universo físico la realidad es descomponible atómicamente en elementos simples, en el mundo no hay tales partes constituyentes últimas. En



contra de la teoría propuesta en el *Tractatus*, Wittgenstein se dio cuenta, al alcanzar la madurez, de lo absurdo que es hablar de elementos atómicos en el mundo:

¿Pero cuáles son las partes constituyentes simples de las que se compone la realidad? –¿Cuáles son las partes constituyentes simples de una silla?– ¿Los trozos de madera con los que está ensamblada? ¿O las moléculas, o los átomos? –“Simple” quiere decir: no compuesto. Y aquí surge luego: ¿‘compuesto’ en qué sentido? No tiene ningún sentido hablar absolutamente de ‘partes constituyentes simples de la silla’. [...] A la pregunta filosófica: “¿Es la figura visual de este árbol compuesta, y cuáles son sus partes constituyentes?”, la respuesta correcta es: “Eso depende de qué entiendas por ‘compuesto’”. (Y esta no es naturalmente una contestación sino un rechazo de la pregunta) (Wittgenstein, 1953: §47).

El propio Wittgenstein creía, en su primera etapa, que el mundo entero es descomponible en proposiciones atómicas. Así lo dice en el *Tractatus*:

La especificación de todas las proposiciones elementales verdaderas describe el mundo completamente. El mundo queda completamente descrito por la especificación de todas las proposiciones elementales más la especificación de las que de ellas son verdaderas y de las que de ellas son falsas (Wittgenstein, 1921: §4.26).

La realidad es que el mundo no se agota en ningún conjunto de representaciones proposicionales y, en consecuencia, el supuesto ontológico es falso. Aunque también se le puede dar la vuelta al razonamiento: dado que la IA simbólica lleva desde su fundación, en los años cincuenta, atascada en los problemas que el propio supuesto ontológico genera, tales como el problema de la cualificación (McCarthy y Hayes, 1969: 34) o el problema de la pertinencia (Dreyfus, 1979: 200), el supuesto ontológico no es verosímil.

Antes de pasar a la sección siguiente, conviene señalar que aquí hemos argumentado la insuficiencia de la IA simbólica para reproducir la inteligencia humana desde el punto de vista de la epistemología, pero no es este el único posible. Otro enfoque muy interesante, pero al que hemos renunciado por razones de espacio, es el de la lógica. Al lector interesado en este le recomendamos acercarse a los textos clásicos de Church y Turing sobre el problema de la decisión (*Entscheidungsproblem*) (Church, 1936; Turing, 1936), por señalar ambas limitaciones formales de las computadoras, y de los que, por cierto, ya nos ocupamos



hace poco en otra parte (Carabantes, 2016: 72-81). También cabe destacar al respecto el teorema de incompletitud de Gödel, que demuestra la existencia de sistemas deductivos incompletos (Gödel, 1931).

#### §8. UN TESORO PARA PAGAR UNA DEUDA

El pronóstico de Descartes sobre los problemas de la IA es, como decimos, acertado de manera definitiva respecto a la IA simbólica. No obstante, su enfoque sobre el funcionamiento de la máquina que hay dentro del cráneo no es simbólico, sino subsimbólico, y en este, como ya hemos dicho, el acierto es solo provisional. Que el enfoque del filósofo francés sobre el cerebro es subsimbólico se observa en muchos lugares de su obra. Por ejemplo, en el *Tratado del hombre* dice que los objetos, ya sean percibidos o rememorados, se descomponen en patrones de activación de los nervios (Descartes, 1633b: 85; AT XI: 175), que es lo característico de la IA subsimbólica (Smolensky, 1989: 239). En consecuencia, la conclusión es que Descartes pronosticó con acierto dos problemas irresolubles de la IA simbólica, pero no porque entendiésemos las causas de un enfoque de la inteligencia, el simbólico, que ni siquiera contemplaba en su modelo mecanicista del cuerpo humano. Su acierto, en tanto que ajeno al conocimiento de las causas, fue solo una casualidad. Lo encontró como aquel que, en su afán de hallar un tesoro, «vagabundea (...) buscando si por casualidad encontraba alguno perdido por algún viajero» (Descartes, 1628: r. IV; AT X: 371).

Lo que impulsó a Descartes a reservar el lenguaje y la flexibilidad del intelecto como facultades exclusivas del alma y no duplicables mediante ningún mecanismo fue, como hemos señalado antes, su fe religiosa. Ante el terreno oscuro, desconocido en aquella época, sobre cómo funciona el universo a escala microscópica, lo que Descartes debería haber hecho es aplicar el método. Pero, en lugar de eso, cayó en el error de permitir que su voluntad, condicionada por el deseo de justificar racionalmente sus creencias religiosas, se extendiera más allá de los límites de su entendimiento. Esta es la causa del error, tal y como él mismo dice en la meditación cuarta: «(Mis errores) nacen de que la voluntad, siendo mucho más amplia y extensa que el entendimiento, no se contiene dentro de los mismos límites, sino que se extiende también a las cosas que no comprendo» (Descartes, 1641: 168; AT VII: 58; AT IX: 46). Sin comprender



cómo funcionaba el universo microscópico, su voluntad de cristiano lo condujo a enunciar un juicio precipitado (Descartes, 1637: 57; AT VI: 15). Ese juicio fue: es moralmente imposible que pueda haber una máquina que tenga sus disposiciones internas «en número suficiente para permitirle obrar en todas las ocurrencias de la vida de la misma manera que nuestra razón nos lo permite» (Descartes, 1637: 93; AT VI: 57).

Desde el principio de su pensamiento, Descartes contrae una deuda con Dios, y es en la imperfección insalvable de los autómatas donde la paga. En su búsqueda de la certeza, la primera verdad hallada, que es la de la propia existencia, deja al filósofo francés encerrado en un solipsismo insuperable. Yo pienso, luego existo; es cierto, pero de ahí es imposible dar un paso hacia afuera y demostrar con igual certeza que existe algo más. A Descartes el *cogito* se le queda “cojito”, dice el provocador filósofo cristiano Carlos Díaz: «un bípedo renqueante, agujero negro capaz de absorber dentro todo excepto un alter ego» (Díaz 1998: 79). Es Dios el único que puede permitirle trascender del yo a lo otro. En tanto que sumamente bueno y absolutamente veraz (Descartes, 1637: 78; AT VI: 40), Dios es el garante de que la inclinación a creer que existen cosas materiales es verdadera (Descartes, 1641: 188; AT VII: 80; AT IX: 63). En las *Meditaciones* contempla, por un instante, la posibilidad de que las ideas recibidas por los sentidos sean causadas por una cierta facultad propia desconocida para uno mismo (Descartes, 1641: 186; AT VII: 77; AT IX: 61), pero enseguida se apresura, dos páginas después, a sentenciar que tal engaño no es posible, porque si fuera el caso, entonces Dios no tendría disculpa (Descartes, 1641: 188; AT VII: 80; AT IX: 63), pues estaría permitiendo un engaño sobre algo de lo que no tenemos ningún medio para conocer la verdad, lo cual es imposible por su definición de ente perfectísimo. Esto le debe Descartes a Dios: la salida del yo a lo otro.

El pago a cambio es el reconocimiento de la *imperfectibilidad* insalvable (Robinet, 1973: 38). El hombre puede reproducir de manera perfecta, es decir, indistinguible, casi cualquier cosa en la naturaleza creada por Dios. Es el caso, por ejemplo, de los animales (Descartes, 1637: 92; AT VI: 56), seres de un alto grado de perfección, ya que fueron creados en los días quinto y sexto (Gn 1, 20-25) en un proceso que es de perfección creciente. De ahí para abajo, la duplicación total es posible. Pero hacia arriba, no, porque más arriba solo está el ser humano, cúspide del Génesis, creado al final del sexto día. Conceder que el hombre pudiera duplicarse a sí mismo de manera perfecta supondría que este se igualase



al Creador a través de la acción. «Por sus obras los conoceréis», dice Mateo el Evangelista (Mt 7, 16). En los mitos de la duplicación del hombre por el hombre la *imperfectibilidad* insalvable es constante: el golem carecía de inteligencia, el homúnculo de Goethe carecía de cuerpo. Siempre falta algo. Para Descartes, faltarían el lenguaje y la flexibilidad intelectual. La deuda queda así saldada: se reconoce la superioridad inigualable del Creador.

Al autómeta antropoide le faltarían también otras cosas, tales como la voluntad (Descartes, 1649: §17; AT XI: 342) y la inmortalidad del yo (Descartes, 1641: 121; AT VII: 14; AT IX: 10), atributos exclusivos del alma. Sin embargo, a Descartes no le habría bastado con reservar estas dos propiedades. La razón es que no son intersubjetivamente observables, es decir, que desde fuera del propio sujeto no habría manera de saber si se trata de un autómeta o de un auténtico ser humano, lo cual generaría una situación de incertidumbre que el filósofo francés rechaza por principio. Ya lo hemos visto a propósito de la existencia del mundo empírico: dado que no hay manera de dirimir si existe o no, tiene que ser verdadero en virtud de la infinita bondad de Dios, que es incompatible con el engaño. En el caso del autómeta impostor (Descartes, 1641: 141; AT VII: 32; AT IX: 25), si no hubiese forma de distinguirlo, entonces Dios estaría condenando al hombre a vivir en la mentira. Por tanto, para Descartes la cuestión es no sólo que haya una diferencia insalvable entre el hombre y la máquina, sino que además debe servir de criterio discriminatorio. El filósofo francés es astuto, y como pago de la deuda no entrega cualquier cosa, sino lo que conviene a su aspiración a la certeza: dos facultades observables.

Ciertamente, la demostración de que el lenguaje y la flexibilidad del intelecto no son duplicables por medios materiales es algo que se echa en falta. En el *Discurso* menciona que ya la ofreció al final de *El mundo*, después del capítulo XVIII, que es el constituido por el *Tratado del hombre*. Dice: «Después de esto, había descrito yo el alma razonable y hecho ver que no puede ser sacada en modo alguno de la potencia de la materia como las otras cosas de que había hablado» (Descartes, 1637: 95; AT VI: 59). Sin embargo, no hizo tal cosa, o por lo menos esos manuscritos no llegaron a Clerselier (Descartes, 1633: 52), como tampoco se conservaron los capítulos XVI y XVII de la misma obra, *El mundo*. En cualquier caso, era moralmente necesario proclamar la existencia de esa tal demostración, aunque en realidad fuese falaz, porque, como venimos argumentando, Descartes necesitaba criterios para distinguir el autómeta del verdadero ser humano. Sus



pretensiones epistemológicas determinan, así, su modelo ontológico de la realidad. Todo queda subordinado a su deseo de saber.

## §9. CONCLUSIÓN

Casi cuatro siglos después de la muerte del padre de la modernidad, la ciencia que él contribuyó a fundar enfoca la luz natural del entendimiento hacia el cerebro con el objetivo de reproducirlo por completo, incluyendo los misterios de lo que él creía que se ocultaba en la glándula pineal. La actitud, sin embargo, es en el fondo la misma: se subordina la ontología a la epistemología. Así, los investigadores del cerebro y de la inteligencia artificial subsimbólica rechazan la ontología del dualismo de sustancias en favor de un enfoque epistemológico *fisicalista*, en tanto que es el único que les garantiza la posibilidad de saberlo todo sobre la sustancia pensante. Observemos la simetría inversa: Descartes renuncia a la explicación científica, en sentido técnico-instrumental, del lenguaje y de la flexibilidad del intelecto para, a cambio de pagar ese precio, poder distinguir el original de la copia. Muy al contrario, la ciencia de nuestro tiempo aspira a saberlo todo sobre la sustancia pensante, lo cual conduce a no poder distinguir, en última instancia, el original de la copia, pues recordemos que uno de los grandes objetivos de la IA es que la máquina supere el test de Turing. En ambos casos, el conocimiento total es imposible. La cuestión es a qué se decide renunciar. El signo de nuestro tiempo es la identidad entre el original y la copia, como bien vio Walter Benjamin, y por tanto es lógico que se sacrifique la discernibilidad del impostor. Las consecuencias fueron exploradas en profundidad por Philip K. Dick, otro genio tan obsesionado como Descartes por la diferencia entre lo real y lo ilusorio.

Así como hemos señalado antes que Descartes estaba condicionado por sus prejuicios religiosos, sería justo apuntar que la ciencia actual se desarrolla bajo la hipóstasis del concepto fundamental de la Ilustración: la identidad del formalismo (Horkheimer y Adorno, 1941: 80). Pero este es un asunto que se aleja demasiado del tema del presente artículo, y tiene que ver más con nuestros pensamientos. Unos pensamientos que son deudores de la escuela de Frankfurt, la cual piensa las consecuencias de la ciencia moderna tal y como fue definida por algunos hombres del Renacimiento entre los que Descartes ocupa un lugar destacado.



## REFERENCIAS BIBLIOGRÁFICAS

- Benítez, A. (2011a). *Fundamentos de inteligencia artificial. Libro primero: Programación en Scheme*. Madrid: Escolar y Mayo.
- Benítez, A. (2011b). *Fundamentos de inteligencia artificial. Libro segundo: Inteligencia artificial clásica*. Madrid: Escolar y Mayo.
- Benítez, A. (2013a). *Fundamentos de inteligencia artificial. Libro tercero: Inteligencia artificial bioinspirada*. Madrid: Escolar y Mayo.
- Benítez, A. (2013b). *Lógica*. Madrid: Escolar y Mayo.
- Canguilhem, G. (1955). *La formation du concept de réflexe aux XVIIe et XVIIIe siècles*. París: PUF [Seguimos la traducción de Juan Rovira: *La formación del concepto de reflejo en los siglos XVII y XVIII*. Barcelona: Avance, 1975].
- Carabantes, M. (2016). *Inteligencia artificial: Una perspectiva filosófica*. Madrid: Escolar y Mayo.
- Church, A. (1936). An unsolvable problem of elementary number theory, *American Journal of Mathematics*, 58(2), 345-363.
- Copeland, B. J. (1993). *Artificial intelligence: A philosophical introduction*. Oxford: Blackwell [Seguimos la traducción de Julio César Armero San José: *Inteligencia artificial: Una introducción filosófica*. Madrid: Alianza, 1996].
- Cottingham, J., Stoothoff, R. & Murdoch, D. (eds. & trans.) (1984a). *The philosophical writings of Descartes*, vol. I. Cambridge: Cambridge University Press.
- Cottingham, J., Stoothoff, R., & Murdoch, D. (eds. & trans.) (1984b). *The philosophical writings of Descartes*, vol. II. Cambridge: Cambridge University Press.
- Cottingham, J., Stoothoff, R., Murdoch, D. & Kenny, A. (eds. & trans.) (1991). *The philosophical writings of Descartes*, vol. III. Cambridge: Cambridge University Press.
- Crevier, D. (1993). *The tumultuous history of the research for AI*. New York: Basic Books.
- Descartes, R. (1628). *Regulae ad directionem ingenii* [Seguimos la edición con traducción y prólogo de Francisco de P. Samaranch: *Discurso del método & Reglas para la dirección de la mente*. Barcelona: Orbis, 1983].
- Descartes, R. (1633a). *Traité du monde et de la lumière* [Seguimos la edición con traducción y prólogo de Salvio Turró: *El mundo. Tratado de la luz*. Madrid: Anthropos, 1989].



- Descartes, R. (1633b). *Le traité de l'homme* [Seguimos la edición con traducción y prólogo de Guillermo Quintás: *El tratado del hombre*. Madrid: Alianza, 1990].
- Descartes, R. (1637). *Discours de la méthode* [Seguimos la edición con traducción y prólogo de Antonio Rodríguez Huéscar: *Discurso del método & Reglas para la dirección de la mente*. Barcelona: Orbis, 1983].
- Descartes, R. (1641). *Meditationes de prima philosophia* [Seguimos la edición con traducción y prólogo de Manuel García Morente: *Discurso del método & Meditaciones metafísicas*. Madrid: Espasa, 2001].
- Descartes, R. (1649). *Les passions de l'âme* [Seguimos la edición con prólogo de José Antonio Míguez y traducción de Consuelo Bergés: *Meditaciones metafísicas & Las pasiones del alma*. Madrid: Orbis, 1985].
- Descartes, R. (1996). *Oeuvres complètes* (nouvelle édition). París: Publiées par Charles Adam et Paul Tannery [Abreviado AT].
- Díaz, C. (1998). *Apología de la fe inteligente*. Bilbao: Desclée De Brouwer.
- Dick, Ph. K. (1968). *Do androids sleep of electric sheep?* New York: Doubleday [Seguimos la traducción de César Terrón: *Blade Runner: ¿Sueñan los androides con ovejas eléctricas?* Barcelona: Edhasa, 1992].
- Dreyfus, H. L. (1979). From micro-worlds to knowledge representation: AI at an impasse. En J. Haugeland (ed.), *Mind design* (pp. 161-204). Cambridge: The MIT Press.
- Dreyfus, H. L. (1992). *What computers still can't do*. Cambridge: The MIT Press.
- Falguera, J. L. & Martínez, C. (1999). *Lógica clásica de primer orden*. Madrid: Trotta.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books [Seguimos la traducción de Leandro Wolfson: *La nueva ciencia de la mente: Historia de la revolución cognitiva*. Barcelona: Paidós, 1988].
- Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik* (38), 173-198.
- Guijarro, V. & González, L. (2010). *La quimera del autómatas matemático*. Madrid: Cátedra.



- Habermas, J. (1968a). *Technik und Wissenschaft als «Ideologie»* [Seguimos la traducción de Manuel Jiménez y Manuel Garrido: *Ciencia y técnica como ideología*. Madrid: Tecnos, 1984].
- Habermas, J. (1968b). *Erkenntnis und Interesse*. Frankfurt am Main: Suhrkamp Verlag [Seguimos la traducción de Manuel Jiménez, José F. Ivars & Luis Martín Santos: *Conocimiento e interés*. Madrid: Taurus, 1990].
- Haugeland, J. (1981). Semantic engines: An introduction to mind design. En J. Haugeland (ed.), *Mind design* (pp. 1-34). Cambridge: The MIT Press].
- Hawkins, J. & Blakeslee, S. (2004). *On intelligence*. New York: Times Books. [Seguimos la traducción de Carmen Martínez Gimeno: *Sobre la inteligencia*. Madrid: Espasa, 2005].
- Horkheimer, M. (1947). *Eclipse of reason*. New York: Oxford University Press. Seguimos la traducción de Jacobo Muñoz: *Crítica de la razón instrumental*. Madrid: Trotta, 2002.
- Horkheimer, M. & Adorno, Th. W. (1944). *Dialektik der Aufklärung*. New York: Social Studies Association [Seguimos la traducción de Juan José Sánchez: *Dialéctica de la Ilustración*. Madrid: Trotta, 2009].
- Jackson, P. (1986). *Introduction to expert systems*. New York: Addison-Wesley.
- Kaku, M. (2011). *Physics of the future: How science will shape human destiny and our daily lives by the year 2100*. New York: Random House [Seguimos la traducción de Mercedes García Garmilla: *La física del futuro*. Barcelona: Debate, 2011].
- Kandel, E. R., Schwartz, J. H. & Jessell, Th. M. (eds.) (1995). *Essentials of neural science and behavior*. Stamford: Prentice Hall.
- López-Muñoz, F. & Álamo, C. (2000). “El tratado del hombre: Interpretación cartesiana de la neurofisiología del dolor”, *Asclepio*, 52(1), 239-267.
- Marcuse, H. (1964). *One-dimensional man: Studies in the ideology of advanced industrial society*. Boston: Beacon Press [Seguimos la traducción de Antonio Elorza: *El hombre unidimensional*. Barcelona: Orbis, 1984].
- McCarthy, J. & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. En B. Meltzer & D. Michie (eds.), *Machine intelligence*, vol. 4. Edinburgh: Edinburgh University Press [Seguimos la edición digital gratuita descargada de [www-formal.stanford.edu/jmc](http://www-formal.stanford.edu/jmc)].
- Michie, D. (1974). *On machine intelligence*. Edinburgh: Edinburgh University Press.



- Parellada, R. (2000). “La naturaleza de las pasiones del alma en Descartes”, *Revista de Filosofía* (23), 235-242.
- Robinet, A. (1973). *Le défi cybernétique*. París: Editions Gallimard [Seguimos la traducción de Carmen García-Trevijano: *Mitología, filosofía y cibernética*. Madrid: Tecnos, 1982].
- Rodríguez, R. (1987). *Heidegger y la crisis de la época moderna*. Madrid: Síntesis.
- Shepherd, G. M. (1988). *Neurobiology*. New York: Oxford University Press.
- Smolensky, P. (1989). Connectionist modeling: Neural computation, mental connections. En J. Haugeland (ed.), *Mind design II* (233-250). Cambridge: The MIT Press, 1997].
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, Series 2, 42 (1936-1937) [Seguimos la reimpresión en J. Copeland (ed.), *The essential Turing* (pp. 58-90). Oxford: Clarendon Press, 2004].
- Turing, A. M. (1950). Computing machinery and intelligence, *Mind* (59) [Seguimos la reimpresión en J. Copeland (ed.). *The essential Turing* (pp. 441-464). Oxford: Clarendon Press, 2004] [Hay edición en español con introducción de Ricardo Álvarez y traducción de Alejandro Bazán: ¿Puede pensar una máquina? Buenos Aires: Almagesto, 1990].
- Wittgenstein, L. (1921). Logisch-philosophische Abhandlung. En W. Oswald (ed.), *Annalen der Naturphilosophie*, vol. 14, Leipzig [Seguimos la traducción de Jacobo Muñoz e Isidoro Reguera: *Tractatus logico-philosophicus*. Madrid: Alianza, 2000].
- Wittgenstein, L. (1953). *Philosophische Untersuchungen* [Seguimos la traducción de Alfonso García Suárez y Ulises Moulines: *Investigaciones filosóficas*. Barcelona: Crítica, 2012].

