

INTELIGENCIA ARTIFICIAL Y ESPECIFICIDAD HUMANA

Inteligência artificial e especificidade humana

Daniel Trapani¹

Resumo

Comprender el pensar humano bajo la forma de un mecanismo, y buscar en el pensamiento la nota distintiva de lo humano, han sido dos tópicos tradicionales de la filosofía. A mediados del siglo veinte la convivencia de ambas ideas se hace tensa, pues el computador digital es visto como mecanismo modélico del pensar humano. ¿Dónde buscar ahora la peculiaridad de lo humano?, ¿o habría que suprimir dicha peculiaridad? El presente trabajo recorre algunas respuestas a estas cuestiones (Daniel Dennett, Jean Baudrillard, John Searle) y propone un modo alternativo de sostener la peculiaridad humana sin buscarla allende la racionalidad.

Palabras clave: Especificidad Humana; Racionalidad; Intencionalidad; Inteligencia; Artificial; Pragmática Moral.

Resumen

Comprender o pensamento humano sob a forma de um mecanismo e buscar no pensamento a nota distintiva do humano têm sido dois tópicos tradicionais da filosofia. Meados do século XX, a convivência de ambas as idéias se faz tensa, pois o computador digital é visto como mecanismo modelo do pensar humano. Onde buscar agora a peculiaridade do humano? Ou haveria de suprir a dita peculiaridade? O presente trabalho recorre a algumas respostas a estas questões (Daniel Dennett, Jean Baudrillard, John Searle) e propõe um modo alternativo de sustentar a peculiaridade humana sem buscá-la além da racionalidade.

Palavras-chave: Especificidade humana; Racionalidade; Intencionalidade; inteligência artificial; Pragmática moral.

¹ Universidad Nacional de Rosario - Argentina
E-mail: danieltrapani@tower.com.ar

Introducción

Una mirada a la historia del pensar nos descubre dos intentos que recurrentemente se hacen presentes. El primero de ellos consiste en comparar el cerebro humano con diversos mecanismos, el segundo en hallar una propiedad característica de lo humano que permita distinguir al hombre del resto de los seres. Respecto al primer intento, comparar el cerebro con un mecanismo, ya algunos fisiólogos griegos pensaron que el cerebro humano funcionaba como una catapulta, y con posterioridad Leibniz lo comparó con un molino². Y ya no sólo el cerebro, sino que, con resonancias cartesianas, el mismo Leibniz asimiló el cuerpo orgánico de un ser vivo a *“una Especie de Máquina divina o Autómata natural.”*³ Y este paradigma mecánico irrumpe con todo su vigor en Julien de la Mettrie, para quien *“el hombre es una máquina tan compleja que resulta imposible formarse primero una idea clara al respecto y luego definirla en consecuencia.”*⁴ El cuerpo humano es la viva imagen del movimiento perpetuo, *“una máquina que pone en marcha sus propios mecanismos.”*⁵ El mecanicismo intentó mantener la diferencia entre lo viviente y el resto de los entes; así Leibniz hacía residir lo específico de un organismo vivo, paradójicamente, en su mayor perfección mecánica (en su ser más y mejor máquina) con relación a las máquinas artificialmente producidas. Las máquinas hechas por el hombre no son máquinas en cada una de sus partes, así, por ejemplo, el diente de una rueda de hierro tiene partes que no guardan relación con la función a la que se destina la rueda. *“Pero las Máquinas de la Naturaleza, es decir, los cuerpos vivos, son, sin embargo, Máquinas en sus menores partes hasta el infinito. Esto es lo que constituye la diferencia entre la Naturaleza y el Arte, es decir, entre el arte Divino y el Nuestro.”*⁶ En La Mettrie la diferencia entre el hombre y

² Cfr. LEIBNIZ, Gottfried, 1999: §17, p.26. Vale señalar que Leibniz utiliza la analogía para descartar la posibilidad de explicar la percepción por razones mecánicas. Pero, tras aclarar esto, el autor afirma en el §17 que si se imagina *“que existe una Máquina, cuya estructura haga pensar, sentir, tener percepción, se le podrá concebir agrandada, conservando las mismas proporciones, de tal manera que se pueda entrar en ella como si fuera un molino”*.

³ LEIBNIZ, Gottfried, 1999: §64, p.42.

⁴ LA METTRIE, Julien, 1962: 35.

⁵ LA METTRIE, Julien, 1962: 39.

⁶ LEIBNIZ, Gottfried, 1999: § 64, p. 42.

los otros vivientes es sólo de complejidad, ya que el hombre *"es al mono y a los más inteligentes animales lo que el péndulo planetario de Huyghens es a un reloj de Julian le Roi"*;⁷ se trata tan sólo de

Un reloj, aunque inmenso y construido con tanto artificio y habilidad que, si la rueda que sirve para marcar los segundos llega a detenerse, la de los minutos gira y sigue siempre su ritmo, así como la rueda de los cuartos y las otras continúan moviéndose cuando las primeras, herrumbradas o descompuestas por lo que fuere, han interrumpido su marcha. (La Mettrie, J., 1962: 92)⁸

Pero, volvamos al cerebro, también La Mettrie apela a un mecanismo para analogarlo, pero prefiere un instrumento musical en lugar del molino leibniciano, y a la hora de comprender el fenómeno del habla afirma que *"así como una cuerda de violín o una tecla de clavicordio vibra y produce un sonido, así las cuerdas del cerebro, heridas por los rayos sonoros, han sido impulsadas a devolver o a repetir las palabras que las tocaban"*.⁹ Más tarde es Freud el que compara al cerebro con un sistema hidráulico y electromagnético, y a principios del siglo veinte la imagen preferida será la de la central telefónica o el sistema telegráfico.

En lo que respecta al segundo intento, el de hallar una característica distintiva de lo humano, puede considerarse la insistencia medieval en la racionalidad como 'diferencia específica' un momento singular de una larga tradición que desde Heráclito hace del *lógos*, de la razón, del pensar, el carácter peculiar del hombre.

Ambos intentos, el de comparar el cerebro con una máquina, y el de hallar en el pensar racional la diferencia esencial del hombre, convi-

⁷ LA METTRIE, Julien, 1962: 91.

⁸ Es interesante advertir esta concepción de La Mettrie del cuerpo humano como un sistema *en paralelo*, no *en serie*, lo cual permite que el sistema (cuerpo humano) pueda funcionar aun cuando algunas partes del mecanismo dejen de hacerlo. Puede verse aquí un anticipo del conexionismo contemporáneo que permite, a diferencia de la Inteligencia Artificial tradicional, que una máquina conexionista pueda continuar funcionando razonablemente bien, aunque algunas de sus unidades estén funcionando inadecuadamente. Para este tema véase: Churchland, Paul (1988) *Matter and Consciousness*, Cambridge, MIT Press; y, también: Bechtel, William y Abrahamsen, Adele (1991) *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Oxford. Blacwell.

⁹ LA METTRIE, Julien, 1962: 53.

vieron, no sin tensiones, hasta mediados del siglo veinte, momento en que surge una nueva imagen como modelo del cerebro, imagen que suscitaría algunos problemas teóricos. Me estoy refiriendo al computador digital, mecanismo como los que la filosofía ya había utilizado, catapulta, molino, violín, clavicordio, sistema hidráulico, o central telefónica, pero con un matiz que producía cierta perplejidad, ya que instaba a “*examinar la siguiente pregunta. ¿Pueden pensar las máquinas?*”.¹⁰ Por primera vez se producía la comparación con una máquina cuya característica aparente era la de ‘pensar’, y esto creaba un conflicto entre los dos intentos mencionados, conflicto que ha de servirnos de guía para desarrollar la trama del presente escrito.

Si la entrada en escena del computador digital como máquina ‘de pensar’, homologable al cerebro humano, generaba una gran confianza en los logros futuros, constituía al mismo tiempo un desafío para toda una tradición que se había esforzado por sustentar el pensar racional como distintivo del ser humano. Se plantean, a partir de esto, dos cuestiones: a) ¿se ha de superar y suprimir la tan mentada diferencia constitutiva de lo humano?; y, en caso de respuesta negativa, b) si hemos de mantener la peculiaridad de lo humano, ¿ha de buscarse esto en una esfera no racional (el sentimiento, por ejemplo), o en un ámbito intra-racional al cual un computador digital, por definición, no pueda acceder? Las respuestas a estas dos preguntas han de organizar la exposición. En II se presenta la teoría de Daniel Dennett quien responde afirmativamente a la primera pregunta. Quienes responden negativamente a), darán lugar a los siguientes apartados. En III se alude a intentos por consolidar un ámbito inaccesible a las computadoras, pero en un terreno extra-racional. En IV, la propuesta de John Searle servirá como intento de mantener la peculiaridad de lo humano en la esfera racional. En V, por último, presentaremos un modo alternativo a los tradicionales (como el de Searle) que encuentra en la racionalidad práctico-moral una vía constitutiva de la especificidad humana.

¹⁰ TURING, A.M. (1950) *Máquinas computadoras e inteligencia*. (“Computing Machinery and Intelligence”, *Mind*, vol. LIX, núm. 236) En ANDERSON, Alan Ross. (1970) *Mentes y máquinas*. México. UNAM. P.11.

*Del termostato y la almeja a la búsqueda de nuestro diseño.
La confusión entre semejanza parcial e identidad.*

Daniel Dennett, al analizar la relación mente-cuerpo, tal como la presenta Descartes, nos presenta un trilema según el cual, o bien negamos que haya cuerpos y hechos físicos (idealismo), o bien negamos que haya mentes y hechos mentales no físicos (materialismo), o sostenemos ambos sin interacción entre ellos (paralelismo psicofísico). De estas tres posibilidades, la segunda parecía estar en mejores condiciones de suministrar a la mente una posición más satisfactoria con relación al corpus científico, fue así que algunos filósofos se inclinaron por identificar las entidades mentales con entidades físicas. Según Dennett esta teoría reduccionista *"de la identidad es errónea"*,¹¹ a pesar de declarar que su *"punto de partida es el mundo objetivo, materialista, tal como lo ve la tercera persona de las ciencias físicas"*,¹² y de proponerse *"ver cómo es la mente desde la perspectiva materialista, exterior, de la ciencia contemporánea"*.¹³ Pero rechazar esta teoría reduccionista no implica volver de la identidad a los antiguos dualismos, o caer prisionero del trilema, para ello hay que revisar un supuesto que suele considerarse evidente, el supuesto *"de que por un lado están las mentes y los hechos mentales, y por el otro, los cuerpos y los hechos físicos"*.¹⁴ La estrategia que permite revisar este supuesto se halla en la senda abierta por Ryle¹⁵ mediante un análisis categorial de los términos utilizados para entidades mentales y para entidades físicas.

Entre los sustantivos comunes hay algunos que pueden aparecer en los más variados contextos, que existen en un sentido *fuerte* (por ejemplo, mesa), y otros sustantivos que están limitados a pocos contextos. Ejemplos de estos últimos son: 'a-fuerza-de', 'por-bien', en estos casos se cometería un error categorial si un fisiólogo intentara identificar la 'a-fuerza-de' con ciertos esfuerzos musculares, *"lo que no significa que un a-fuerza-de sea un acompañamiento secreto, no físico de tales esfuerzos"*.¹⁶ Este tipo de sustantivos parecen ser más bien expresiones idio-

¹¹ DENNETT, Daniel, 1996: 24.

¹² DENNETT, Daniel, 1998: 18.

¹³ DENNETT, Daniel, 1998: 20.

¹⁴ DENNETT, Daniel, 1996: 18.

¹⁵ Cfr. RYLE, Gilbert. (1967). *El concepto de lo mental*. Buenos Aires. Paidós.

¹⁶ DENNETT, Daniel, 1996: 26.

máticas antes que sustantivos plenos. Pero, ¿qué ocurre con el vocabulario mental? Los pensamientos, dolores y deseos parecen de mayor robustez óptica que los ‘a-fuerza-de’. Dennett busca analogar entonces los sustantivos mentales con algún ejemplo de mayor complejidad, eligiendo para ello al término ‘voz’. Decimos “*está usted forzando su voz’ y ‘he perdido la voz’*. Ahora bien, la voz ¿es una cosa? Y si lo es, ¿qué clase de cosa es?”¹⁷ No puede identificarse la voz con nuestras cuerdas vocales, pues la voz puede transmitirse a distancia y sobrevivirnos grabada, y nada de esto ocurre con las cuerdas vocales. Tampoco coincide con las vibraciones producidas en el aire, pues éstas no pueden forzarse o reconocerse como sí puede serlo la voz. Y es que “*la palabra ‘voz’ como se descubre en su propio conjunto peculiar de contextos, no se adapta plenamente a la dicotomía físico-no físico (...) pero no por eso es una palabra vaga, ambigua o insatisfactoria*”.¹⁸ Dennett intenta defender un fisicalismo respecto de las voces sin identificar las voces con cosas físicas. Esto requiere de algunas distinciones ontológicas. Para ello intenta aclararse cómo funcionan los contextos de existencia (hay *x*) con diversos tipos de sustantivos (por ejemplo, ‘mesas’, ‘a-fuerzas-de’ y ‘voces’), analizando si son contextos legítimos para el sustantivo en cuestión. A través de este análisis y aprovechando las similitudes entre el sustantivo ‘voz’ y el vocabulario mental, el autor introduce la siguiente distinción, se propone llamar:

referenciales a los sustantivos o nominalizaciones que denoten, nombren o refieran a cosas existentes (en el sentido fuerte desarrollado antes), y *no referenciales* a los demás sustantivos o nominalizaciones, tales como ‘porbien’, ‘milla’ y ‘voz’. Las palabras y frases no referenciales son, pues, aquellas que dependen en alto grado de ciertos contextos restringidos, en particular, no pueden aparecer apropiadamente en contextos de identidad y, en forma concomitante, no tienen fuerza óptica o significación. (Dennett, D., 1996: 32-33)

Esta distinción promueve una alternativa promisoria, la de considerar las entidades mentales como *no referenciales*, consistiendo la tarea “*en suministrar una explicación científica de las diferencias y las*

¹⁷ DENNETT, Daniel, 1996: 27.

¹⁸ DENNETT, Daniel, 1996: 28. Ni tampoco se hace necesario generar un dualismo garganta-voz que sea una nueva saga del dualismo cartesiano mente-cuerpo.

similitudes en las condiciones en virtud de las cuales diferentes enunciados del lenguaje mental son verdaderos o falsos".¹⁹ Así, la tarea no reside en identificar una creencia con algún estado físico del cerebro, sino en señalar las condiciones bajo las cuales es posible dilucidar la verdad o falsedad del enunciado '*X está creyendo y*'. De este modo, es posible caracterizar la tarea.

como la búsqueda de una explicación de la mente que *esté unificada con*, que sea *consistente con* y ciertamente *una parte de* la ciencia como un todo, pero que elude –al menos inicialmente– la obligación de encontrar entre los *objetos* de la ciencia algún referente para los términos del vocabulario mental. (Dennett, D., 1996:37)

Pero a esta posible estrategia se le opone un obstáculo de peso, aquel que estima que la indagación de la verdad o falsedad de ciertos enunciados del lenguaje mental están fuera del dominio de las ciencias físicas, este obstáculo se presenta como: el problema de la Intencionalidad,²⁰ y fue defendido por Franz Brentano.²¹

Según Brentano, era posible establecer una distinción neta entre los fenómenos mentales y los fenómenos físicos, pues los primeros exhiben un rasgo peculiar: la Intencionalidad que abre un abismo infranqueable entre lo mental y lo físico. La tesis de Brentano rezaría así: *Todo fenómeno mental se caracteriza por la inexistencia Intencional de un objeto, y por la referencia a un contenido, y una dirección hacia un objeto*. En esta caracterización de los fenómenos mentales pueden apreciarse dos aspectos: - la dirección hacia un objeto, y, - la relación con un contenido o significado. En cuanto al primer aspecto, la *dirección hacia un objeto*, no se puede querer sin querer *algo*, esperar sin esperar *algo*, y sin embargo el *objeto* no tiene por qué existir como los objetos de acciones físicas, como tocar o levantar. Puedo esperar que surja una cura para el Alzheimer, aunque este objeto no exista, al menos de modo

¹⁹ DENNETT, Daniel, 1996: 37.

²⁰ Respeto aquí el uso de Intencionalidad con mayúscula tal como lo hace Dennett (1996), a fin de no confundirlo con la noción más habitual de intención o propósito ligado a la acción.

²¹ Cfr. BRENTANO, Franz. *Psychologie vom Empirischen Standpunkt*, Leipzig, 1874, vol.I, Libro II, cap. i, "The Distinction between Mental and Physical Phenomena", selección en R. Chisholm, *Realism and the Background of Phenomenology*, Glencoe, 1980.

ordinario, y esto es posible, según Brentano, porque el objeto (la cura del Alzheimer) *inexiste*. En referencia al segundo aspecto, la *relación con un contenido*, en lugar de haber un objeto al cual nos dirigimos, aparece un enunciado relacionado con el objeto, en este sentido *creemos que*, o *esperamos que* (enunciados de actitud proposicional), resultan distintos a *creemos x*, o *esperamos x*. Dennett considera que la tesis de Brentano es pasible de muchos contraejemplos por lo que decide conferir a la Intencionalidad una asociación más amplia con lo *psicológica*, en lugar de la restricta asociación brentaniana de lo Intencional con lo mental. Decide, además, hacer residir lo específico de la Intencionalidad en el segundo rasgo señalado (la referencia de las cláusulas proposicionales) y no en el primer rasgo (el estar dirigida a objetos); intentando transformar enunciados sobre objetos en enunciados de actitud proposicional, así por ejemplo, '*Creo en los duendes*' se transformaría en '*Creo que hay duendes*'.

La tesis Intencionalista ha sido fuertemente criticada por el conductismo, basándose en que si no puede lograrse que los giros Intencionales encajen en el marco de la ciencia vigente, entonces deben abandonarse. Sin embargo, el conductismo no ha logrado producir una austera ciencia del estímulo y la respuesta que prescindiera de lo Intencional. Según Dennett

la dificultad con la que ha tropezado el conductista es básicamente ésta: mientras que resulta claro que un experimentador puede predecir, por ejemplo, la velocidad de aprendizaje a partir de las condiciones iniciales de sus laberintos y de la historia empírica de sus animales, ¿cómo puede especificar exactamente *qué* aprenden? (Dennett, D., 1996: 56)

Pareciera que lo que aprende, por ejemplo una rata de laboratorio, es 'dónde está la comida', pero ¿cómo se puede caracterizar esto desde un enfoque no Intencional? El fracaso del intento conductista.

Constituye una fuerte indicación, aunque no una prueba, por supuesto, de que los fenómenos psicológicos *deben* caracterizarse en forma Intencional si es que han de ser explicados y predichos, y de que ninguna ciencia de la conducta puede progresar sin los giros idiomáticos Intencionales. (Dennett, D., 1996: 57)

Las explicaciones Intencionales se oponen al conductismo, pues mientras éste trata de hallar regularidades que expliquen la adecuación de las conexiones entre acontecimientos antecedentes y consecuentes a la conducta a explicar, para la explicación Intencional, en cambio, el hecho de que un acontecimiento (Intencional) sea seguido adecuadamente por otro, no está sujeto a explicación. *“La intención de alzar el brazo no sería la intención de alzar el brazo si no estuviese seguido, salvo interferencia, por el alzamiento del brazo, de modo que la cuestión de por qué una sigue a la otra es superflua”*.²² Allí donde la explicación en un sistema Intencional da por terminada su tarea, el resto de la ciencia (física) ve sólo un punto de partida requerido para comenzar la explicación.

Todo lo anterior nos coloca ante un nuevo dilema: por un lado, los indicios de fracaso de una psicología conductista nos llevan a rechazarla; pero, por otro, *“la alternativa de una psicología enteramente Intencional implicaría una reestructuración catastrófica de la ciencia en general”*.²³ Pero, para Dennett, es posible salir del dilema. El punto débil en el que se asienta el dilema es la tesis de la *irreductibilidad extensional*²⁴ de los enunciados Intencionales. Pero para evitar este dilema Dennett propone una especie de escalera de Jacob que nos permita pasar de la tierra del movimiento y la materia al cielo del contenido y el propósito, en un ida y vuelta que establezcan formulaciones de este tipo *“el*

²² DENNETT, Daniel, 1996: 61.

²³ DENNETT, Daniel, 1996: 62.

²⁴ Los enunciados Intencionales son caracterizados como enunciados intensionales, no extensionales. La extensión de un término es la clase de las cosas a las que se aplica el término. La intensión de un término es el conjunto de características relevantes que promueven la aplicación de ese término a un conjunto de cosas. La lógica que funciona en el discurso científico es extensional, ya que es ciega a las distinciones intencionales (de significado), razón por la cual *“la intersubstitución de términos coextensionales, cualesquiera fuesen sus intensiones, no afecta el valor de verdad (verdad o falsedad) del enunciado que los incluye. Además, el valor de verdad de un enunciado complejo es siempre función de los valores de verdad (verdad o falsedad) del enunciado que los incluye”* DENNETT, Daniel 1996: 52. Consideremos ahora el siguiente ejemplo: ‘Pelé’ y ‘Edson Arantes do Nascimento’ nombran al mismo individuo, del enunciado ‘Pelé jugó en el Santos’ se deduce el enunciado ‘Edson Arantes do Nascimento jugó en el Santos’; pero del enunciado ‘Santiago cree que Pelé jugó en el Santos’, no se deduce que el enunciado ‘Santiago cree que Edson Arantes do Nascimento jugó en el Santos’ sea verdadero, pues Santiago puede ignorar que Pelé y Edson Arantes do Nascimento sean una y la misma persona.

estado físico *S* tiene la significación (o significa, o tiene el contenido) *p*".²⁵ Es a partir del uso de esa escalera de Jacob como Dennett comienza a borrar la distinción entre el ser humano y otros entes, al englobar a ambos bajo la categoría de '*sistema intencional*'. Al presentar su enfoque acerca de qué ha de entender por '*sistema intencional*', Dennett supone que habrá resistencias a aceptarlo, pues la idea de que ningún autómatas podrá ser consciente, como nosotros, está muy popularizada. Idénticos reparos suscitó, en su momento, la física académica, en aquellos puntos en que se oponía a la física popular.²⁶ Si la física académica es fuertemente contraintuitiva, una psicología que genere rechazos tiene, por ello mismo, fuertes indicios a su favor, ya que "*cualquier teoría que progrese está destinada a ser inicialmente contraintuitiva*".²⁷

Tanto la física popular como la psicología popular funcionan en los aspectos más importantes de nuestra vida. "*Gracias a la física popular podemos mantenernos abrigados y bien alimentados y evitar los choques, y gracias a la psicología popular colaboramos en proyectos multipersonales, aprendemos los unos de los otros y disfrutamos de períodos de paz local*".²⁸ Ambas, la física popular y la psicología popular, funcionan en lo cotidiano, pero con esta diferencia, mientras la física popular ha quedado desacreditada en algunas de sus intuiciones en situaciones claramente estipuladas; la psicología popular, en cambio, no cuenta con muchas situaciones en las que quede resueltamente desacreditada. Dennett se propone explicar el éxito habitual de la psicología popular sosteniendo la siguiente tesis: "*nos comprendemos los unos a los otros adoptando la actitud intencional*",²⁹ y es esto lo que nos permite adquirir una capaci-

²⁵ DENNETT, Daniel, 1996: 63.

²⁶ "*La física popular es el sistema de expectativas sensatas que todos tenemos acerca de cómo los objetos físicos de tamaño mediano de nuestro mundo reaccionan ante los acontecimientos de mediana importancia. Si vuelco un vaso con agua en la mesa de la cena, usted salta de su silla porque espera que el agua se derrame por el costado y le empape la ropa*". DENNETT, Daniel, 1998: 20.

²⁷ DENNETT, Daniel, 1998: 19. Esta destinación a la que alude Dennett parece algo forzada. Está generalizando inapropiadamente, si bien muchas teorías físicas académicas son fuertemente contraintuitivas, no parece ser ésta una característica indispensable a la hora de formular hipótesis. Más que la invención popperiana sería recomendable recordar aquí la insistencia peirceana en la abducción a la hora de generar hipótesis, proceso inferencial que se nutre de la cantera de analogías con que la educación haya enriquecido al científico.

²⁸ DENNETT, Daniel, 1998: 23.

²⁹ DENNETT, Daniel, 1998: 24.

dad general para interpretar a nuestros semejantes y a algunos desemejantes. Pero, ¿esta 'actitud intencional' en qué consiste?; se trata de asumir una cierta estrategia consistente "en tratar al objeto cuyo comportamiento se quiere predecir como un **agente racional con creencias y deseos y otras etapas mentales** que exhiben lo que Brentano y otros llaman intencionalidad".³⁰ Es a partir de este tratar como un agente racional a cualquier objeto o sistema cuyo comportamiento resulte bien pronosticado por esta estrategia, como comienza Dennett a diluir la especificidad de lo humano, ya que de cualquier objeto o sistema con el que resulte exitosa la estrategia puede afirmarse que "es, en el más **completo** sentido de la palabra, un **creyente**".³¹ Ser un verdadero *creyente* es ser un sistema intencional, un sistema cuyo comportamiento se puede predecir confiablemente por medio de la estrategia intencional. La actitud intencional funciona de este modo:

Primero se decide tratar al objeto cuyo funcionamiento hay que predecir como un agente racional; luego se deduce qué creencias debería tener ese agente, dada su posición en el mundo y su objetivo. Más tarde se deduce qué deseos tendría que tener siguiendo las mismas consideraciones, y por fin se predice que este agente racional actuará para conseguir sus metas a la luz de sus creencias. (Dennett, D., 1998: 28-29)

Pero, ¿por qué afirmamos que esta estrategia diluye la peculiaridad de lo humano? ¿No es acaso éste el modo en que interpretamos la conducta de otras personas? Quizás sea así, pero no es esto lo que pretendemos discutir aquí, sino la indiferenciada y abusiva extensión de la estrategia intencional, y la desmesurada consideración como '*creyente*' de cualquier objeto con el que la estrategia funcione exitosamente, ya que "este proceso puede «encarnarse», «ejemplificarse», o realizarse de diversas maneras, no necesariamente en un tejido nervioso viviente. No sólo serían aptos los cerebros, sino también los ordenadores y hasta los fantasmas."³²

³⁰ DENNETT, Daniel, 1998: 27. Énfasis mío.

³¹ *Ibidem*. Énfasis mío.

³² BUNGE, Mario, 2003: 230.

La actitud intencional consiste en una doble atribución: a) atribución de aquellas *creencias* que el sistema *tendría que tener*, y, b) atribución de aquellos *deseos* que el sistema *tendría que tener* (los que el sistema *considera* buenos para sí). La cuestión es la siguiente, ¿es la adopción de una actitud intencional ante un objeto lo que constituye a un sistema en intencional, esto es, en un verdadero creyente?, o, ¿es el reconocer que nos hallamos ante un sistema intencional el que nos lleva a aplicarle la actitud intencional? A juzgar por algunos pasajes de sus escritos, pareciera que Dennett se inclina por la primera alternativa: hemos de considerar *sistema intencional* a aquellos entes ante los que adoptemos la estrategia y constatemos que la misma funciona. La gama de entes ante los que se puede adoptar esa estrategia es muy amplia, y esto implica que la clase de los verdaderos creyentes y de los verdaderos deseantes abarca un extenso registro en la frondosa imaginación de Dennett, ya que la estrategia intencional

Funciona también en la mayoría de los mamíferos casi siempre. Por ejemplo, se puede usar para diseñar mejores trampas para cazar esos mamíferos, razonando sobre lo que la criatura sabe o cree acerca de distintas cosas, qué prefiere, qué quiere evitar. La estrategia funciona con los pájaros, con los peces, con los reptiles y con los insectos y arañas y hasta con criaturas tan inferiores y poco emprendedoras como las almejas. (Cuando una almeja cree que hay algún peligro cerca, no afloja su apretón sobre su concha cerrada hasta que se convence de que ha pasado el peligro). También funciona con algunos *artefactos*: el *ordenador* que juega al ajedrez no se comerá mi caballo porque *sabe* que hay una línea de juego siguiente que le llevaría a perder su torre, y no *quiere* que eso suceda. Más modestamente el *termostato* apagará la caldera en cuanto llegue a *creer* que la habitación ha alcanzado la temperatura deseada.

La estrategia funciona hasta con las plantas. En un lugar con tormentas tardías de primavera, habría que plantar variedades de manzana que son especialmente *cautelosas* en lo que se refiere a *llegar a la conclusión* de que es primavera, que es cuando *quieren* florecer, por supuesto. También funciona con fenómenos tan inanimados y *aparentemente no intencionales* como el rayo. (Dennett, D., 1998: 32-33, énfasis mío)³³

³³ Es interesante notar que el autor sólo coloca cursivas allí donde se refiere a las plantas como *cautelosas* o *llegando a una conclusión*, pero no parece abandonar la literalidad cuando alude al 'saber', 'creer' o 'querer' de artefactos como un computador o un termostato.

Ante esta heterogénea muestra de atribuciones de creencias, unas 'serias', otras dudosas, y otras, aparentemente, metafóricas, cabría preguntarse si es posible deslindar un caso de otro, si cabe distinguir dentro de los sistemas intencionales aquellos que *realmente* tienen creencias y deseos. "*Pero ese sería un trabajo de Sísifo*"³³ condenado al fracaso *ab initio*, lo cual no parece preocupar al autor, pues lo importante es "*que hasta en el peor de estos casos, aun cuando estuviéramos seguros que la estrategia funciona por razones equivocadas es sin embargo cierto que sí funciona*".³⁵

Ahora bien, esta respuesta de Dennett parece dejar expedito el camino para atribuir intencionalidad a cualquier objeto. El autor considera esta objeción y se pregunta si su "*definición de sistema intencional excluye algún objeto*",³⁶ ¿qué es lo que descalificaría, por ejemplo, a un 'atril' como sistema intencional? Su respuesta es que con el 'atril' la estrategia no funciona, pues "*no obtenemos de ella ningún poder predictivo que ya no tuviéramos anteriormente*".³⁷ El aspecto analizado por Dennett en cuanto a la conducta del 'atril' está referido a su movimiento local, el tratar al atril como un sistema intencional que *crea* estar sobre un escritorio y que *desea* permanecer allí, no añade nada a lo que ya sabíamos que iba a hacer – esto es, *nada*. Pero, del mismo modo, ¿no sabíamos ya que un rayo puede ser desviado, sin necesidad de atribuirle creencias ni deseos, ni considerar que al desviarlo lo estamos engañando? Además, si con el 'atril' no funciona la 'estrategia intencional', ¿qué decir del agua del mar en su pleamar y en su bajamar, o de una nube bombardeada eléctricamente por nosotros?³⁸ Nos parece que en Dennett se unen el espíritu arcaico antropomorfizante con un hálito de moderno

³⁴ DENNETT, Daniel, 1998: 33.

³⁵ *Ibidem*.

³⁶ *Ibidem*.

³⁷ *Ibidem*.

³⁸ Este gesto teórico antropologizante de Dennett es común, según Bunge, a la psicología informacionista, y el error reside en confundir semejanza parcial con identidad. Así BUNGE, Mario, 2003: 194-195. "*Ningún astrónomo afirmarí que los planetas calculan sus órbitas al moverse alrededor del Sol. Y, con todo, los computacionistas nos piden que creamos algo parecido: que las sustancias químicas calculan al reaccionar unas con otras y que los cerebros computan al percibir, sentir o pensar (a consecuencia de lo cual y de manera similar a ciertos robots programados se les debe atribuir conciencia, sentimientos y, tal vez, hasta una conciencia moral). El computacionista confunde la semejanza parcial con la identidad.*"

pragmatismo. Si una antropomorfización sirve, entonces no hay que plantearse nada más. Así, por ejemplo, algunos de entre nosotros, los ingenieros, podríamos prescindir de la estrategia intencional y no precisar esta antropomorfización para entender cómo funciona un termostato, o los meteorólogos para predecir lo que hará una nube. Sin esta limitante epistémica la estrategia intencional resultaría superflua, sólo justificada por la ausencia de un mejor recurso; en este sentido *"la decisión de adoptar la actitud intencional es libre, pero los hechos acerca del éxito o del fracaso de la actitud, si se la adoptara, son perfectamente objetivos"*.³⁹ Robert Nozick⁴⁰ ha visto este déficit de la teoría dennettiana, planteando la posibilidad de seres de inteligencia superior que nos trataran como los ingenieros a un termostato, no precisando de la actitud intencional para predecir nuestra conducta; con lo cual, nuestro carácter de creyentes (sistemas intencionales) sería una mera concesión de un espectador poco informado.⁴¹

Ante las acusaciones del carácter puramente opcional que tendría en nuestras relaciones la actitud intencional, Dennett se propone defender lo *inevitable* de la actitud intencional a la hora de tratar con uno mismo y con nuestros congéneres inteligentes. Para ello deja de destacar *"nuestra afinidad con las almejas y los termostatos para enfatizar un punto de vista del status lógico de la atribución de creencias"*,⁴² enfoque que le permitirá reconocer las diferencias evidentes, y acallar así el reclamo siguiente: *ser un verdadero creyente no es más que ser un sistema predecible mediante la estrategia intencional, esto es, no hay más que creer que algo es un sistema munido de creencias*. Pero este nuevo enfoque es más proclamado que efectivizado por el autor, pues las creencias y deseos son tan sólo consecuencia de la misma estrategia intencional.

No se trata de que atribuyamos (o debemos atribuir) creencias y deseos sólo a las cosas en las que no encontramos representaciones internas, sino que cuando descubrimos algún objeto para el cual la estrate-

³⁹ DENNETT, Daniel, 1998: 34.

⁴⁰ Cfr. NOZICK, Robert (1981) *Philosophical Explanations*. Cambridge, MA, Harvard University Press.

⁴¹ Dennett intenta responder a la objeción planteada, aunque de modo poco afortunado. Cfr. DENNETT, Daniel, 1998: 35-49.

⁴² DENNETT, Daniel, 1998: 38-39.

gia intencional funciona, nos esforzamos por interpretar algunos de sus estados o procesos internos como representaciones internas. Lo que hace que algún rasgo interno de algo sea una representación sólo podría ser su papel en la regulación de un sistema intencional. (Dennett, D., 1998: 41)

Y es que

los principios y problemas de interpretación que descubrimos cuando atribuimos creencias a la gente son los *mismos* principios y problemas que descubrimos cuando miramos el problema ridículo, pero gloriosamente sencillo, de atribuirle creencias a un termostato.⁴³

La diferencia (entre un termostato y una persona) en cuanto a estrategia intencional, aparece cuando nos preguntamos el 'por qué' funciona la estrategia. En el caso del termostato la razón se encuentra en que "*el termostato está bien diseñado*",⁴⁴ y este buen diseño consiste en que fue proyectado para ser comprendido fácilmente desde esta actitud. El por qué funciona un termostato es, en realidad, una respuesta acerca de *cómo funciona el mecanismo*. Mas en el caso de que el sistema intencional sea una persona, sólo sabemos que la estrategia funciona, pero no sabemos "*cómo funciona el mecanismo con el que la naturaleza nos ha dotado*",⁴⁵ ¿será acaso porque no conocemos el diseño del mecanismo?, y ¿será esto subsanable? La respuesta de Dennett es afirmativa para ambas preguntas. En cuanto a la primera, si bien piensa al hombre en términos mecanicistas, no introduce la hipótesis leibniziana de la distinción entre el arte divino y el nuestro.⁴⁶ En cuanto a remediar el desconocimiento del diseño, Dennett estima que alguna versión de la ciencia cognitiva dará con la respuesta, quizás el que podamos diseñar sistemas más complejos que un termostato, nos pondrá en situación de comprender nuestro propio diseño, borrando el hiato que Leibniz defendía entre un Autómata natural y un Autómata artificial. La síntesis de esta pretensión, para ascender en la escala de complejidad, puede plantearse así:

⁴³ DENNETT, Daniel, 1998: 41.

⁴⁴ DENNETT, Daniel, 1998: 42.

⁴⁵ *Ibidem.*

⁴⁶ Cfr. LEIBNIZ, Gottfried, 1999: §64, p.42.

Para cada creencia predictivamente atribuible habrá un estado interno funcionalmente notable del mecanismo, que se puede descomponer en partes funcionales de casi la misma forma en que la oración que expresa la creencia se puede descomponer en parte, es decir, en palabras o términos. Las inferencias que les atribuimos a los seres racionales serán reflejadas por procesos causales físicos en el hardware; la forma *lógica* de las proposiciones en que se cree, se copiarán en la forma *estructural* de los estados en correspondencia con ellos. Esta es la hipótesis de que hay un *lenguaje del pensamiento* codificado en nuestros cerebros, y que eventualmente nuestros cerebros serán entendidos como sistemas manipuladores de símbolos en una *analogía* por lo menos aproximada con los ordenadores. (Dennett, D., 1998: 43)

Irónicamente, y a pesar del declamado materialismo de autores como Dennett o Churchland, *“se trata de dualistas y hasta de platónicos, puesto que escriben acerca de cerebros que almacenan y procesan símbolos y programas informáticos que están «ejemplificados» (encarnados) en cerebros o en robots”*⁴⁷, sin necesidad de un viviente, erigiendo una barrera entre el cerebro y su mente, *“lo que es como estudiar el caminar sin prestar atención a las piernas”*.⁴⁸

Del vedado placer

Refiramos ahora la postura que coloca la especificidad de lo humano en lo ‘no racional’. La alternativa pareciera rezar así: *‘Si homologamos nuestro cerebro a un computador digital (máquina de pensar) tendremos que buscar en una esfera no racional (sentimiento, emoción) la diferencia específica del hombre’*. Ante las afirmaciones de Turing en un artículo de 1950, según las cuales era posible hablar sin inconvenientes del ‘pensar’ de las máquinas como similar al nuestro, se levantaron voces discrepantes, negándose a tal asimilación, y recurriendo como antídoto exorcizante contra Turing a la Oración de Lister de 1949 del Profesor Jefferson:

⁴⁷ BUNGE, Mario, 2003: 193-194.

⁴⁸ BUNGE, Mario, 2003: 229.

Hasta que una máquina no sepa escribir un soneto o componer un concierto a base de los pensamientos y las *emociones que siente* (...) no podremos estar de acuerdo en que la máquina pueda ser igual que un cerebro (...) Ningún mecanismo podría *sentir alegría* de sus éxitos, *tristeza* cuando sus válvulas se fundieran, *placer* al ser adulado y *sentirse desgraciado* a consecuencia de sus errores, *encantado* por el sexo, *enfadado* o *deprimido* al no lograr lo que desea. (Turing, A.M., 1950: 26-27, énfasis mío)⁴⁹

Quizás pudiera haber una máquina de pensar, pero nunca una máquina de sentir. Es el recurso a la esfera sensitivo-afectiva la que define de la especificidad humana, y esta apelación se hace explícita en un artículo de Michael Scriven publicado en *Mind* en 1953, el autor se plantea si *“¿hay alguna diferencia esencial entre los robots y los seres humanos?”*⁵⁰ y encuentra en la conciencia senso-emotiva la distinción, pues los robots *“hacen lo que hacen los seres humanos, pero no pueden ser lo que son los seres humanos”*.⁵¹ Es sólo en el ‘hacer’ en el que se nos asemejan, no en nuestro ser, estando nuestro ‘ser’ en la capacidad de ‘sentir placer’ o ‘autoconmiseración’ más que en nuestra capacidad de habla:

Podemos hacer que las máquinas aren y recojan la cosecha, actúen e imiten, pero no que *sientan placer* o *autoconmiseración*. *Estas cosas son posibles sólo para los seres conscientes*, y por más ingenioso que sea el mecanismo, por más complejo que sea el comportamiento de una máquina, aunque hable o juegue al ajedrez, no es más consciente que un reloj. (Scriven, M., 1953: 51, énfasis mío)⁵²

Años después, Paul Ziff es quien de modo más claro define de esta tesis de la especificidad humana asentada en lo ‘sentimental’. La cuestión que se plantea Ziff es si un robot podría tener sentimientos, esto es, si podría sentirse cansado o aburrido, por ejemplo. Y sólo si esto fuera posible, habríamos suprimido la diferencia entre la máquina y el

⁴⁹ Turing está citando a JEFFERSON, G. (1949) “The Mind of Mechanical Man”, *British Medical Journal*, vol.I, pp.1105-1121.

⁵⁰ SCRIVEN, Michael. *El concepto mecánico de la mente*. (“The Mechanical Concept of Mind”, *Mind*, vol.LXII, núm. 246, 1953). En ANDERSON, Alan Ross, 1970: 49.

⁵¹ SCRIVEN, Michael, 1953: 50.

⁵² Cabe aclarar que en un Apéndice de 1963, Scriven modifica su posición inicial, concluyendo *“que es posible construir una supercomputadora de tal modo que sea absolutamente irrazonable negar el hecho de que tenga sentimientos”*. P.63.

hombre: "el problema no es si podemos borrar los límites entre un hombre y una máquina y atribuir así sentimientos a esta última, sino si podemos **atribuir sentimientos** a la máquina y borrar así la **diferencia** entre ésta y un hombre."⁵³ El tema no es, para Ziff, la prohibición de borrar los 'límites' entre hombre y máquina, sino la posibilidad o no de 'atribuir sentimientos', de la cual el emerger de un límite es una consecuencia. La respuesta negativa del antecedente (¿podemos atribuir sentimientos a las máquinas?) está signada por estas dos premisas: '*Solamente los seres vivos pueden tener sentimientos*', y '*los robots no pueden ser seres vivos*'. El plantearse el cansancio o el aburrimiento de un robot no difiere demasiado del preguntarse si "*¿Podría una piedra sentirse cansada? ¿Podría el número 17 sentirse cansado?*"⁵⁴

Si bien Scriven admitía una posible identificación en la acción, insistiendo en la diferencia óptica, Ziff limita aún más la asimilación, incluyendo además el nivel de las acciones:

Ningún robot podrá hacer jamás todo lo que un hombre es capaz de hacer, sin que importe cómo esté construido o lo complejos y variados que sean sus movimientos y operaciones (...) Tal vez puedan matar, pero no literalmente asesinar. Podrán pedir disculpas, pero literalmente no lo harán. Hay acciones que pueden ser efectuadas sólo por personas. *Ex hypothesi* los robots no son personas. (Ziff, P., 1959: 144)

Y esto impide que sensatamente podamos afirmar que los robots sienten algo, pues "*no hay verdades psicológicas acerca de robots sino solamente acerca de sus constructores humanos*".⁵⁵

Los logros ulteriores de la tecnología robótica parecieron agudizar la imposibilidad de distinguir entre hombre y máquina. La antropología, según Baudrillard, ha de llamarse a silencio ante el nuevo dilema: "*¿Soy un hombre, soy una máquina? Ya no hay respuesta a esta pregunta*

⁵³ ZIFF, Paul. (1959) *Los sentimientos de los robots*. ("The feelings of robots", Analysis, vol.XIX, número 3. En ANDERSON, Alan Ross (1970). P.143. Énfasis mío. En este escrito Ziff se opone a la afirmación de MacKay, M. en el sentido de que los robots eran capaces de tener sentimientos. Cfr. MACKAY, M. (1956) "The Epistemological Problem for Automata", en *Automata Studies*, Princeton, Princeton University Press, p.235.

⁵⁴ ZIFF, Paul. (1959). En ANDERSON, Alan Ross, 1970: 145.

⁵⁵ ZIFF, Paul. (1959) En ANDERSON, Alan Ross, 1970: 147.

antropológica. Así pues significa en cierta manera el final de la antropología, confiscada subrepticamente por las máquinas y las tecnologías más recientes.⁵⁶ La indistinción hombre-máquina toma la forma de un ambiguo circuito integrado:

En la relación con las máquinas tradicionales no existe ambigüedad. El trabajador es, en cierto sentido, extraño a la máquina, y por tanto alienado por ella. Mantiene su cualidad preciosa de hombre alienado. Pero las nuevas tecnologías, las nuevas máquinas, las pantallas interactivas no me alienan en absoluto. Forman conmigo un circuito integrado, (Baudrillard, J., 1991: 65),

ni siquiera existe ya *“la alienación del hombre por el hombre, sino una homeostasis del hombre por la máquina.”*⁵⁷ Para Baudrillard, las máquinas pueden confiscar el pensamiento humano, objetivarlo y transformar al hombre en un espectador *“pues lo que ofrecen esas máquinas es el espectáculo de pensamiento, y los hombres, al manipularlas, se entregan al espectáculo del pensamiento más que al mismo pensamiento”*.⁵⁸ Puede anticiparse un hombre futuro sin pensamiento: *“De la misma manera que se puede vaticinar que las gafas o las lentes de contacto se volverán un día la prótesis integrada de una especie de la que habrá desaparecido la mirada, también es de temer que la inteligencia artificial y sus soportes técnicos se vuelvan la prótesis de una especie de la que habrá desaparecido el pensamiento.”*⁵⁹ Pero aun así, el autor cree encontrar una carencia hedonística que sostiene la diferencia hombre-máquina, la Inteligencia artificial es siempre una *machine célibataire*. Confiscado el confiscable pensamiento surge inconfiscable el inconfiscado placer, y esto por siempre:

Lo que *siempre diferenciará* el funcionamiento del hombre del de las máquinas, incluso las más inteligentes, es la *ebriedad del funcionar, el placer*. Inventar máquinas que sientan placer es algo que, afortunadamente, sigue estando fuera de los poderes del hombre. Todo tipo de prótesis puede ayudar a su placer, pero es incapaz de inventar alguna que disfrute en su lugar. Inventa aquellas que trabajan, ‘piensan’ o se

⁵⁶ BAUDRILLARD, Jean, 1991: 65.

⁵⁷ BAUDRILLARD, Jean, 1991: 66.

⁵⁸ BAUDRILLARD, Jean, 1991: 58.

⁵⁹ BAUDRILLARD, Jean, 1991: 59.

desplazan mejor que él, o en su lugar, pero no hay prótesis, técnica o mediática, del *placer del hombre*, del *placer de ser hombre*. (Baudrillard, J., 1991: 60, énfasis mío)

Las máquinas se han especializado en *operar*, y lo que caracteriza a la operación, a diferencia de la acción, es que está necesariamente regulada en su desarrollo, las máquinas pueden operar como un humano, mas no pueden actuar como nosotros. En el hombre, placer o sufrimiento sintetizan un irónico incremento de funcionamiento que lo aleja de la mera operación reglada,⁶⁰ y permite hallar en el *vedado placer* a las máquinas, la instancia no-racional de alteridad humana.

Advertimos aquí cómo los sentimientos se han erigido en la nueva 'diferencia específica', y en romántico reducto de la condición humana. Ahora bien, cuáles son los supuestos de quienes como Jefferson, Scriven, Ziff o el mismo Baudrillard, han optado por recluirse en la no racionalidad del sentimiento como sello antropológico distintivo. Estos autores parecen cautivos del éxito abrumador de la lógica simbólica de principios del siglo veinte. El modo en que Frege, Russell y Wittgenstein diseñaron la lógica como modelo del pensar con sentido, constituye la base de la cibernética y de la inteligencia artificial. La primera filosofía analítica y el neopositivismo elevaron la forma lógica a paradigma del pensamiento, paradigma que obviamente no abarcaba al sentimiento; y veneraron a la lógica simbólica como *e/* mecanismo de la racionalidad. Este criterio logicista, corroborado por sus éxitos, condicionó la decisión de quienes optaron por el sentimiento como reducto antropológico al que un pensar lógico (compartido por hombres y máquinas) no pudiera llegar.

La mente como programa: ¿ un cuento chino?

Ya el pragmatismo americano a principios del siglo veinte pretendía subsanar un hiato que resultaba preocupante, el que se abría entre la visión científica del mundo heredada de la modernidad y aquellos valores que loamos, entre los cuales figura nuestra propia autocomprensión. Es precisamente este hiato entre la concepción científica y

⁶⁰ Cfr. BAUDRILLARD, Jean, 1991: 60.

nuestra autoconcepción el punto de inicio mediante el cual John Searle aborda la cuestión de la especificidad humana:

Tenemos una cierta representación de sentido común de nosotros mismos como seres humanos que es muy difícil casar con nuestra concepción 'científica' global del mundo físico. Nos pensamos a nosotros mismos como agentes *conscientes, libres, cuidadosos, racionales* en un mundo del que la ciencia nos dice que consta enteramente de partículas físicas carentes de mente y de significado. (Searle, J., 1985: 17)

Durante el siglo veinte hubo intentos de salvar este hiato entre la visión científica (neurofisiología) y nuestra autoconcepción de sentido común (psicología intencionalista), buscando hallar "*algo que fuera una ciencia de la conducta humana que no fuera psicología de sentido común ni tampoco fuera neurofisiología*".⁶¹ Entre estos esfuerzos, fallidos según Searle, pueden citarse al conductismo, la cibernética, el estructuralismo, el postestructuralismo, o más recientemente la sociobiología. Pero los principales candidatos de hoy, los que parecen contar con mejores credenciales, son los que apelan a analogías entre los seres humanos y los computadores digitales,⁶² de este modo refiere Searle sus pretensiones:

La teoría más ambiciosa para salvar el hiato es la que dice que la investigación en psicología cognitiva y en inteligencia artificial ha establecido que la mente es al cerebro como el programa del computador es al *hardware* del computador. La siguiente ecuación es muy común en la literatura: mente/cerebro = programa/*hardware*. Para distinguir este punto de vista de versiones más cautelosas de la inteligencia artificial, lo he llamado 'inteligencia artificial fuerte' ('IA fuerte', para abreviar). De acuerdo con la IA fuerte, un computador adecuadamente programado, con los *inputs* y *outputs* correctos, tiene literalmente una mente en el mismo sentido en que usted y yo la tenemos. (Searle, J., 1989: 414-415)

⁶¹ SEARLE, John, 1989: 414.

⁶² En especial se recurre a programas de 'comprensión de relatos'. Uno de los programas al que se refiere Searle, recibe como *input* un relato simple sobre el comportamiento de un hombre que ingresa a un restaurante, sin informarle al computador si el hombre comió o no en el restaurante. El computador, al ser indagado sobre este particular, entrega como *output* la respuesta acerca de si comió o no el hipotético hombre.

Searle insiste en lo difundida que está la ecuación mente/cerebro = programa/hardware, y adjudica esta extendida creencia a *"que la gente no puede ver otra forma de resolver el problema mente-cuerpo sin recurrir al dualismo"*.⁶³

El autor se enfoca en la ecuación claramente formulada mente/cerebro = programa/hardware, y se propone someterla a una, a su juicio, simple y decisiva refutación. Uno de los corolarios que se seguiría de esa ecuación, si la misma fuera correcta, es que las máquinas comprenden mensajes tal como lo hacemos los humanos. La estrategia refutatoria de Searle consiste en imaginarse él mismo como una máquina encerrada en una habitación. El argumento es conocido como *'el caso de la habitación china'*, y Searle lo sintetiza del modo siguiente:

Supóngase que estoy encerrado en una habitación. En esa habitación hay dos grandes cestos llenos de símbolos chinos, junto con un libro de reglas en español acerca de cómo aparear los símbolos chinos de una de las cestas con los símbolos chinos de la otra cesta. Las reglas dicen cosas como: 'Busque en la canasta 1 y saque un signo garabateado, y póngalo al lado de un cierto signo garabateado que saque de la canasta 2'. Adelantándonos un poco, esto se llama una 'regla computacional definida en base a elementos puramente formales'. Ahora bien, supóngase que la gente que está fuera de la habitación envía más símbolos chinos junto con más reglas para manipular y aparear los símbolos. Pero esta vez sólo me dan reglas para que les devuelva los símbolos chinos. Así que estoy aquí, en mi habitación china, manipulando estos símbolos. Entran símbolos y yo devuelvo los símbolos de acuerdo con el libro de reglas. Ahora bien, sin yo saberlo, quienes organizan todo esto fuera de la habitación llaman a la primera cesta un 'guión-de-restaurante' y a la segunda cesta un 'relato acerca del restaurante', a la tercera hornada de símbolos la llaman 'preguntas acerca del relato', y a los símbolos que les devuelvo 'respuestas a las preguntas'. Al libro de reglas lo llaman 'el programa', ellos se llaman 'los programadores' y a mí me llaman 'el computador'. Supóngase que después de un tiempo soy tan bueno para responder esas preguntas en chino que mis respuestas son indistinguibles de las de los nativos hablantes del chino. Con todo, hay un punto muy importante que necesita ser enfatizado. Yo no comprendo una palabra del chino, y no hay forma de que pueda llegar a entender el chino a partir de la

⁶³ SEARLE, John, 1989: 416.

instanciación de un programa de computación, en la manera en que la describí. Y éste es el quid del relato: *si yo no comprendo chino en esa situación, entonces tampoco lo comprende ningún otro computador digital sólo en virtud de haber sido adecuadamente programado, porque ningún computador digital por el solo hecho de ser un computador digital, tiene algo que yo no tenga*. Todo lo que tiene un computador digital, por definición, es la instanciación de un programa formal de computación. Pero como yo estoy instanciando el programa, como suponemos que tenemos el programa correcto con los *inputs* y *outputs* correctos, y yo no comprendo el chino, entonces no hay forma de que cualquier otro computador digital *sólo en virtud de instanciar el programa* pueda comprender el chino. (Searle, J. 1989: 418-419)

El argumento searleano se funda en la conocida distinción entre sintaxis y semántica de un lenguaje, esta distinción es justamente la que pierde de vista la simulación del comportamiento cognitivo de la IA fuerte. El rechazo de la ecuación tan mentada entre mente y programa reside en que *"tener una mente es algo más que instanciar un programa de computación"*,⁶⁴ y esto en razón de que las mentes tienen contenidos semánticos, y los programas no. El programa requiere sólo una definición formal, y esto en un orden solamente sintáctico.

Cabe distinguir entre la IA fuerte y un intento más austero por superar el hiato entre el sentido común y la ciencia. Este intento se denomina 'cognitivismo', y se deriva del trabajo en psicología cognitiva e inteligencia artificial, y forma la corriente principal de la 'ciencia cognitiva'. Al igual que la IA fuerte concibe *"al computador como la representación correcta de la mente, y no sólo como una metáfora. Pero a diferencia de la IA fuerte, no afirma, o al menos no tiene que afirmar, que los computadores tienen literalmente pensamientos y sensaciones."*⁶⁵ El programa del cognitivismo podría sintetizarse así: *"pensar es procesar información, pero procesar información es solamente manipulación de signos. Los computadores manipulan símbolos. Así, la mejor manera de estudiar el pensar (o, como ellos prefieren llamarlo, 'la cognición') es estudiar los programas computacionales de manipulación de símbolos, ya estén en los computadores o en los cerebros"*.⁶⁶

⁶⁴ SEARLE, John, 1989: 420.

⁶⁵ SEARLE, John, 1985: 50.

⁶⁶ *Ibidem*. Cfr. BUNGE, Mario, 2003: 193-194.

El cognitivismo ha suscitado gran adhesión, Searle atribuye esto a diversas razones: 1) nuestra tentación a utilizar la última tecnología disponible para intentar comprender al cerebro; 2) la suposición de evidencia psicológica que lo avala, sea, a) por la diferencia en tiempos de reacción que tiene una persona sometida a una tarea intelectual, fenómeno que se reitera también en una computadora, o, b) por la idea de que las reglas gramaticales que seguimos al hablar son semejantes a las reglas formales que sigue una computadora; 3) la suposición de que si el resultado de un sistema es significativo, como lo es nuestra competencia lingüística, entonces tiene que haber alguna teoría internalizada que subyace a esa actividad; 4) como no puede verse otro modo de entender las relaciones mente/cerebro, se hecha mano del modelo programa/hardware de la computadora. Quisiera detenerme en las razones por las que Searle rechaza 2b), la que se refiere al seguimiento de reglas. Si bien los seres humanos seguimos, aunque no siempre, reglas en nuestro comportamiento, no es éste el caso con las computadoras,

Hay una *diferencia crucial*. En el caso de los seres humanos, siempre que seguimos una regla, *estamos siendo guiados por* el contenido efectivo o *el significado de la regla*. En el caso de seguir una regla, los significados causan la conducta. Ahora bien, desde luego no causan la conducta enteramente por sí misma, pero ciertamente juegan un papel causal en la producción de la conducta. (Searle, J. 1985: 53-54, énfasis mío)

La discusión acerca de las pretensiones del cognitivismo es expresada por Searle bajo la siguiente moraleja:

en el sentido en que los seres humanos siguen reglas (y, a propósito, los seres humanos siguen reglas en mucha menor medida de la que los cognitivistas dicen que las siguen), en este sentido los computadores no siguen reglas en absoluto. Solamente actúan de acuerdo con ciertos procedimientos formales. (...) en el sentido literal en el que los seres humanos siguen reglas los computadores no siguen reglas, actúan solamente como si estuviesen siguiendo reglas. (Searle, J. 1985: 50)

El cognitivismo es preso de una metáfora, confundiendo el procesamiento psicológico de la información, que incluye estados mentales, con un tipo diverso de procesamiento de información (el de las compu-

tadoras) en el que acontecen procesos que son *como si (as if)* ocurriese algún procesamiento mental de información. Creer que hay en el procesamiento de información *como si* algo relevante psicológicamente (como es el caso en Dennett) sería tan descaminado como suponer que si el agua cuesta abajo puede usarse para obtener información sobre la línea de menor resistencia en la ladera de una montaña, entonces el agua cuesta abajo y la acción gravitatoria sobre ella, tendrían alguna relevancia psicológica.⁶⁷ Pues confundir la mente con sus modelos conceptuales sería semejante a que *“los físicos afirmaran que la luz calcula su trayectoria –vale decir, que integra las ecuaciones pertinentes– al moverse. O como si de los exploradores que se aventuran en tierras no cartografiadas, se esperase que descubrieran mapas en lugar de llanuras, colinas y ríos.”*⁶⁸

Ahora bien, descartadas las ecuaciones de la IA fuerte y del cognitivismo, no pareciera aún haberse cerrado el hiato con el que iniciamos esta sección. Sólo se hizo referencia a los sucesivos fallos, deteniéndonos en el más novedoso propuesto por la ciencia cognitiva. Pero, ¿qué hay del hiato?, ¿cómo cerrarlo? La respuesta searleana es que no hay tal hiato, y ésta es la razón por la cual fallan los intentos por salvarlo: *“No hay hiato alguno entre el nivel de las explicaciones intencionales y el nivel de las explicaciones neurofisiológicas”*.⁶⁹ Sin embargo, no es simplemente declamando que no hay hiato como desaparece el mismo, siguen habiendo características de nuestra concepción folk sobre la mente que se muestran recalitrantes al cierre del hiato. Searle señala estas características: a) *conciencia* (la dificultad consiste en concebir que sistemas meramente físicos tengan conciencia), b) *intencionalidad* (¿cómo pueden ser *acerca de* algo los procesos que se encuentran en un cerebro que, físicamente, sólo son ‘átomos en el vacío?’), c) *subjetividad* (¿cómo acomodar el hecho de que mi actual estado mental sea sólo accesible para mí, con la asociación de lo objetivo de una imagen científica del mundo?), d) *causalidad intencional* (suponemos que nuestros pensamientos y sensaciones tienen importancia con relación a nuestro comportamiento, pero, ¿cómo pueden entidades mentales tener una influencia física?). Cualquier teoría satisfactoria de lo mental ha de tomar

⁶⁷ Cfr. SEARLE, John, 1985: 56-58.

⁶⁸ BUNGE, Mario, 2003: 230.

⁶⁹ SEARLE, John, 1989: 421.

en cuenta estos cuatro rasgos. Searle cree hallar una teoría que, calzando con nuestra concepción científica, hace lugar, al mismo tiempo, a estas cuatro características.

Esa teoría incluye dos tesis que Searle afirma al unísono, y que resuelven, además, el problema mente cuerpo. La primera tesis reza así: "*Los fenómenos mentales, todos los fenómenos mentales, ya sean conscientes o inconscientes, visuales o auditivos, dolores, cosquilleos, picazonas, pensamientos, toda nuestra vida mental, están efectivamente causados por procesos que acaecen en el cerebro.*"⁷⁰ Mientras que la segunda tesis afirma lo siguiente: "*Los dolores y otros fenómenos mentales son sólo rasgos del cerebro (y quizá del resto del sistema nervioso central).*"⁷¹ Ahora bien, que los cerebros causen las mentes, y que las mentes sean sólo rasgos de los cerebros son proposiciones que *prima facie* parecen excluirse; sin embargo, Searle defiende que ambas proposiciones son verdaderas. Esta defensa requiere utilizar un concepto de *causa* más sofisticado que el modelo habitual (modelo del relámpago que causa el trueno), pues si utilizáramos el modelo habitual, las relaciones cerebro-mente se entenderían desde el dualismo, esto es, que eventos del reino 'físico' causan eventos de otro reino, el 'mental'. Y el defender que ambas proposiciones son verdaderas requiere además recurrir a una distinción, muy habitual en física, para poder sustentar la compatibilidad de ambas tesis acerca de lo mental. La distinción utilizada en física es la que se presenta entre micropropiedades y macropropiedades de los sistemas.

Todo sistema (pensemos en un escritorio o en un arroyo) está compuesto por micro-partículas y cada una de éstas tiene características (micropropiedades) a nivel molecular, atómico y subatómico. Pero también cada sistema tiene ciertas propiedades que son propiedades de superficie (macropropiedades), como la solidez de la mesa o la fluidez del arroyo. Algunas propiedades de superficie pueden explicarse causalmente sobre la base del comportamiento de los elementos de micronivel. Así, la solidez de la mesa es explicada (causalmente) por la estructura reticular de la que está compuesta la mesa. O, por ejemplo, la fluidez del agua es explicada (causalmente) por el comportamiento de los movimientos de las moléculas de H₂O. En estos ejemplos

⁷⁰ SEARLE, John, 1985: 22.

⁷¹ SEARLE, John, 1985: 24.

no tenemos ninguna dificultad en decir que los fenómenos de superficie son *causados por* el comportamiento de elementos del micronivel y, al mismo tiempo, que los fenómenos de superficie *sólo son* rasgos (físicos) de los sistemas en cuestión (...) la característica de superficie F es *causada por* el comportamiento de los micro-elementos M, y al mismo tiempo está *realizada en* el sistema de los micro-elementos. Las relaciones entre F y M son causales pero al mismo tiempo F es, simplemente, una característica de nivel más elevado del sistema que consiste en los elementos M. (Searle, J. 1989: 431-432)

En la historia de la ciencia suele ocurrir que una expresión que es definida originariamente en términos de las características de superficie es luego definida en función de la microestructura que la causa.⁷² Así, por ejemplo, la solidez de la mesa, esto es, que es rígida, que resiste la presión, que sostiene objetos, configura la noción de solidez de sentido común. Con el tiempo, y desde una perspectiva científica, la solidez fue entendida como una microestructura reticular que causa esos toscos rasgos observables por el sentido común. Y puede entonces afirmarse que la solidez es sólo la estructura reticular del sistema molecular y que la solidez causa la resistencia al tacto; o, puede decirse que la solidez consiste en la resistencia al tacto, y que está causada por elementos del micronivel. Searle cree conveniente aplicar estas lecciones que provee la historia de la ciencia al caso de las relaciones mente-cerebro:

Así como la fluidez del agua está causada por el comportamiento en el micro-nivel y es al mismo tiempo una característica realizada en el sistema de microelementos, exactamente en el mismo sentido de 'causado por' y 'realizado en', los fenómenos mentales son causados por los procesos que suceden en el cerebro en el nivel neuronal o modular, pero están realizados en el mismo sistema que consiste en neuronas organizadas en módulos. (Searle, J. 1989: 432-433)

Searle no pretende, de ningún modo, haber desentrañado cómo funciona el cerebro, más aún, indica que no tiene idea al respecto, tan sólo estima que,

⁷² Los ejemplos siguientes ilustran esta cuestión: el rojo es causado por emisiones de fotones con una longitud de onda de 600 nanómetros: *rojo sólo es una emisión de fotones de 600 nanómetros*. El calor es causado por movimientos de moléculas: *el calor es sólo la energía cinética media de los movimientos de las moléculas*. Sobre este tema ver SEARLE, John, 1989: 431-432.

aunque hay enormes misterios empíricos sobre cómo funciona con detalle el cerebro, no hay obstáculos lógicos, o filosóficos, o metafísicos, para dar cuenta de la relación entre la mente y el cerebro en términos que son completamente familiares para nosotros a partir del resto de la naturaleza. Nada hay más común en la naturaleza que el que rasgos superficiales de un fenómeno sean a la vez causados por y realizados en una microestructura, y éstas son exactamente las relaciones que se exhiben en la relación de la mente con el cerebro. (Searle, J. 1985: 27)

Mente y cuerpo interactúan sin ser cosas diferentes, pues los fenómenos mentales son sólo rasgos del cerebro. Puede verse esto como la afirmación conjunta de fisicalismo y mentalismo. Si se define '*fisicalismo ingenuo*' como el punto de vista de que todo lo que existe en el mundo son partículas físicas con sus propiedades y relaciones, y si se caracteriza al '*mentalismo ingenuo*' como la perspectiva que afirma que los fenómenos mentales existen realmente. A partir de esta distinción, Searle concluye que "*el mentalismo ingenuo y el fisicalismo ingenuo son perfectamente coherentes entre sí. Es más, en la medida en que sabemos algo sobre cómo funciona el mundo no son solamente coherentes; ambos son verdaderos.*"⁷³

De la semántica a la pragmática. De la exactitud a la imprecisión

Tal como señalamos al final de III, el éxito del proyecto logicista fue enfrentado mediante el recurso a lo no racional (sentimientos) como sello de lo humano. Pero es en la ulterior reflexión de uno de los mentores de este proyecto logicista, Ludwig Wittgenstein, en la que hallamos una reconsideración de las pretensiones de la lógica simbólica. Wittgenstein previene contra la imagen de una máquina lógica como un mecanismo etéreo que lo penetrara todo,⁷⁴ revestida de las cualidades de lisura y dureza que caracterizan a la lógica formal. Opta en su lugar por una noción de exactitud más lábil, más cercana a los diversos tipos

⁷³ SEARLE, John, 1985: 27.

⁷⁴ Cfr. WITTGENSTEIN, Ludwig, 1987: Par. 119.

de lógica. Ya no existe entonces una forma lógica detentadora de la inexorable dureza de un super-mecanismo, sino diversas racionalidades unidas a las distintas formas de vida. Tanto la dureza como el extremo rigor de la forma lógica no se compadecen con la vitalidad que nuestras expresiones lingüísticas adquieren en sus innumerables usos.

Estas consideraciones nos colocan ante una alternativa que podría enunciarse así: *'Aun homologando, a nivel sintáctico, nuestro cerebro a un computador digital (máquina de pensar), podremos hallar dentro de la racionalidad la marca distintiva de lo humano'*, este aserto es compartido tanto por la propuesta de Searle (IV) como por la que aquí presentamos (V). Quienes sostienen esta postura se han mantenido fieles a una tradición que ha centrado en el *lógos* la nota esencial del hombre. Sin embargo, no es el *lógos* racional que la lógica simbólica imponía como perfecto lenguaje formalizado, sino una racionalidad que campea por diversas geografías lógicas. Esta es la perspectiva que elige la filosofía analítica del lenguaje ordinario, la cual, influida por el segundo Wittgenstein, considerará insuficiente la atención al rigor formal de una lógica estudiada sólo sintácticamente, para abrirse a los matices que entraña una perspectiva semántica, y más aún, a los que acompañan a la riqueza contextual del nivel pragmático. Ya no basta, por tanto, la mera relación formal entre los símbolos, ésta es quizá la función más rigurosa de nuestra racionalidad, pero no la más perfecta; participa del rigor de lo inerte, pero carece de la perfección de lo viviente. Es este rigor de la lógica simbólica el que un computador digital puede imitar, pero no la necesaria ambigüedad que reviste nuestro lenguaje cotidiano, modelado sobre presuposiciones, creencias, unidades culturales y contextos diversos, y no sólo sobre la rígida formalización de un plano simbólico. Es en razón de esto que John Searle (tal como vimos en IV) se opone a la idea de analogar la mente humana a un programa computacional, pero no ya dirigiendo su atención a ámbitos extrarracionales como el sentimiento, sino atendiendo a una perspectiva intrarracional. Searle encuentra en la razón misma la marca que nos permite seguir distinguiéndonos de una inteligencia artificial, ya que el pensamiento humano no es reducible enteramente a la mera manipulación sintáctica de símbolos, sino que incluye además contenidos semánticos significativos a los que, por definición, no accede una máquina.

Podría pensarse quizás que, desde esta perspectiva, el hombre vuelve a ser definido como un animal racional, y que tendríamos en la

racionalidad la diferencia específica según la concibió Aristóteles o la escolástica. Sin embargo, lo que ha variado aquí es el punto de referencia en función del cual se busca esta marca distintiva. No son ahora las rocas, las plantas o los animales los que han suscitado esta búsqueda, sino una máquina 'de pensar'. Y esto no constituye un mero cambio en el punto de comparación, antes una planta ahora un computador, sino que con ello ha variado fundamentalmente la misma noción de racionalidad. Los medievales contrastaban el conocimiento racional con una facultad cognoscitiva animal (la *cogitativa*) e insistían en el modo en que la razón corrige en sus juicios la apreciación de los sentidos; es decir, eran la perfección y el rigor los que distinguían el conocimiento humano del conocimiento animal. Pero en la postura de la filosofía del lenguaje ordinario, no es la perfecta precisión la que nos distingue, sino, por el contrario, la *necesaria imprecisión* que acompaña a nuestra racionalidad. Es Ludwig Wittgenstein quizás el primero en señalar esto con claridad. Las reglas de inferencia, revestidas de una inexorabilidad mayor aún que la de las leyes naturales, pueden impedir que veamos con claridad que es posible diferenciar "*leyes de inferencia **inexorables**, es decir, precisas, de leyes de inferencia **imprecisas**, o sea, de aquellas que nos permiten una **alternativa***".⁷⁵ Si nos guiamos por las primeras, precisas e inexorables, entonces sí un programa de computador digital es homologable al pensar humano. Si, en cambio, nos atenemos a las inferencias imprecisas, surge una *racionalidad alternativa* que, por definición, excede las posibilidades de un computador digital.

Es en este tópico de las inferencias imprecisas donde podemos hallar una racionalidad alternativa de la razón lógico-formal, razón ésta que legitimada por sus éxitos pareció constituir la médula del pensar humano. Y es aquí precisamente donde radica, a mi juicio, el aporte que la Ética analítica ha realizado para la instauración de una racionalidad alternativa. La Ética analítica abrió un espacio en el cual fuera posible una argumentación moral o un legítimo desacuerdo valorativo, superando el plano meramente emotivo o persuasivo. Para ello trató de hallar nuevos modos de inferencia, modos que sin tener la rigurosidad del cálculo de enunciados, tuvieran la maleabilidad necesaria para plegarse al contexto moral. Hemos de centrarnos en dos autores británicos que tomaremos como muestras de lo que aquí queremos significar. En pri-

⁷⁵ WITTGENSTEIN, Ludwig, 1987: Par. 118.

mer lugar nos referiremos a John Austin, conocido por su teoría de los actos de habla, pero olvidado a menudo como Profesor de Filosofía Moral, y en segundo término a un filósofo moral coetáneo de Austin, Patrick Nowell-Smith.

Austin sostiene que hay ciertas cosas que no podemos decir aunque ellas no sean contradictorias; por tanto, la categoría de implicación lógica que se articula sobre las nociones de tautología y contradicción, es insuficiente para analizar estos modos en los que no podemos decir algo. Así, junto a la *implicación lógica* que me impide decir ‘todos los hombres se sonrojan pero algunos hombres no se sonrojan’, señala Austin la *presuposición* y lo que llama simplemente *implicación*.⁷⁶ La presuposición regula un uso del lenguaje que nos impide abusar del mismo, por ejemplo, si digo ‘todos los hijos de Juan son calvos’, esto presupone que Juan tiene hijos, y si no los tuviera, no habría incurrido en contradicción alguna, pero el enunciado carecería de referencia.⁷⁷ Más importante aún es la categoría lógica a la que Austin denomina *implicación*, no se trata de una relación entre proposiciones sino entre hechos. Así, por ejemplo, el *hecho* de que yo diga ‘el gato está sobre el felpudo’ implica realmente el *hecho* de que yo creo que está allí. No existe contradicción lógica si digo: ‘El gato está sobre el felpudo, pero yo no lo creo’, sin embargo, esto es algo que no anda bien, pues la aserción implica una creencia. Este caso presenta analogías con el de una expresión relevante en cuestiones morales, cual es la de decir ‘prometo’, el *hecho* de decir ‘prometo’ implica para Austin ‘la intención de hacer lo prometido’. De igual modo, algo anda mal en ‘*Creo que x es bueno, pero no lo apruebo*’, pues aunque no sea autocontradictorio, tampoco puede asignársele un ‘*perfecto buen sentido*’.⁷⁸

⁷⁶ Cfr. AUSTIN, John, 1982: 90-95; y, AUSTIN, John, 1975: 77-79.

⁷⁷ Cabe distinguir, de un modo en que Austin no lo hizo, entre dos tipos de presuposición. La presuposición semántica: un enunciado *e* presupone que *p* si, y sólo si, *e* implica *p* y la negación de *e* también implica *p*. Y la presuposición pragmática: un hablante presupone que *p* si este hablante tiene la disposición a actuar como dando por sentado que *p* es información compartida por todos los participantes en el intercambio lingüístico. Las presuposiciones pragmáticas conforman el contexto, esto es, el conjunto proposicional que los participantes asumen como dados. Dado este sentido básico de presuposición pragmática de los *hablantes*, se puede caracterizar la presuposición pragmática de las *expresiones*: una expresión *e* presupone *pragmáticamente* que *p*, si y sólo si, sería inapropiado usar *e* en un contexto en el que los hablantes no presuponen pragmáticamente que *p*. Austin se habría centrado en la presuposición semántica. Wittgenstein, en cambio, apeló, implícitamente, a la presuposición pragmática al presentar su teoría de los juegos de lenguaje. Para esta distinción véase STALNAKER, R. (1999).

⁷⁸ Cfr. AUSTIN, John, 1975: 79.

Para Nowell-Smith, nuestro segundo filósofo moral, los enfoques tradicionales llegaron a negar la posibilidad de conocimiento o de opinión racional en la ética; este indeseable efecto se debió al intento de aplicar al discurso práctico “*un aparato lógico inadecuado*”.⁷⁹ Este aparato inadecuado había sido elaborado atendiendo exclusivamente a las matemáticas o a las ciencias naturales, bajo el supuesto de que las ciencias son las que mejor ejemplifican el pensamiento humano. Guiados por el dogma lógico asentado en la dureza del rigor formal, se supuso que la lógica de todos los tipos de discurso tenía que ser idéntica, y se aplicó al discurso práctico (moral) un vocabulario de términos tales como “*proposición, seguirse necesariamente, contradecir, analítico, sintético, etc.*”.⁸⁰

Según Nowell-Smith, en el análisis de las expresiones propias del discurso práctico (expresar elecciones, valorar, aconsejar, recomendar y prometer) en el que se utilizan adjetivos como ‘bueno’ y ‘justo’, habrá que recurrir a términos diversos a los de la lógica tradicional de enunciados. Con este criterio, propone reemplazar la noción de ‘implicación lógica’ por la de ‘implicación contextual’, y la noción de ‘contradicción lógica’ por la de ‘extrañeza lógica’. La ‘extrañeza lógica’ es la penalización a la que nos hacemos acreedores cuando usamos ciertas expresiones que suponen razones, sin que concurran las razones normalmente relevantes al caso, o más aún, cuando usáramos la expresión, y, no obstante, desmintiésemos la razón correspondiente. Así, si yo insistiera a un mismo tiempo en que algo es horrible y que me agrada dicho algo, podríamos decir que incurro en una expresión ‘lógicamente extraña’. Este concepto de ‘extrañeza lógica’ es propuesto como alternativo del de ‘contradictorio en sí mismo’, en regiones en las que la sutileza del discurso práctico no es analizable en términos de una ‘lógica veritativo-funcional’. De este modo, la relación entre las expresiones ‘esto es malo’ y ‘esto no es malo’ puede analizarse en términos de contradicción, pero no así si dijera: ‘esto es malo pero no lo desapruebo’. La otra noción alternativa que aporta Nowell-Smith es la de ‘implicación contextual’⁸¹ a la que

⁷⁹ NOWELL-SMITH, Patrick, 1976: 80.

⁸⁰ *Ibidem*.

⁸¹ La idea de Nowell-Smith es que concebir a las personas como procesadores de símbolos puros es olvidar que la comprensión está requerida del conocimiento de un contexto adecuado. En este sentido se expresa también Mario Bunge: “*Si se cortan los lazos semánticos, pragmáticos y emotivos ligados a una expresión lingüística, la inteligibilidad se esfuma. En otras palabras, el procesamiento del lenguaje emerge de procesos subsimbólicos.*” BUNGE, Mario, 2003: 230.

enuncia del modo siguiente: “*un enunciado p implica contextualmente un enunciado q si cualquiera que conociese las convenciones normales del lenguaje tendría derecho a inferir q de p en el contexto en que ambos aparecen*”.⁸² Un caso de implicación contextual en el que no existe ninguna implicación lógica es el siguiente: Si Juan *dice* ‘está lloviendo’, Pedro tiene derecho a inferir (pues se halla implicado contextualmente en el *decir* de Juan) que Juan cree que está lloviendo, aunque ‘Juan cree que está lloviendo’ no se siga lógicamente de ‘está lloviendo’.

No es mi intención detenerme en un análisis minucioso de estas categorías de una lógica informal, sino tan sólo mostrar como dos filósofos morales hallaron en el análisis del lenguaje un terreno de inferencias en el cual sí cabe la *alternativa* que, según Wittgenstein, no permite un modelo inferencial demasiado preciso.

Hemos señalado que el escollo que representaba el computador digital para la pretendida exclusividad de la razón humana ha sido superado en el seno del análisis del lenguaje ordinario, y hemos hecho referencia a la contribución realizada por la ética de corte analítico en tal empresa. Ahora bien, más allá del pensamiento o no de los computadores y del aporte de la ética a la decisión de esta cuestión, estimo que existen vías ulteriores de reflexión, abiertas tangencialmente por lo hasta aquí considerado. Cuáles sean estas reflexiones, sugeridas por la exposición, es lo que pretendo señalar ahora.

En primer lugar, es posible revalorizar el aporte de la filosofía analítica, tan cuestionada como logocentrista, y que, sin embargo, nos ha mostrado la luminosa penumbra de una racionalidad no axiomatizable, y por esto mismo, profundamente humana. Por tanto, la opción racionalismo-irracionalismo no es válida en estos términos; no se trata de acudir al sentimiento, la emoción o la intuición como facultades suprarracionales, sino de ampliar los límites de la racionalidad, siendo conscientes de que estos límites tienen sus contornos borrosos como los del espacio visual. Esto no significa que se haya renunciado a la racionalidad, que todo valga y que la precisión rigurosa haya sido expulsada de la razón. Los lenguajes formales, precisos y rigurosos, cumplen su función en sectores acotados de nuestro lenguaje, y desconsiderar esto es no haber comprendido que “*nuestro lenguaje puede verse como una vieja ciudad: una maraña de callejas y plazas, de viejas y nuevas casas, y de casas con*

⁸² NOWELL-SMITH, P., 1976: 98.

anexos de diversos períodos; y esto rodeado de un conjunto de barrios nuevos con calles rectas y regulares y con casas uniformes".⁸² Estos suburbios lineales de nuestro lenguaje, el simbolismo químico, la notación infinitesimal o un lenguaje máquina como el Fortran IV, han sido diseñados por los urbanistas de nuestro lenguaje, químicos como Mendeleiev, matemáticos como Leibniz o cibernéticos como Norbert Wiener, pero este nuevo planeamiento urbanístico ha crecido en torno a la vieja ciudad de nuestro lenguaje materno, sin eliminar sus plazas y callejas irregulares. Según Wittgenstein, "cuanto más de cerca examinamos el lenguaje efectivo, más grande se vuelve el conflicto entre él y nuestra exigencia"⁸⁴ de considerar cristalina la pureza de la lógica, exigencia que imponíamos nosotros mismos y que sin embargo pretendíamos dada como resultado. La fascinación que esta exigencia de pureza cristalina nos produce se debe a que nos coloca en un hielo resbaladizo en donde falta la fricción, siendo las condiciones en cierto sentido ideales, pero impidiendo, por eso mismo, que podamos avanzar. "Queremos avanzar; por ello necesitamos la **fricción**"⁸⁵, por eso el consejo de Wittgenstein es tan elocuente: *Volvamos a terreno áspero*.

En segundo lugar, y por lo que a la ética se refiere, es posible advertir que las categorías lógicas propuestas por Austin y por Nowell-Smith promueven un cambio de perspectiva en el tratamiento que habitualmente hacemos de las cuestiones morales. Por una parte, no habrá que tener solamente en cuenta las relaciones entre las proposiciones: contradicción, implicación lógica, derivabilidad, etc., sino además, y fundamentalmente, las relaciones que se establecen entre los *hechos* lingüísticos, no lo implicado por los contenidos proposicionales, sino lo que se sigue del *hecho* de decir tales contenidos. Si esto es pertinente para la ética, como lo creo, el análisis de una decisión moral requerirá de un marco más rico que el que nos ofrece la moral clásica. Por otra parte, el sistema moral de un hombre no podría homologarse a un sistema axiomático, en el que se parta de ciertos principios-axiomas y del que se derive un juicio de conducta racional, de tal modo que si en un punto de la cadena argumentativa se entrara en conflicto con el axioma, tal juicio conductual sería irracional, precisamente por entrar en *contradicción* con el principio. En lugar de esto, las reflexiones expuestas permiten calibrar el ethos humano en tér-

⁸³ WITTGENSTEIN, Ludwig, 1988: I, par. 18.

⁸⁴ WITTGENSTEIN, Ludwig, 1988: I, par. 107.

⁸⁵ *Ibidem*.

minos de fidelidad-infidelidad a los principios, y no de coherencia o contradicción con los axiomas. Advirtiendo, por cierto, que fidelidad e infidelidad no son aquí categorías que supongan una apelación al sentimiento, sino más bien goznes en torno a los cuales gira la racionalidad moral.

Pero aun acordando con esto, podría alguien lamentarse diciendo: 'Bien, hemos encontrado en esta razón alternativa un nuevo sello antropológico, pero en lugar de una razón que detentaba la luminosa claridad de una llanura despejada, se nos ofrece ahora esta razón alternativa, apesada en los repliegues de un oscuro laberinto'. Y aun más, podría alguien detestar la aterradora intemperie a la que nos expone el renunciar al paradigma científico como modelo de la ética. Quizás nos ayude evocar una imagen mítica, me refiero a Jano, divinidad latina de dos rostros indisolublemente unidos y con diversos horizontes al mismo tiempo. Si es Palas Atenea, la de la mirada penetrante, la que ha encarnado para occidente la profundidad del intelecto humano, es, en cambio, Jano, el bifronte dios latino, quien a mi juicio mejor recoge la riqueza de una razón alternativa contextualizada como la que he pretendido presentar. Es la peculiaridad de Jano, consistente en mirar hacia delante y hacia atrás, la que permite al dios, según Séneca, considerar un asunto en todos sus aspectos. Y esto es precisamente lo que la racionalidad humana comparte con Jano, el no limitarse a la frontalidad sintáctico-semántica, sino el dirigir también su aguda mirada al horizonte pragmático, permitiendo así que en torno a esta bifrontalidad crezca la vida del *lógos* humano.

Referências

- ANDERSON, Alan Ross. **Mentes y máquinas**. México: UNAM, 1970.
- AUSTIN, John. **Ensayos filosóficos**. Madrid: Revista de occidente, 1975.
- AUSTIN, John. **Cómo hacer cosas con palabras**. Barcelona: Paidós, 1982.
- BAUDRILLARD, Jean. **La transparencia del mal**. Barcelona: Anagrama, 1991.
- BUNGE, Mario. **Emergencia y convergencia. Necesidad cualitativa y unidad del conocimiento**. Barcelona: Gedisa, 2003.
- DENNETT, Daniel. **Contenido y conciencia**. Barcelona: Gedisa, 1996.
- DENNETT, Daniel. **La actitud intencional**. Barcelona: Gedisa, 1998.

JEFFERSON, G. **The Mind of Mechanical Man**, British Medical Journal, v. 1, pp.1105-1121, 1949.

LA METTRIE, Julien. **El hombre máquina**. Buenos Aires: Eudeba, 1962.

LEIBNIZ, Gottfried. **Monadología**. Barcelona: Folio, 1999.

NOWELL-SMITH, Patrick. **Ética**. Estella (Navarra): Verbo Divino, 1976.

RABOSI, Eduardo (compilador). **Filosofía de la mente y ciencia cognitiva**. Barcelona: Paidós, 1995.

SCRIVEN, Michael. El concepto mecánico de la mente. ("The Mechanical Concept of Mind", Mind, vol.LXII, núm. 246). En ANDERSON, Alan Ross. **Mentes y máquinas**, pp. 47-63, 1970.

SEARLE, John. **Mentes, cerebros y ciencia**. Madrid: Cátedra, 1985.

SEARLE, John. Mentes y cerebros sin programas. Aparecido como "Minds and Brains Without Programs", en **Mindwaves**. Colin Blakemore y Susan Greenfield (comps.) Oxford: Blackwell, 1989, pp. 209-223. En RABOSI, Eduardo (compilador). **Filosofía de la mente y ciencia cognitiva**, pp. 413-443, 1995.

STALNAKER, R. **Content and Context**. Oxford: Oxford University Press, 1999.

TURING, A.M. (1950) Máquinas computadoras e inteligencia. ("Computing Machinery and Intelligence", Mind, vol. LIX, núm. 236) En ANDERSON, Alan Ross. **Mentes y máquinas**. México: UNAM. pp. 9-45, 1970.

WITTGENSTEIN, Ludwig. **Observaciones sobre los fundamentos de la matemática**. Madrid: Alianza, 1987.

WITTGENSTEIN, Ludwig. **Investigaciones filosóficas**. Barcelona: Grijalbo, 1988.

ZIFF, Paul (1959) Los sentimientos de los robots. ("The feelings of robots", Analysis, vol.XIX, número 3. En ANDERSON, Alan Ross. **Mentes y máquinas**, pp. 141-154, 1970.

Recebido em/ Received in :21/11/2005
Aprovado em/Approved in :04/12/2005