

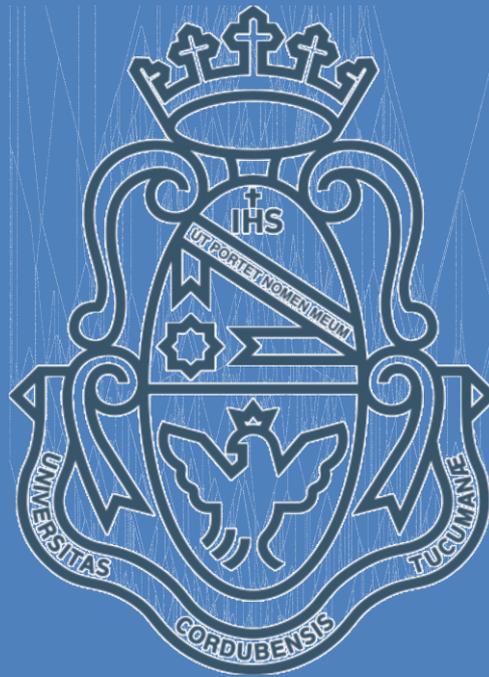
EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS V JORNADAS

1995

Alberto Moreno

Editor



ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES
UNIVERSIDAD NACIONAL DE CÓRDOBA



Esta obra está bajo una Licencia Creative Commons atribución NoComercial-SinDerivadas 2.5 Argentina



ACTITUD INTENCIONAL Y MODELOS COGNITIVOS

En la actualidad, hay dos modelos distinguibles, en el campo de la inteligencia artificial, que pretenden dar cuenta del ámbito de lo mental a partir de una de las tesis básicas de la psicología cognitiva. En términos de Haugeland (1981, p.2), ésta consiste en que "en un nivel adecuado de abstracción, una teoría de la inteligencia 'natural' debe tener la misma forma básica que las teorías que explican sistemas computacionales sofisticados". El fenómeno de la intencionalidad, esto es, la relación entre los estados mentales y el mundo, es abordado por ambos modelos.

Mi intención es mostrar en qué medida, los modelos conexionistas¹ son compatibles (en un sentido amplio) con la actitud intencional (AI), a diferencia de los modelos basados en la Teoría Representacional/Computacional de la Mente (TRCM; Fodor, 1975). Y, siguiendo la estrategia delineada por Dennett (1987, p.149) para la psicología, mostrar ciertas pautas de investigación que se desprenden de la AI (en tanto teoría de la competencia acerca de la capacidad o los 'poderes representativos' de un sistema) en dirección a los modelos de procesamiento de la información (o los 'medios representativos' de dichas capacidades).

1-AI y TRCM

Dennett (1987) propone tratar a aquellos sistemas de los cuales queremos predecir su comportamiento como 'sistemas intencionales' (SI). Un sistema es intencional si le es apropiada una estrategia o AI. Una AI le es apropiada a un sistema si podemos hacer predicciones adecuadas que no podríamos hacer desde ninguna otra actitud. Esto es, la AI consiste en una estrategia que adoptamos los seres humanos cuando no nos es posible predecir el comportamiento de un sistema desde la 'actitud física' (en base a la determinación de la constitución física del sistema y la utilización de las leyes físicas) o desde la 'actitud de diseño' (en base al fin para el que está diseñado o, de manera más explícita, apelando a una descripción de su funcionamiento).

La AI, tal como se mencionó, se utiliza para la predicción del comportamiento. Esto se realiza en virtud de la atribución de racionalidad, creencias y deseos al sistema en cuestión. Es decir, se atribuyen las creencias y deseos que un sistema debería tener, dadas las circunstancias y los objetivos del mismo, y se predice que se comportará de manera racional, esto es, en virtud de tales creencias y deseos (y del resto de las creencias y deseos del sistema).

Sin embargo, no todo sistema para el cual es posible adoptar la AI, es un 'verdadero' SI (o un "verdadero creyente"). Y, por otro lado, no sólo cabe adoptar la AI debido a nuestra falta de conocimiento en relación a la constitución física o el diseño de un sistema.

¹ Cuando se habla de modelos conexionistas, se alude, de manera global y arbitraria, a diferentes modelos cognitivos de procesamiento distribuido y en paralelo.

El criterio pragmático que avala la adopción de la AI (puesto que no habría ninguna restricción 'de principio' para considerar a cualquier sistema como intencional) consiste en el poder predictivo de la misma. La diferencia entre los sistemas genuinamente intencionales y los que no lo son, no consiste en una diferencia en relación a la constitución interna o a la complejidad del diseño sino en el fenómeno objetivo del éxito o fracaso de la AI. Hay modelos que describen el comportamiento humano desde la AI, que permiten obtener generalizaciones y predicciones de gran poder informativo para la vida cotidiana. De manera que, parece ser que sólo aquellos sistemas para los que hay modelos del comportamiento basados en la AI (seres humanos, animales y computadoras) pueden ser considerados 'verdaderos creyentes'.

Por otro lado, la AI es independiente de cualquier modelo de diseño. Podemos describir el comportamiento de un sistema desde la AI, aun cuando tengamos conocimiento del mismo desde la actitud física o de diseño, puesto que dicho conocimiento no nos permite obtener las generalizaciones, o dar cuenta del poder predictivo, de la AI. Tampoco la AI arroja ninguna luz sobre los estados o procesos mecánicos subyacentes que describen el funcionamiento de un sistema. La AI es una teoría o modelo acerca de la competencia de un sistema, esto es, una idealización de lo que el sistema debería hacer. Por lo tanto, la AI no se compromete con ninguna actitud de diseño.

Si tomamos en cuenta los niveles de análisis de la teoría psicológica cognitiva, la teoría pura del SI se ubica en el nivel 'ecológico' de Sterelny (1990). En este nivel se especifica la competencia cognitiva o función del sistema sin especificar los mecanismos de procesamiento de la información que computan o ejecutan dicha función. Es decir, se describe lo que el sistema puede hacer en virtud de su capacidad representacional en relación a aspectos del entorno (información) que el mismo puede detectar, usar, transformar, etc. Este nivel no se compromete con ninguna explicación acerca de cómo el sistema puede ejecutar la capacidad especificada por el nivel ecológico.

La teoría del SI, en tanto teoría de la competencia cognitiva, tal como se mencionó más arriba, posee un poder explicativo y predictivo sustentado en las generalizaciones que se pueden hacer desde la AI. Generalizaciones que permiten captar e interpretar de manera intencional ciertos patrones salientes del comportamiento de los seres humanos y así, atribuir creencias y deseos compartidos, sin necesidad de postular estados de creencia físicamente estructurados de forma parecida en las cabezas de los individuos. Estados intencionales (semánticamente valiables) que juegan un papel causal interno (en el funcionamiento de la mente) en relación al comportamiento.

Dennett no está dispuesto a aceptar esta caracterización de la naturaleza de las creencias. En este sentido, la teoría de la competencia intencional no se compromete con ningún tipo de contrapartida en el diseño, de las entidades postuladas por la primera (por ejemplo, las creencias). Del hecho de que utilicemos modismos intencionales, discriminativos en relación a los significados de sus partes, que sirven para describir el comportamiento, ésto no significa que dichas discriminaciones tengan que ser proyectables a estados internos físicos, específicos e individualizados, que hagan posible el comportamiento.

Las adscripciones intencionales no son una caracterización interna del sistema puesto que los significados pueden variar sin que varíe nada intrínseco al diseño, como lo ilustra el ejemplo de la

máquina de 2 bits (Dennett, 1987, pp. 256-61). Una misma máquina ante el input 'dolar' entraría en el estado Q que significa algo así como "acepto una moneda norteamericana auténtica de 25 centavos en este momento". Si la máquina es trasladada a Panamá (no pudiendo distinguir entre un dolar y un balboa del mismo valor) entraría en el estado QB que significaría "acepto un balboa panameño auténtico...". Lo cierto es que el contenido del estado interno varía en virtud de la inserción de la máquina en otro entorno sin que haya ninguna diferencia en su diseño.

Por otro lado, sería posible la atribución de creencias compartidas en sistemas con grandes diferencias de diseño, como lo ilustra el ejemplo de las réplicas de María (Dennett, 1977). Tanto a María, Ruth (una réplica exacta de los procesos computacionales psicológicamente reales de María) y Sally (un robot con un programa equivalente al de Ruth) se les puede atribuir la misma creencia a pesar de que las realizaciones individuales en el comportamiento no se correspondan (en un sentido no muy relevante de la acción). En este caso, la AI es igual de efectiva puesto que no se trata de predecir en detalle el comportamiento sino alguna especie de disposición a la conducta que permitiría atribuir las mismas creencias ante las mismas disposiciones. Esto es, la AI es caracterizada como un conductismo holista en el sentido de que lo que sucede dentro del sistema no es tenido en cuenta, es el sistema total el que es sujeto de las atribuciones intencionales. Estas no pueden ser adscriptas aisladamente sino sólo como parte de un entramado de creencias y deseos.

Dennett no es un realista acerca del contenido psicológico (hemos visto que no es un realista acerca de las creencias, tal como las entiende la TRCM), de modo que, sostiene que no hay ningún dato que permita dar cuenta realmente de lo que creemos. En este sentido, la teoría de la competencia intencional pretende caracterizar a los que creen que X sin comprometerse con un X en el cerebro o en el mundo. No hay un criterio último de individuación de contenidos, ni en la estructura de un sistema (tal como propone la TRCM) ni un hecho en el mundo (tal como propone el realismo esencialista) que nos dice cuál es el significado de un término o el contenido exacto de una creencia.

Para Dennett, las creencias no son parte de la ontología del mundo. Tienen un status especial de 'abstracta', esto es, 'constructos lógicos o en contextos de cálculo', a diferencia de la realidad de los estados y procesos neurofisiológicos. Dennett es instrumentalista en relación a las entidades postuladas en el nivel 'computacional' (Sterelny, 1990) en el que se describen los procesos que llevan a cabo la función del sistema. Las creencias no se comportan ni se individualizan tal como propone TRMC. En este sentido, los argumentos de Dennett que hemos visto se dirigen contra las siguientes tesis del modelo cognitivo TRCM²:

1) Hay un sistema representacional interno (estructurado oracionalmente) compuesto de representaciones mentales explícitas. Esto es, objetos físicamente estructurados (fórmulas de un sistema de elementos) en el lugar funcionalmente pertinente del sistema para los que hay una semántica o interpretación (Dennett, 1987, pp. 194-95).

² Las tesis mencionadas no agotan las tesis sostenidas por TRCM. Por otro lado, los argumentos de Dennett también podrían ser utilizados para atacar tesis de TRCM no mencionadas o especificadas en detalle en este trabajo.

2) Los procesos mentales consisten en secuencias de operaciones computacionales (en términos algorítmicos), en virtud de las propiedades formales de las representaciones mentales.

3) Tener una creencia (o actitud proposicional) es estar en alguna relación computacional con una representación mental interna explícita.

4) Tesis del realismo intencional: hay estados mentales (las actitudes proposicionales) "con propiedades intencionales genuinas que están implicadas de manera causal en la producción del comportamiento" (Lower & Rey, 1991, xiv).

2-AI y Modelos Conexionistas

Si bien es cierto que la teoría de la competencia es independiente del diseño del sistema, tal como se mencionó más arriba, Dennett trae argumentos conceptuales relacionados con el nivel de la competencia que juegan un papel heurístico a la hora de concebir el tipo de funcionamiento de un sistema (por ejemplo, el de la atribución holística de las creencias o el de la naturaleza de las diferentes representaciones mentales, que se mencionará luego). Por otro lado, hay argumentos que Dennett trae, en oposición a la postura realista, en el nivel del procesamiento de la información, 'computacional' o de diseño, conectados a los modelos conexionistas (sobre los que se volverá más adelante).

Las creencias son 'abstracta', en el sentido de que son interpretaciones de fenómenos reales cuando se adopta la AI. Dennett está pensando en ciertos elementos internos, en el nivel del diseño, que portan contenido informativo ('contenido sin creencias', cf. Dennett, 1987, pp.73, 201-02, 205-06). Pero, en este nivel de descripción, parecería que ya no tendría sentido hablar de un realismo intencional, puesto que los contenidos de los portadores de contenido, no darían lugar, por sí mismos, a una semántica intencional (acerca de cosas y hechos del mundo exterior) sino a una semántica interna (acerca de las operaciones y otros estados del sistema). La intencionalidad de los estados del sistema no surge del sistema mismo sino en un nivel más elevado de descripción, en el que la función que algo puede cumplir en un sistema se especifica globalmente por el papel que juega dentro del esquema de propósitos mayor del sistema. Y, en tanto que hay indeterminación funcional, "no puede haber realismo acerca de los significados sin realismo acerca de las funciones" (Dennett, 1987, pp.283-84).

Parecería que los portadores de contenido o vehículos de significado no serían representaciones en el sentido de ser acerca de algo del mundo, sino informaciones que no estarían representadas explícitamente (en este sentido, se podrían relacionar con la noción de 'constructos en contexto de cálculos', mencionada más arriba) y que cumplen el papel de informar a los subsistemas. Por otro lado, esos contenidos son asignados o individualizados de manera holística, en términos intencionales, a partir de la tarea que cumple todo el sistema (cf. Dennett, 1977, p.72, "lo que el ojo de la rana le dice al cerebro de la rana no es lo que el ojo de la rana le dice a la rana").

Dennett (1987, cap.6) está pensando en sistemas, a la manera de los modelos conexionistas, en los que la intencionalidad (o las propiedades semánticas externas) no surge por relaciones computacionales. Esto es, no surge por la manipulación de representaciones explícitas internas sino por relaciones emergentes holísticas, quizá, a partir de representaciones tácitas (portadoras de

información pero incorporadas a la capacidad del sistema, sin encarnadura física)³. De manera que, Dennett no se opone a la postulación de un sistema representacional en el cerebro ("ítems cargados de contenido", 1987, p.210), pero sí se opone a su naturaleza oracional y al tipo de relación computacional que postula TRCM.

Si de lo que se trata es de explicar el funcionamiento del cerebro, las creencias (como el resto de las nociones representacionales) no parecen tener un papel causal interno (de manera directa), en la medida en que no parece haber un mapeo directo, biunívoco, entre las creencias y los estados internos que hacen posible el comportamiento. Las representaciones a las que Dennett alude parecen consistir en vehículos sintácticos, puesto que el cerebro es una máquina sintáctica (Dennett, 1987, p.65) que portan información, se discriminan por sus propiedades formales y covarían con el significado atribuido, de manera holística, por un intérprete externo al sistema.

Ahora bien, el contenido de las representaciones mentales, tal como afirma Dennett (1977, p.76), es una "función de una función", es decir, es la función 'contenido' de la función 'portador de contenido'. En ciertas estructuras (como en los modelos conexionistas), dicha función de función puede no estar realizada por tal estructura pero intervenir o estar incorporada a la capacidad del sistema (en el sentido de la representación tácita de Dennett).

En los modelos conexionistas existen unidades ocultas (entre las unidades de input y las de output) que intervienen en el funcionamiento de un sistema pero que no poseen interpretación. En líneas generales, en dichos modelos, la información se codifica en los pesos de las conexiones entre las unidades (rotuladas) que forman redes interconectadas. No hay ningún estado particular de la red (algún conjunto de unidades) que represente un aspecto específico del entorno. Cada input nuevo que ingresa a la red modifica las conexiones entre los nodos, de manera global. Esto significa, en términos toscos, que no hay un conjunto de nodos que signifiquen 'perro' y al ingresar el input 'negro' se agregue otro conjunto de interconexiones entre nodos sino que se modifica toda la configuración de la red. Y, los pesos de las conexiones de toda la red sirven para codificar diferentes 'proposiciones'. En este sentido, el agregado de una proposición no nos permite distinguir qué estado o características del sistema codifican dicha proposición. Por ende, no tiene sentido preguntarse qué papel causal juega una proposición en particular puesto que no se corresponden con ninguna parte funcionalmente discreta del sistema (cf. Ramsey y col. 1991).

Dennett (1987, cap.6) trae a los modelos conexionistas en favor de su postura en relación a las actitudes proposicionales, pues muestran, en el nivel del procesamiento de la información, que es posible la modelización cognitiva sin ningún tipo de elementos que puedan ser identificados con las actitudes proposicionales. Sin embargo, también aclara que la semántica de dichos sistemas es interna. De manera que, con la alusión de Dennett a un nivel superior de descripción para la atribución de propiedades semánticas externas, la atribución de contenidos parecería otorgarse a la configuración estable final del sistema.

Así, parecería que hay aspectos de la postura de Dennett que lo llevan a optar, en el nivel del

³ Los argumentos en favor de la distinción entre los diferentes estilos de representación mental y la descripción detallada de lo básico de los modelos conexionistas no se desarrollan por razones de extensión del trabajo. Ver Dennett, 1987, cap.6.

diseño, por los modelos conexionistas, cuyas tesis (al menos aquellas que se han desarrollado) son compatibles con la teoría pura del sistema intencional. Las mismas se pueden resumir de la siguiente manera (cf. Tienson, 1987):

1) El sistema representacional consiste en una red de nodos interconectados, cada uno con un valor de activación. Las representaciones mentales consisten en un conjunto de nodos (representaciones distribuidas) que no representan, por sí mismos, ningún rasgo específico del entorno o alguna proposición en particular, sino que "juegan papeles múltiples, drásticamente equivocados, y en los que la desambiguación se produce sólo 'globalmente'" (Dennett, 1987, p.205)

2) Los procesos computacionales son estadísticos, en virtud del conjunto de las fuerzas de las conexiones ('pesos') entre los nodos.

3) Tener una creencia consiste en estar en un patrón de activación estable, en virtud de una secuencia causal (no sintáctica).

4) La información está en el sistema, pero no de manera explícita (o activa) sino potencialmente. Está almacenada en las fuerzas de las conexiones que se activan ante los estímulos (no se almacenan en un banco de datos, a la manera de oraciones de un lenguaje).

3-Algunas conclusiones

Hemos visto algunos de los argumentos de Dennett en oposición a TRCM y sus razones para acercarse a los modelos conexionistas. En ambos la semántica es interna al sistema. Para Dennett, no parece haber manera de que la sintaxis determine a la semántica (externa) de un sistema. Lo que se puede esperar es el diseño de sistemas que se asemejen a máquinas que discriminan significados, con elementos portadores de información que co-varíen con lo que, en términos intencionales, denominamos 'significados'. Sin embargo, hemos visto que en los modelos conexionistas (a los que Dennett acude) difícilmente se puedan encontrar elementos que no impliquen la consideración de todo el sistema en relación a la cuestión del significado.

De lo expuesto se desprende que la intencionalidad sólo parece ser derivada, dependiente de un intérprete u observador externo. En este sentido, para Dennett, los seres humanos somos como las computadoras puesto que somos máquinas, en nuestro caso, diseñadas por la naturaleza. Ahora bien, si la evolución da cuenta de por qué somos intencionales, se podría pensar que, en el nivel de diseño, se podría dar cuenta de cómo llegamos a ser SI. Pero, si en el nivel de diseño no parece haber manera (al menos por ahora, dados los argumentos conceptuales de Dennett y los modelos cognitivos actuales) de establecer algún tipo de relación (que no sea vía una teoría de la interpretación) entre los estados mentales y el mundo. Y, si esto último es lo que se entiende por el problema de la intencionalidad (es decir, la relación entre las propiedades sintácticas y las semánticas (externas) de los estados mentales), parecería que lo único que tenemos, por ahora, es la AI y los modelos de diseño compatibles con las tesis de la primera.

Asumir las consecuencias de la AI, en tanto teoría de la competencia intencional, significa, entre otras cosas, asumir que somos máquinas con el programa formal correcto y una interpretación adecuada. Así, podemos ser máquinas semánticas porque manejamos significados, actuamos 'como si' entenderamos. Nuestra intencionalidad es derivada, al igual que la de las computadoras. Luego, parece ser que la intencionalidad no sería esa X que les falta a las computadoras para asemejarse a nosotros (cf. Haugeland, 1981).

REFERENCIAS

- DENNETT,D. (1977) "A Cure for the Common Code?", en N. Block, Readings in the Philosophy of Psychology, vol. II, Cambridge, Mass., Harvard University Press, 1981, pp.65-77.
- DENNETT,D. (1987) La actitud intencional, Barcelona, Gedisa, 1991.
- FODOR,J. (1975) El lenguaje del pensamiento, Madrid, Alianza, 1984.
- HAUGELAND,J. (1981) "Semantics Engines: An Introduction to Mind Design", en J. Haugeland (ed.), Mind Design, Cambridge, Mass., The MIT Press, pp.1-34.
- LOWER,B. & REY,J. (1991) Meaning in Mind. Fodor and his Critics. Oxford, Blackwell.
- RAMSEY,W.; STICH,S.; GARON,J. (1991) "Connectionism, Eliminativism, and the Future of Folk Psychology", en W. Ramsey; S. Stich y D. Rumelhart (eds.), Philosophy and Connectionism Theory, Hillsdale, NJ, Lawrence Erlbaum, pp.199-228.
- STERELNY,K. (1990) The Representational Theory of Mind, Oxford, Basil Blackwell.
- TIENSON,L. (1987) "An Introduction To Connectionism", en The Journal of Philosophy, vol. xxvi, suppl., pp. 1-16.