

EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS VI JORNADAS
(1996)

Marisa Velasco
Aarón Saal
Editores



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



LA CUESTION DE LOS NIVELES EXPLICATIVOS DE UNA TEORIA COGNITIVA

Los modelos desarrollados en el marco de la Inteligencia Artificial (IA) pueden considerarse, y en general los investigadores en ciencia cognitiva así lo creen, como un ámbito de investigación para la psicología cognitiva y la filosofía de la mente. La línea directriz que comparten con el resto de las ciencias cognitivas consiste en que una teoría de la inteligencia comprenderá tanto a las personas como a las computadoras pues ambos sistemas son manifestaciones de un mismo fenómeno subyacente.

La tarea que se intenta realizar en IA consiste en caracterizar a un sistema cognitivo de manera de entender cómo puede lograr X (X refiere a cualquier logro cognitivo humano). En relación a esta caracterización del sistema, algunos investigadores consideran que el objetivo principal de una teoría psicológica consiste en la especificación de la competencia del sistema. Es decir, una formulación abstracta de la función cognitiva que computa. En este sentido, defienden, en términos de Sterelny (1990, p.198), una 'autonomía metodológica': no hay que preocuparse por la implementación computacional o física, pues, en principio, no aportaría nada sustancial a la descripción de la capacidad del sistema.

O bien, algunos defienden una 'autonomía epistémica' (Sterelny 1990, p. 198-9): las teorías psicológicas sólo requieren de evidencias psicológicas. Aunque esto no significa afirmar que la evidencia de otras ciencias (i.e. las neurociencias) es irrelevante, es una manera más de afirmar que los detalles relacionados con la implementación no enriquecen la teoría psicológica.

Ambas versiones de la autonomía parecen asentarse en la distinción de Marr relacionada con los niveles explicativos de una teoría cognitiva. Quiero empezar por analizar esta distinción mostrando que corresponde a una metodología de investigación presupuesta en los modelos cognitivos clásicos y que es en base a ella que se sostienen las versiones de la autonomía mencionadas. Luego, quisiera sugerir que los modelos desarrollados dentro del marco conexionista parecen desfigurarse tal distinción, o en sentido más débil, invertirla, con la consecuencia de que tales versiones de la autonomía no tendrían cabida en sus posibles teorías cognitivas.

La distinción y el cognitivismo clásico

D Marr (1982, 1977) distingue tres niveles de explicación desde los cuales entender la/s tarea/s que lleva a cabo cualquier sistema de procesamiento de información:

1-Nivel computacional: análisis abstracto (formal) del problema que va a ser computado. Se relaciona con la pregunta de qué va a ser computado y por qué. El análisis se dirige a la naturaleza del problema para el cual la teoría es una solución. La teoría es entendida a la manera de una teoría de la competencia (en términos chomskianos).

2-Nivel algorítmico: cómo se implementa la teoría. Se especifica la representación para el input y el output y el algoritmo para la transformación.

3-Nivel del hardware: cómo puede realizarse físicamente la representación y el algoritmo.

Muchos autores se han servido de esta distinción respetando a grandes rasgos los lineamientos de Marr pero, al mismo tiempo, acentuando el carácter de independencia relativa de cada nivel. El nivel 1 es independiente o neutral en relación al nivel 2. También aceptan que el nivel 2 depende del 3, y que, en algún sentido, tanto 2 como 3 imponen restricciones al nivel 1 (i.e. en términos del límite de información que se puede almacenar o si el sistema funciona en serie o en paralelo).

Una de las cuestiones que quiero tratar radica en si el nivel 1 sugiere la forma de procesamiento, es decir, si afecta al nivel 2. En principio, acepto que una teoría en el nivel 1 es independiente del resto de los niveles puesto que se puede describir la función de un sistema como una idealización de lo que el sistema debería hacer y, en este sentido, capturar generalizaciones que le permitan explicar y predecir su comportamiento, de manera de no comprometerse con ningún modelo en los niveles restantes.

Sin embargo, hay varios puntos que quisiera aclarar y que parecen sugerir que sería una actitud ingenua sostener que las teorías del nivel 1, en el caso en que se puedan obtener, no guían, en algún sentido, la búsqueda y la forma de las teorías del nivel 2.

En primer lugar, la distinción de Marr responde a una estrategia metodológica utilizada por la investigación clásica en IA, denominada 'de arriba-abajo'. Esta consiste en aislar un problema, formular una teoría computacional, construir un algoritmo que la implemente y mostrar empíricamente que el algoritmo tiene éxito. Si uno acepta esto, tiene un criterio para evaluar los logros. Uno puede preguntarse qué problema interesante se aisló y si hay una teoría clara para el mismo y cuáles son sus argumentos.

Si uno se toma el trabajo de analizar algunas de las teorías del nivel 1 que se basan en esta estrategia, tal como lo hace Clark (1990), se puede observar que hay una especie de conexión lógica entre ellas y las del nivel 2. Por otro lado, Marr (1982) sostiene que los niveles están "relacionados lógicamente y causalmente"¹, punto que parece haber sido soslayado por sus lectores. Por su parte, Dennett (1987) dice que el sueño clásico es una "cascada triunfal a través de los tres niveles de Marr"².

El tipo de teoría 1, en IA, no consiste sólo en especificar la función a computarse, es decir, dado un input el sistema dará un output específico. Analizar la tarea de un sistema sólo en estos términos no parece conducir a una teoría interesante de la competencia y menos en el espíritu del cognitivismo. Un tipo de teoría 1 interesante, en la IA clásica, no parece asimilarse

¹Citado en Rumelhart y MacClelland (1986) p. 122.

²Citado en Clark (1990), pp. 281-308

a esto puesto que se realiza una mayor especificación de lo que el sistema tiene que hacer, con lo cual se orienta la investigación en el nivel 2.

La teoría 1 clásica, además de especificar la función, especifica un cuerpo de información usado por alguna clase de algoritmos. Esto es, definen procesos computacionales que el sistema tiene que llevar/lleva a cabo (cf. Clark, 1990). La ambigüedad en el verbo 'llevar' carga en sí todo el peso de la conexión entre los niveles 1 y 2. Puesto que el cognitivismo clásico (Fodor y Pylyshyn, 1988) postula en el nivel 1, la existencia de un lenguaje del pensamiento (LDP), esto es, una estructura de datos o representaciones simbólicas que tienen una sintaxis y una semántica combinatoria y reglas de transformación que se aplican a las representaciones en virtud de su forma, el nivel 2 tiene que incluir determinadas operaciones computacionales y simbólicas que se aplican a representaciones explícitas (instanciadas físicamente en el sistema).

La postulación de un LDP implica que hay que hacerlo intervenir en una clase de algoritmos que reflejan el conjunto de reglas de transformación. Estas últimas son, por ejemplo, del tipo: Si (A y B), entonces (A) y en el caso de la formación de verbos en el tiempo pasado en inglés. Si (raíz + desinencia), entonces (raíz+ -ed)³. La teoría de la competencia chomskiana también parece suponer un grado más allá del nivel 1, pues describe la forma de procesamiento del lenguaje en términos de un sistema de principios y parámetros asociados (cf. Chomsky, 1986).

Clark (1990, p.284) afirma que las teorías 1 de la cognición clásica son, en verdad, teorías de nivel 1.5. Esto trae como consecuencia, quizá indeseada, pretender que el sistema refleje de manera casi literal las especificaciones postuladas en el nivel 1.5. No parece haber manera de que el sistema cognitivo realice la función especificada sin postular un tipo de mecanismo de procesamiento que realiza determinadas operaciones sobre símbolos depositados en su 'cabeza'. Símbolos que son la contrapartida del contenido de las adscripciones de sus estados cognitivos.

Antes de pasar al siguiente apartado, quisiera hacer algunas observaciones. Uno puede sostener que el hecho de que las teorías en IA sean de nivel 1.5 es, justamente, una cuestión de hecho que no atañe al tema de cómo tiene que ser una teoría de la competencia. Pero, el punto es que no tengo claro que una teoría de la competencia tenga que ser de determinada manera. Veremos que para los desarrollos conexionistas, una teoría de la competencia puede ser entendida en términos distintos.

Por otro lado, el hecho de que diferentes algoritmos pueden llevar a cabo la misma función especificada en las teorías de la competencia, no es indicio fuerte de la independencia absoluta del nivel 1 puesto que lo que estoy diciendo no es que el nivel 1 determina un/os algoritmo/s específico/s que llevan a cabo la función.

Quisiera aclarar que el nivel algorítmico no hay que entenderlo como la elección de un algoritmo a secas, sino que consiste en especificar la representación de la información y los procesos involucrados en la cognición. Es el nivel en donde nos comprometemos con hipótesis

³La cláusula si-entonces especifica la operación a la manera de una gramática pero no tiene que estar representada explícitamente, tal como Fodor y Pylyshyn (1988, p. 61) afirman.

empíricas acerca de cuáles propiedades de un sistema le permiten llevar a cabo las capacidades cognitivas (cf. Rumelhart y MacClelland, 1986, p. 122).

La distinción y el conexionismo

Quisiera explorar la idea de que en el marco conexionista los niveles de análisis están íntimamente relacionados y se entremezclan, de manera que la distinción no tiene mucha significación metodológica. O bien, el orden de la estrategia del cognitivismo clásico se invierte. O bien, habría que ver la relación entre las teorías de niveles 1 y 2 desde otra perspectiva.

Los conexionistas, en líneas generales, se ocupan del diseño de sistemas cuya explicación se desarrolla en el nivel algorítmico. Consisten en una red de nodos interconectados en términos de excitación e inhibición. Cada nodo posee un valor numérico de activación. El nivel de activación de cada nodo da lugar a patrones de activación que junto con los valores numéricos de las conexiones da lugar al estado de la red. Dado un input determinado, la activación se propaga por toda la red y concluye asentándose en un estado estable.

En los modelos que hacen uso de representaciones distribuidas, la interpretación semántica se asigna a los estados de la red (a un conjunto de nodos). Los nodos individuales pueden no tener una función representativa o tan sólo consistir en micro-características que no constituyen el análisis de la característica representada por la totalidad de la red. Los objetos de la computación no son las representaciones (como en el modelo clásico) sino que éstas se producen en respuesta a determinados inputs. En este sentido, la distinción entre la información procesada y los mecanismos de procesamiento no tiene sentido puesto que la información se va reflejando en las distintas configuraciones del sistema y no se almacena en una estructura de datos sino que están en las conexiones y se configuran dado el input pertinente.

Algunos teóricos clásicos consideran que el conexionismo deja sin explicar los fenómenos cognitivos puesto que no parecen poseer ninguna teoría del nivel 1 que les permita un entendimiento de los problemas que el sistema intenta resolver y por ende, sólo se trata de un nuevo medio de implementación. La estrategia de los conexionistas para defenderse de esta objeción parece ramificarse:

(1) Aquellos que afirman que hay una teoría computacional implícita pero que aun no está desarrollada de la manera en que lo está la teoría del nivel 1 del cognitivismo clásico (cf. Rumelhart y MacClelland, 1986) y por eso, aun no pueden arrojar luz sobre la estructura de los fenómenos cognitivos.

Una corriente afirma que la teoría 1 conexionista implica la inversión de la estrategia de Marr. El conexionismo empieza con teorías del nivel 0.5, es decir, con un mínimo de entendimiento de la tarea, pasan al nivel 2/3 entrenando una red para realizarla y luego hay que intentar la búsqueda de principios generales que nos permitan entender el qué y el por qué de la actividad de la red.

En este sentido: (a) Se acepta que una explicación en el nivel 2 (en el que se especifica la configuración de la red en términos de sus nodos y las fuerzas de las conexiones) no parece ser una explicación cognitiva (cf. Clark 1990). (b) Se afirma que las propiedades del sistema cognitivo (por ejemplo, la capacidad de las redes de poder manejarse con datos incompletos o

erróneos, el procesamiento sensorial en tiempo real, la posibilidad de realizar generalizaciones) restringen y condicionan la manera de conceptualizar la tarea cognitiva. Por otro lado, de acuerdo a los elementos primitivos que se postulan en el nivel 2, su interacción puede dar lugar a la emergencia de ciertos elementos que no hace falta postular en el nivel 1, con lo cual las formas de las teorías 1 pueden modificarse. A partir de (b), se podría pensar que (c) Una teoría psicológica idealizada, sin tomar en cuenta las restricciones empíricas, nos lleva a soluciones elegantes, bien entendidas desde un punto de vista intelectual, pero alejadas de soluciones naturales para problemas de diseños naturales. No serán un modelo de la cognición humana (cf. Dennett 1984; Clark, 1989).

Otra corriente afirma que las teorías 1 no se asemejarán a las del cognitivismo clásico, no serán en términos de un conjunto de proposiciones bien definidas sino más bien con el estilo pictórico de una "descripción geométrica de la forma de una parte del terreno cognitivo" (Clark, 1990, p. 302; Churchland, 1989) o explicaciones matemáticas-estadísticas de cómo emergen los fenómenos de alto nivel.

(2)-La primacía asignada al nivel 1, por los cognitivistas clásicos, no es tal. Las explicaciones interesantes desde el punto de vista psicológico se encuentran en el nivel 2 de las hipótesis empíricas que explican los datos psicológicos.

Por extraño que parezca, es posible sacar consecuencias disímiles de esta afirmación. (a) El cuestionamiento de hasta dónde llega la cognición. Cómo lo que en principio pareció no ser cognitivo o no entrar en el ámbito de las explicaciones cognitivas pasa a formar parte de las mismas, ya sea con pleno derecho o afectando el carácter de lo que se considera cognitivo. Esto se relaciona con el hecho de que los cognitivistas clásicos sostienen que los sistemas conexionistas permiten dado un input determinado generar las representaciones correctas y que la tarea cognitiva recién comienza cuando la información ya está representada de manera simbólica. (b) No estoy segura de que pueda querer decir esto pero quizá la cognición no pueda entenderse en sus propios términos. Según Cummins (1989), difícilmente se encuentre una teoría científica con leyes propias de cada dominio cognitivo dadas las dificultades en la especificación de las funciones cognitivas. Si la cognición es simulación (via instanciación) de funciones naturales y, si no es posible especificar de manera independiente las funciones naturales⁴, entonces el estudio científico de la cognición caerá en leyes generales de la física o en reglas que nos permitan dar cuenta de cómo realizar las tareas pero sin tener una teoría especial en el nivel 1.

(3) Los niveles 1 y 2 se relacionan a la manera en que la mecánica newtoniana se relaciona con la física cuántica. Las teorías 1 clásicas o de nivel macroscópico son aproximaciones al nivel microscópico. El sistema se comporta como si manipulara transformaciones entre representaciones, al igual que el universo, bajo ciertas condiciones, se comporta como si fuera newtoniano. La teoría macroscópica posee un grado de adecuación descriptiva y predictiva del comportamiento de los sistemas cognitivos, pero, a nivel

⁴Las funciones naturales como el reconocimiento de caras, por ejemplo, difícilmente están gobernadas por leyes especiales que permitan una autonomía, es decir, ser formuladas en sus propios términos.

microscópico, la actividad del sistema es otra. Los fenómenos de la macroteoría emergen de la actividad descrita en la microteoría. (cf. Rumelhart y MacClelland, 1986, Clark, 1990).

(4) Los niveles están relacionados y se restringen entre sí de manera tal que cualquier nivel se puede aplicar en cualquier estadio de la investigación (cf. P.S. Churchland, 1989; Dennett, 1987; Bechtel, 1990, Rumelhart y MacClelland, 1986). Tanto las cuestiones relacionadas con la implementación como las evidencias de las neurociencias intervienen en la elaboración de las teorías cognitivas. En términos precisos, no hay un único nivel que sea 'el nivel cognitivo'. Puede haber teorización cognitiva en niveles inferiores al simbólico tradicional del cognitivismo clásico, al mismo tiempo que una caracterización en términos neuronales o implementacionales de las actividades de alto nivel cognitivo. La idea de que la naturaleza de la investigación se desarrolla de acuerdo al nivel en el que ocurre no parece el principio metodológico adoptado actualmente. Según Bechtel (1990) la diferencia entre las investigaciones cognitivas y neuronales (y extendiendo su afirmación a las cuestiones de implementación) depende más de las clases de preguntas y herramientas que se utilicen.

Conclusiones

Tanto la estrategia 'arriba-abajo' de Marr como la idea muy similar de lo que tiene que ser una explicación psicológica en Fodor (1968) parecían querer guiar a la IA hacia la concepción de teorías cognitivas abstractas e idealizadas de todo condicionamiento de diseño de un sistema cognitivo.

Sin embargo, la formulación de una teoría que establece computaciones sobre un LDP nos lleva a buscar diseños que reflejen dicho lenguaje. Hay una relación directa entre los elementos postulados en el nivel 1 y los del nivel 2. Esto nos hace pensar que, por un lado, hay un solo nivel de implementación, lo cual es simplificar la cuestión. Marr (1977) reconoce que una teoría que involucra una descripción en cada nivel puede involucrar teorías intermedias para las cuales puede o no haber una descripción en términos de los tres niveles. Si pensamos en la organización del sistema neuronal, en cada nivel estructural se puede preguntar por la función.

Por otro lado, desde el punto de vista clásico, la relación directa entre los niveles 1 y 2 hace que naturalmente se piense que el nivel 2 no aporta nada que no se haya especificado en el nivel 1. Sin embargo, en los modelos conexionistas, el nivel 2 parece querer decirnos que no hace falta postular los primitivos de una teoría 1 clásica puesto que las propiedades cognitivas surgen de la actividad del sistema.

En este sentido, las restricciones o evidencias de los sistemas biológicos y artificiales pueden enriquecer la teoría de la cognición pues aluden a las capacidades que tiene que tener un sistema para ser cognitivo. Y, es de esto último, al fin y al cabo, de lo que se ocupan los que investigan en ciencia cognitiva.

REFERENCIAS

- Bechtel, W. (1990) "Connectionism and the Philosophy of Mind: An Overview", en W. Lycan (ed.) Mind and Cognition, Cambridge, Mass., Basil Blackwell, pp. 252-273.
- Clark, A. (1990) "Connectionism, Competence, and Explanation", en M. Boden (ed.) The Philosophy of Artificial Intelligence, Oxford University Press, pp.281-308.
- Cummins R. (1989) Meaning and Mental Representation, Cambridge, Mass., The MIT Press.
- Churchland P. S. y Sejnowski T. (1989) "Neural Representation and Neural Computation", en W. Lycan (ed.) Mind and Cognition, Cambridge, Mass., Basil Blackwell, 1990, vol. 1, pp. 224-252
- Chomsky N (1986) El conocimiento del lenguaje, Madrid, Alianza, 1989.
- Dennett D. (1987) La actitud intencional, Barcelona, Gedisa, 1991.
- Dennett D. (1984) "Cognitive Wheels: The Frame Problem of IA", en M. Boden (ed.) The Philosophy of Artificial Intelligence, Oxford University Press, 1990, pp.147-170.
- Fodor J. (1968) La explicación psicológica, Madrid, Cátedra, 1991.
- Fodor J. y Pylyshyn Z. (1988) "Connectionism and cognitive architecture: A critical analysis", Cognition, 28:3-71.
- Marr D. (1977) "Artificial Intelligence-A personal view", en J. Haugeland (ed.) Mind Design, Cambridge, Mass., The MIT press, 1981, pp. 129-142
- Marr D. (1982) Vision, San Francisco, Freeman.
- Rumelhart D. y MacClelland J. (1986) "PDP Models and general issues in Cognitive Science", en D. Rumelhart y J. MacClelland (eds.) Parallel Distributed Processing, Cambridge, Mass., The MIT Press, vol 1, pp.110-146.
- Sterelny K. (1990) The Representational Theory of Mind, Oxford, Basil Blackwell.