



Sophia, Colección de Filosofía de la Educación

ISSN: 1390-3861

ISSN: 1390-8626

revista-sophia@ups.edu.ec

Universidad Politécnica Salesiana

Ecuador

Bernache Maldonado, Fabián  
La función explicativa de la noción de representación interna  
Sophia, Colección de Filosofía de la Educación, núm. 31, 2021, Julio-, pp. 265-290  
Universidad Politécnica Salesiana  
Cuenca, Ecuador

DOI: <https://doi.org/10.17163/soph.n31.2021.10>

Disponible en: <https://www.redalyc.org/articulo.oa?id=441867990010>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto

# LA FUNCIÓN EXPLICATIVA DE LA NOCIÓN DE REPRESENTACIÓN INTERNA

## The explanatory function of the notion of internal representation

FABIÁN BERNACHE MALDONADO\*

Universidad de Guadalajara, Guadalajara, México

[fabian.bernache@academicos.udg.mx](mailto:fabian.bernache@academicos.udg.mx)

Código Orcid: <https://orcid.org/0000-0001-7158-892X>

### Resumen

El objetivo de este trabajo es presentar una objeción a uno de los principios centrales de la Teoría Representacional de la Mente (TRM): la idea de que la noción de representación interna tiene una función primordial en la explicación de la actividad cognitiva. De acuerdo con la TRM, la vida cognitiva de un organismo consiste esencialmente en la formación, procesamiento y almacenamiento de representaciones internas. Tales representaciones son vistas como objetos o eventos concretos capaces de influir causalmente en los procesos cognitivos de los organismos. Expuesta en forma de dilema, la objeción pretende mostrar las dificultades inherentes a la postulación de representaciones internas, dada la manera en que estas representaciones y su operación han sido concebidas en el marco mismo de la TRM. La cuestión que introduce al dilema es: ¿en virtud de qué propiedades una representación interna influye en la actividad cognitiva de un organismo? Dos respuestas son posibles: en virtud de sus propiedades representacionales o en virtud de sus propiedades no representacionales. Empleando una metodología argumentativa, se mostrarán las consecuencias problemáticas de cada una de estas respuestas al dilema formulado y se discutirán dos ejemplos importantes de la literatura para ilustrar estas dificultades. La principal conclusión del artículo es que la noción de representación interna es incapaz de satisfacer la función explicativa que le ha sido asignada por la propia TRM.

### Palabras clave

Cognición, explicación, información, función, normatividad, causalidad.

**Forma sugerida de citar:** Bernache, Fabián (2021). La función explicativa de la noción de representación interna. *Sophia, colección de Filosofía de la Educación*, 31, pp. 265-290.

\* Profesor investigador titular en el Departamento de Filosofía de la Universidad de Guadalajara. Doctor en Filosofía y Ciencia Cognitiva por el Instituto Jean Nicod, París, Francia. Miembro del Sistema Nacional de Investigadores del Consejo Nacional de Ciencia y Tecnología de México. Sus intereses de investigación se centran en la noción de representación interna y en las teorías filosóficas y empíricas sobre el razonamiento y la argumentación.

### Abstract

The aim of this paper is to present an objection to one of the main principles of the Representational Theory of Mind (RTM): the idea that the notion of internal representation has a central function in the explanation of cognitive activity. According to the RTM, the cognitive life of an organism basically consists in the formation, processing, and storage of internal representations. Such representations are viewed as concrete objects or events that are able to causally influence the cognitive processes of organisms. Presented as a dilemma, the objection aims to show the intrinsic difficulties of the postulation of internal representations, given the way in which these representations and their operation have been conceived in the framework of the RTM itself. The question that introduces to the dilemma is: in virtue of which properties does an internal representation influence the cognitive activity of an organism? Two answers are possible: in virtue of its representational properties or in virtue of its no representational properties. Employing an argumentative methodology, the problematic consequences of both answers to the dilemma will be shown and two important examples from the literature will be discussed to illustrate these difficulties. The main conclusion of the paper is that the notion of internal representation is unable to satisfy the explanatory function that has been assigned to it by the RTM itself.

### Keywords

cognition, explanation, information, function, normativity, causality.

266



## Introducción

De acuerdo con la Teoría Representacional de la Mente (TRM), la vida cognitiva de un organismo consiste esencialmente en la formación y almacenamiento de representaciones y en la aplicación de diversas operaciones a dichas representaciones, de acuerdo con autores como Sterelny (1990), Fodor (1998) y Fodor y Phyllyshyn (2015). Dado un individuo o una categoría A, la capacidad de representarse A consiste esencialmente en la posibilidad de formar —en la mente o en el cerebro— representaciones de A. Pensar en A, en un momento  $t$ , es formar una representación de A, en el momento  $t$ , en la memoria de trabajo. Conservar información acerca de A es conservar, en la memoria de largo plazo, ciertas representaciones compuestas (proposicionales) que incluyen entre sus componentes representaciones de A. Creer que hay un A en mi bolsillo es tener la disposición a emplear, en circunstancias y procesos específicos, una representación interna cuyo contenido es ‘hay un A en mi bolsillo’ de una manera determinada que difiere sistemáticamente de la manera en que esta misma representación (u otra con el mismo contenido) sería empleada si, en vez de creer, quisiera que hubiera un A en mi bolsillo.

Más allá de las dificultades que se presentan cuando estas ideas son discutidas en detalle, uno de los grandes atractivos de la TRM es la posibilidad que ofrece de reducir el problema filosófico de la cognición a una sola cuestión fundamental: ¿qué explica, en última instancia, que algo sea una representación? Como fue señalado, para los partidarios de la TRM,

la posesión de capacidades representacionales deriva esencialmente del poder de formar y procesar representaciones de cierto tipo (internas o mentales). Por consiguiente, los mecanismos o propiedades que, en última instancia, explican que algo sea una representación no pueden, desde esta perspectiva, depender de la posesión previa de capacidades representacionales. ¿En qué pueden, entonces, consistir tales mecanismos o propiedades? De acuerdo con una concepción clásica de la representación, la semejanza entre ideas y objetos es lo que, en última instancia, explica que algo sea una representación, pues, por un lado, dicha semejanza conecta las ideas con sus objetos y, por el otro, las ideas son representaciones originarias, es decir, representaciones de las que derivan los otros tipos de representaciones, particularmente las representaciones lingüísticas, como Fodor y Lepore (1991) y otros autores sostienen.

La semejanza entre ideas y objetos parece satisfacer las exigencias de la TRM, pues no es obvio que la mera existencia de tal semejanza implique que los organismos en cuyas mentes las ideas han sido formadas posean ya capacidades representacionales. De ahí que aún sea posible encontrar en algunos autores contemporáneos, como Cummins (2010) y Johnson-Laird (2006), cierta adhesión a formas sofisticadas de la concepción clásica de la representación. La concepción clásica tiene, sin embargo, sus dificultades. Por ejemplo, dado que la semejanza admite grados, ¿qué grado de semejanza debe existir entre una idea y un objeto para que la idea pueda ser considerada una representación del objeto? Además, la relación de semejanza es simétrica: si A es semejante a B, B es semejante a A. Por consiguiente, como Fodor (1984) ha señalado, si una idea es una representación de un objeto en virtud de la semejanza que existe entre ambos, el objeto tendría también que ser considerado una representación de la idea.

Pero existen otras propuestas influyentes que no hacen uso de la noción de semejanza. Una de ellas es la semántica informacional de Fred Dreske (1981). De acuerdo con la semántica informacional, aquello que explica que un evento de tipo R represente un evento de tipo A son las relaciones nomológicas que existen entre las instancias de R y las instancias de A pues, en virtud de tales relaciones, las instancias de R son capaces de vehicular información proveniente de las instancias de A. Supóngase, por ejemplo, que R sea la activación de una estructura neuronal que no puede (nomológicamente) producirse sin que un evento de tipo A, a través de cierta estimulación sensorial, la haya causado. Dada tal situación, podemos afirmar que R es un indicador de A, es decir, que R vehicula la información según la cual un A ha acontecido. Según la semántica informacional, en virtud de esta relación entre R y A, R puede ser considerada una representación de A.



Al igual que la concepción clásica de la representación, la semántica informacional enfrenta problemas importantes. El más conocido de ellos es el problema del error, señalado por Fodor (1984): mientras que las representaciones pueden ser verídicas o erróneas, las relaciones nomológicas pueden existir o no existir, pero no pueden ser incorrectas o equivocadas. Por consiguiente, no parece ser cierto que, en virtud de las meras relaciones nomológicas que existen entre R y A, R pueda ser considerada una representación de A.

Varias nociones complementarias han sido introducidas para reforzar la semántica informacional y dar una solución al problema del error, como las nociones de causa estructurante (Dretske, 1988), dependencia causal asimétrica (Fodor, 1990) o causas incipientes (Prinz, 2002). Sin embargo, el objetivo de este trabajo no es discutir estas propuestas u otras distintas, sino formular una objeción general a la TRM. Dicha objeción, que será presentada en forma de dilema, está dirigida a su tesis central: la idea de que el fundamento explicativo de la cognición es la noción de representación interna.

Desde luego, diversas objeciones ya han sido planteadas a la TRM, algunas de las cuales han sido consideradas retos importantes por los teóricos representacionalistas mismos, como Godfrey-Smith (2006). Entre las objeciones recientes más discutidas está el llamado ‘problema difícil del contenido’ (*Hard Problem of Content*) que plantean Hutto y Myin (2013). Este problema es también presentado en forma de dilema. Hutto y Myin sostienen, por un lado, que las concepciones de la información aceptables desde una perspectiva naturalista son insuficientes para fundar una auténtica noción de representación interna y, por el otro, que las concepciones de la información que permitirían fundar una auténtica noción de representación interna son inaceptables desde una perspectiva naturalista. Así, según estos autores, en el estado actual de la investigación sobre la noción de representación interna, o bien se conserva el naturalismo y se abandona la TRM, o bien se conserva la TRM y se abandona el naturalismo. El problema que se plantea aquí es distinto y más radical aún, pues no depende de la aceptación del naturalismo, ni se basa en los detalles de las concepciones existentes de la representación interna. Lo que se pretende es mostrar que la noción de representación interna, independientemente de la manera en que es fundamentada, no es capaz de cumplir la función explicativa que le atribuyen los partidarios de la TRM. En la sección 1 de este trabajo se procederá a la presentación del dilema planteado a los partidarios de la TRM y se discutirán detalladamente las consecuencias problemáticas de cada una de las dos respuestas que pue-



den darse. En la sección 2 se presentarán dos ejemplos de representantes importantes de la TRM para ilustrar estas dificultades. Estas discusiones serán la base de la conclusión central del artículo, a saber: que la noción de representación interna es incapaz de satisfacer la función explicativa que le ha sido asignada por la propia TRM.

## El dilema de la representación interna

Admítase, tal como lo hacen los partidarios de la TRM, que el poder de formar y manipular representaciones internas es lo que, fundamentalmente, explica la posesión de capacidades representacionales. ¿De qué manera tales representaciones cumplen sus diversas funciones en la actividad cognitiva? En el marco de la TRM, el valor explicativo de la noción de representación interna no se reduce a la mera inteligibilidad que resulta de la descripción de los procesos cognitivos de un organismo en términos de manipulación de representaciones internas, sino que se fundamenta en la capacidad que poseerían estas representaciones, en cuanto entidades realmente existentes, de influir causalmente en dichos procesos.

Si bien la noción de representación interna puede ser vista como un componente de un modelo explicativo de la cognición inspirado en el funcionamiento de representaciones públicas, para los partidarios de la TRM, una representación interna no es un mero instrumento explicativo sin implicaciones ontológicas, sino una entidad situada en el orden de las relaciones causales que es capaz de participar en ellas, como lo han señalado Ramsey (2007), Fodor y Phyllyshyn (2015), Neander (2017) entre otros.

Dada esta concepción de las representaciones internas, cabe preguntarse: ¿en virtud de qué propiedades una representación interna ejerce su influencia causal en la actividad cognitiva? Dos respuestas a esta pregunta, y únicamente dos, parecen posibles: o bien una representación interna tiene poder causal en virtud de sus propiedades representacionales, o bien es de sus propiedades no representacionales que deriva dicho poder.<sup>1</sup> Ambas opciones conducen a serias dificultades para la TRM. A continuación, se explicará en qué consisten estas dificultades.

### *Primera parte del dilema*

Supóngase que R es una representación de A formada en el sistema nervioso de un organismo O y que la influencia causal de R en la actividad cognitiva de O es posible gracias a las propiedades representacionales de R. Así, de acuerdo con este planteamiento, es fundamentalmente del he-



cho, como tal, de que R represente A que deriva el poder causal de R en la actividad cognitiva de O.

Ahora bien, ¿cómo de un hecho semántico, como tal, puede derivar semejante poder causal? En otras palabras, ¿cómo un hecho semántico puede relacionarse causalmente, en cuanto hecho semántico, con otro tipo de hechos?

Una respuesta simple es: a través de un proceso de interpretación y comprensión. Para poder admitir que es gracias a sus propiedades representacionales, y no a propiedades de otro tipo, que R influye causalmente en la actividad cognitiva de O, parece necesario asumir que el sistema nervioso de O, de algún modo, ‘comprende’ R y que esta comprensión de R determina, al menos en parte, el curso de los procesos cognitivos efectuados en O. Puesto que R representa A, comprender R no es otra cosa que representarse A gracias a R.

Pero ¿cómo el sistema nervioso de O puede hacer tal cosa? Respecto del problema de la comprensión de representaciones lingüísticas, los partidarios de la TRM suelen tener una respuesta más bien clara: comprender un enunciado E que pertenece a una lengua L es formar una representación interna cuyo contenido es el mismo que el contenido convencionalmente asociado a E en L y asociar dicha representación a E. Sin embargo, ¿qué decir de la comprensión de una representación interna? Cuando el problema es explicar la comprensión de representaciones internas, es obvio que no es posible apelar a la formación de nuevas representaciones internas, pues no se haría más que postergar la cuestión.

El punto discutido está relacionado con la famosa advertencia de Dummett (1993) según la cual toda teoría del significado (o de la representación) que pretenda ser satisfactoria debe estar acompañada de una teoría de la comprensión. Asumir que R es una representación de A, independientemente de cuál sea su localización espacio-temporal y su composición material, es asumir que R está destinada a ser comprendida, es decir, que, en cuanto representación de A, R debe poder permitir a algún interprete representarse A. Así pues, si se postula la existencia de representaciones internas, debe también postularse la existencia de procesos internos de comprensión, pues de lo contrario sería incongruente asumir que las entidades cuya existencia está siendo postulada son auténticas representaciones. Pero, dado que dichos procesos internos de comprensión no pueden consistir en procesos de producción de nuevas representaciones internas, ¿cómo deben ser entendidos?

Comprender representaciones es algo que ciertos organismos pueden hacer: por ejemplo, los seres humanos pueden comprender los



enunciados de la lengua que han aprendido en la infancia. Sin embargo, cuando los partidarios de la TRM hablan de representaciones internas, pretenden hacer referencia a representaciones a las que ningún organismo tiene propiamente acceso, pues se trata, o al menos eso es lo que se supone, de estructuras neuronales, es decir, de estructuras literalmente situadas en el interior de los organismos y no en el espacio interno (metafóricamente hablando) de sus mentes. Y aunque un organismo puede tener acceso sensitivo a algunos de sus procesos internos como, por ejemplo, los procesos digestivos, lo que acontece en el sistema nervioso no forma parte del conjunto de procesos internos a los que un organismo puede sensitivamente acceder. Así pues, si hay comprensión de representaciones internas, tal comprensión debe ser efectuada por un componente interno del organismo y no por el organismo como tal. Pero ¿cómo un simple componente de un organismo puede poseer la capacidad de comprender representaciones?

Una objeción a estas ideas sería la de señalar que hablar de comprensión de representaciones tiene sentido únicamente si se hace referencia a capacidades que un organismo, como tal, puede poseer y no a capacidades de meros componentes de un organismo. Suponer que una representación interna debe ser comprendida sería, por consiguiente, un error. No obstante, si hablar de comprensión es inadecuado en el caso de las representaciones internas, es posible entonces concluir que el uso que los partidarios de la TRM hacen de la noción de representación interna es, más bien, un uso metafórico que permitiría, tal vez, formarse una imagen simple e inteligible de la actividad cognitiva, pero no explicar dicha actividad bajo los criterios requeridos por la propia TRM. La inteligibilidad de esta imagen derivaría del hecho de que utiliza como modelo la comunicación lingüística, cuya realización implica efectivamente el uso de representaciones, pero también de capacidades de interpretación y comprensión. Para contribuir al esclarecimiento de la actividad cognitiva, en cuanto noción de representación en sentido estricto, la noción de representación interna tendría que estar, pues, acompañada de una noción plausible de comprensión interna. Esta última noción, sin embargo, no es más fácil de elucidar que la primera, pues ninguna de las dos es inteligible independientemente de la otra. La postulación de representaciones internas no parece, por consiguiente, ser útil para la comprensión de la actividad cognitiva si se asume que es de las propiedades representacionales de estas que deriva su poder causal.

Sin embargo, a este planteamiento se puede tal vez objetar que una estructura interna puede tener cierto tipo de funcionamiento que puede







legítimamente ser caracterizado como un funcionamiento representacional, aun si dicho funcionamiento no presupone la realización de procesos internos de comprensión como los mencionados. Como señala Ramsey (2007), sería legítimo asumir que ciertas estructuras en el sistema nervioso de un organismo son representaciones si se logra de algún modo mostrar que estas estructuras funcionan como representaciones y, por consiguiente, si se logra establecer que caracterizarlas de este modo es crucial para la comprensión de los procesos en los que participan. Pero ¿qué función debe llevar a cabo una estructura interna para que se pueda legítimamente asumir que dicha estructura funciona como una representación? En sentido estricto, una estructura, evento u objeto R funciona como una representación de A, no solamente si representa A, sino también si permite a un sistema u organismo representarse A. Pero, para que un sistema u organismo pueda representarse A gracias a R, parece ser indispensable que el sistema u organismo en cuestión sea capaz de interpretar y comprender R. Así pues, si este razonamiento es correcto, toda postulación de estructuras internas que funcionan como representaciones nos compromete con la existencia de procesos internos de interpretación y comprensión de representaciones.

En concordancia con estas ideas, Godfrey-Smith (2006) reconoce que los partidarios de la TRM no pueden conformarse con la mera postulación de representaciones internas, sino que deben también admitir la necesidad de postular ‘lectores’ o ‘intérpretes’ internos de dichas representaciones. Para Godfrey-Smith, la actividad de estos mecanismos ‘lectores’ debe poder ser reconocida como una auténtica actividad de interpretación, sin ser empero tan sofisticada que pueda poner en riesgo el valor explicativo de la postulación de representaciones internas (presuponiendo, de algún modo, la posesión previa de capacidades representacionales).

Pero ¿qué tipo de actividad puede satisfacer semejantes criterios? Este problema constituye una preocupación central en la propuesta de Garrett Millikan (1984, 2000, 2017), quien fundamenta su concepción de la representación interna no solamente en la especificación de la relación que debe existir entre R y A para que R represente A, sino también en el tipo de procesos a los que R debe ser sometida y que permitirían al organismo en el que R ha sido formada —o a un mecanismo ‘lector’ interno— identificar aquello que R representa.

La propuesta de Millikan es compleja y será analizada con más detalle en la subsección 3.1, donde se intentará mostrar su inadecuación. Desde la perspectiva de este artículo, es un error asumir que es posible comprender el valor explicativo de la noción de representación interna a

partir de la noción de interpretación, pues ninguna de estas dos nociones es más fácil de integrar en el orden de la explicación causal que la otra.

El hecho de pasar de la noción de representación interna a la noción de interpretación no constituye un avance, sino que más bien muestra que se ha llegado a un *impasse*. Así, la conclusión que se puede extraer de esta primera parte del dilema (y que será reforzada más adelante) es que la noción de representación interna parece incapaz de cumplir la función explicativa que le ha asignado la TRM.

### *Segunda parte del dilema*

Pero tal vez sea posible eludir estos problemas si se admite que el poder causal de una representación interna deriva, no de sus propiedades representacionales, sino de sus propiedades no representacionales. Tal es la postura explícitamente asumida por Fodor (1995, 2008), pero implícitamente asumida por muchos autores que, como Mercier y Sperber (2017), sostienen que una representación interna es un objeto concreto que posee, como todo objeto concreto, poderes causales.<sup>2</sup> Igualmente, Neander (2017) adopta esta postura cuando afirma que “si naturalizamos las representaciones mentales en términos de ciertos fenómenos no intencionales, la fuerza explicativa de las representaciones postuladas será en el caso más básico la fuerza explicativa de estos fenómenos no intencionales” (p. 85).<sup>3</sup>

Supóngase, pues, que R es una representación de A formada en el sistema nervioso de un organismo O y que es gracias a sus propiedades no representacionales que R influye causalmente en la actividad cognitiva de O. En este caso, no parece necesario postular la existencia de procesos internos de interpretación y comprensión de R, pues no es del hecho, como tal, de que R represente A que deriva el poder causal de R, sino de las propiedades que R posee en cuanto mera estructura neuronal. En otras palabras, la cuestión de cómo un hecho semántico se relaciona causalmente con otro tipo de hechos no se plantea aquí, pues los hechos causalmente relacionados que constituyen la actividad cognitiva de O serían, todos, hechos puramente materiales. Sin embargo, dado tal planteamiento, es posible preguntar: ¿en qué medida se puede sostener que las propiedades no representacionales de R ejercen su influencia causal en cuanto propiedades no representacionales de una representación y no en cuanto propiedades de una simple estructura neuronal? Dicho de otro modo, ¿por qué se tendría que suponer que el hecho de que las propiedades no representacionales de R sean propiedades de una representación



tiene alguna relevancia para la influencia causal que estas ejercen sobre la actividad cognitiva de O?

Como fue mencionado, cuando los partidarios de la TRM hablan de representaciones internas, suelen asumir, como Fodor y Pylyshyn (2015), que se trata de objetos o eventos concretos, individuables del mismo modo que pueden serlo “marcas de tiza sobre una pizarra, marcas de tinta sobre el papel, oraciones proferidas, eventos neuronales, etcétera” (p. 7).<sup>4</sup> Otra manera común de asumir esta idea se manifiesta en la distinción entre representaciones mentales (en el cerebro o en la mente) y representaciones públicas (enunciados o imágenes), pues se suele suponer que ambos grupos de representaciones están formados de objetos discretos que difieren básicamente en su localización y composición material. Admítase, pues, que R es un objeto discreto, esto es, en este caso particular, una estructura neuronal N (o, si se prefiere, la activación de N), individual y localizable como un componente (o evento) más o menos bien delimitado del sistema nervioso de O. ¿Qué relevancia tiene, para la explicación causal de la actividad cognitiva de O, caracterizar ciertas propiedades de N como propiedades no representacionales de una representación? La respuesta es: ninguna.

El fragmento de masa cerebral identificado como N puede ciertamente poseer propiedades causalmente relevantes para la actividad cognitiva de O. Sin embargo, caracterizar estas propiedades como propiedades no representacionales de una representación no permite comprender su función en la actividad cognitiva de O, pues el hecho de que N sea una representación (es decir, el hecho de que N sea R) no tiene valor explicativo alguno en dicha actividad, dado que se ha rechazado que las propiedades representacionales de N tengan influencia causal en los procesos cognitivos de O. Esta incapacidad explicativa es una simple consecuencia del hecho de que, por un lado, se ha delimitado la estructura neuronal N asumiendo que se trata de una representación, pero, por el otro, se ha asumido que las propiedades representacionales de N (es decir, las propiedades de N en cuanto representación) no son causalmente relevantes. La estructura neuronal N no tiene, por consiguiente, significación funcional auténtica en la actividad cognitiva de O, ni tampoco sus propiedades, al menos en la medida en que son caracterizadas como propiedades no representacionales de una representación. De forma general, admitir que una representación es un objeto (o evento) concreto con poderes causales no nos obliga a asumir que todos los efectos que produce este objeto (o evento) tienen necesariamente que ver con el hecho de que se trata de una representación. Así pues, aun si N es una representación, no



se puede concluir, sin razones adicionales, que los efectos que N produce tienen que ver con ello.

Veamos el asunto en detalle. El sistema nervioso de O, como todo objeto espacio-temporal, está sometido a diversas fuerzas que pueden afectarlo nomológicamente de distintas maneras. Pero también, además de ser un objeto físico sometido, como cualquier objeto físico, a influencias causales, el sistema nervioso de O cuenta con canales a través de los cuales puede captar flujos causales específicos que afectan su estructura interna de manera más precisa y controlada. Obsérvese, sin embargo, que estos flujos causales no son, por sí mismos, flujos de información. Para que los flujos causales internos puedan ser transformados en auténticos flujos de información, es necesario que sean explotados adecuadamente por el sistema nervioso de O. Para aclarar esta idea, considérese el ejemplo de los anillos de crecimiento de un árbol.

Como es bien sabido, los anillos de crecimiento que forman el tronco de los árboles se correlacionan con la edad de los árboles, de manera que un sistema u organismo con las capacidades requeridas podría extraer de estos anillos información acerca de la edad de un árbol. Esto no significa, sin embargo, que los anillos de crecimiento deban ser vistos, por sí mismos, como vehículos transmisores de información, pues los anillos de crecimiento son simples estructuras físicas que resultan de una determinada serie de acontecimientos causales. Otro ejemplo es el de una huella de animal impresa en el barro: por sí misma, una huella no es un vehículo transmisor de información, sino una simple estructura física producto igualmente de una serie de acontecimientos causales. Para que una estructura de este tipo pueda transformarse en un auténtico vehículo transmisor de información, debe haber un organismo o sistema capaz de explotar las conexiones nomológicas que existen entre la estructura en cuestión y los acontecimientos del entorno (pasados, presentes o futuros).

Desde luego, tal forma de explotación no necesariamente implica comprensión de la información, pues se trata simplemente de un proceso a través del cual la información es utilizada con el fin de efectuar una determinada función del organismo o sistema explotador. En resumen, no hay flujo real de información, ni vehículos transmisores, sin procesos de explotación de relaciones nomológicas; y no hay procesos de explotación de relaciones nomológicas sin utilización de estas relaciones para satisfacer determinadas funciones.<sup>5</sup>

Así, los flujos causales en el sistema nervioso de O, por sí mismos, no son más que acontecimientos causalmente conectados y nomológicamente determinados. Suponer que estos flujos causales internos son



flujos de información es asumir que el sistema nervioso de O, adicionalmente, está estructurado de forma tal que es capaz de explotar las relaciones nomológicas que existen entre sus flujos causales internos y diversos acontecimientos externos con el fin de satisfacer determinadas funciones. Dichas funciones pueden ser cognitivas, como la percepción visual o auditiva, pero también de otro tipo, como la regulación de la temperatura corporal o del ritmo cardíaco.

Ahora bien, volvamos a R, esto es, una representación de A formada en el sistema nervioso de O. Como fue señalado, en cuanto objeto (o evento) concreto, R es una estructura neuronal N (o una activación de N). ¿Es relevante, para la explicación de la actividad cognitiva de O, caracterizar las propiedades de N como propiedades no representacionales de una representación? Si se ve la actividad del sistema nervioso de O como un simple flujo causal, no es posible caracterizar N como una representación, ni *a fortiori* sus propiedades como propiedades no representacionales de una representación, a menos que se esté dispuesto a aceptar una especie de pansemantismo, es decir, la idea de que el puro orden nomológico natural basta para producir representaciones. Pero ¿qué pasa si se ve la actividad del sistema nervioso de O, no como un simple flujo causal interno, sino como un auténtico flujo de información? ¿Es relevante, en este caso, caracterizar las propiedades de N como propiedades no representacionales de una representación?

Para empezar, ¿qué función precisa se puede suponer que N realiza en el sistema nervioso de O visto como un sistema estructurado? Se ha asumido que N es una representación de A. La pregunta es: ¿puede esta caracterización permitir comprender la función de N en la actividad estructurada del sistema nervioso de O? Es razonable suponer que las propiedades de N cumplen alguna función, tal vez crucial, en esta actividad. Sin embargo, como fue señalado, no se puede simplemente asumir que la importancia de estas propiedades en la realización de dicha función —cualquiera que sea— tenga que ver con el hecho de que se trata, como se ha admitido, de propiedades no representacionales de una representación. Para ello, se tendría primero que mostrar que tiene relevancia explicativa para el funcionamiento del sistema nervioso de O caracterizar ciertas estructuras neuronales como representaciones, aun si el poder causal de estas estructuras deriva exclusivamente de sus propiedades no representacionales. El mero hecho de admitir que el sistema nervioso de O es un sistema estructurado no permite, por sí mismo, establecer tal relevancia. Así, la cuestión que se plantea es: ¿cómo mostrar la relevancia



explicativa de la caracterización de ciertas estructuras o eventos neuronales como representaciones?

Una opción es sostener que caracterizar ciertas estructuras o eventos neuronales como representaciones es esencial, o tal vez simplemente útil, para explicar el proceso mismo de estructuración del sistema nervioso. Tal opción no es recomendable, pues conduce de nuevo al pansemantismo, en la medida en que asume la idea de que, en el flujo causal interno previo a la estructuración del sistema nervioso, existen representaciones que, en virtud de sus propiedades no representacionales, contribuyen a dicha estructuración.

Una segunda opción es asumir que el proceso de estructuración del sistema nervioso, o al menos de algunas de sus partes, es precisamente un proceso en el cual son formadas representaciones internas. El problema con esta opción es que no aclara realmente la función de las representaciones formadas, pues se podría suponer que la simple estructuración del sistema nervioso basta para explicar la actividad cognitiva de un organismo, independientemente de si hay representaciones que resultan, de algún modo, del proceso mismo de estructuración. Dicho de otra manera, la formación de representaciones, en este caso, podría no ser más que un epifenómeno.

Una tercera opción, tal vez la más convincente, es negar que la simple estructuración del sistema nervioso sea suficiente para explicar la actividad cognitiva de los organismos y admitir que no es posible, o al menos que es difícil, explicar adecuadamente dicha actividad sin caracterizar ciertos componentes o eventos del sistema estructurado como representaciones. Nótese que esta afirmación no es un planteamiento obvio, sino un compromiso sustantivo. Admitamos, no obstante, su verdad. La primera pregunta que debe plantearse es: ¿qué componentes o eventos del sistema estructurado deben (o pueden) ser caracterizados como representaciones? La respuesta más común ha sido la de caracterizar los vehículos transmisores de información como representaciones. Esta respuesta no es, sin embargo, satisfactoria. Para empezar, hemos asumido, en concordancia con la TRM, que una representación interna es un objeto o evento más o menos bien delimitado: una estructura neuronal N (o una activación de N). Pero un vehículo transmisor de información no necesariamente debe ser visto de esta manera. La existencia de un flujo de información, como fue señalado, supone ante todo la explotación de relaciones nomológicas con el fin de satisfacer determinadas funciones. Pero la explotación de relaciones nomológicas no supone forzosamente la delimitación de objetos o eventos que puedan servir de vehículos transmisores.



res de información. La explotación de relaciones nomológicas es efectuada cuando, ante un estímulo externo relevante, una respuesta que contribuye a la satisfacción de una función del sistema u organismo explotador es generada. La estructuración que posibilita tal respuesta puede ser vista como la consolidación de un vínculo causal interno entre la estimulación y la respuesta, sin que sea necesario suponer que, en la cadena de acontecimientos que constituyen dicho vínculo, un objeto o evento en particular deba ser delimitado. Desde esta perspectiva, hablar de vehículos transmisores de información sería simplemente una manera de señalar que el sistema u organismo reacciona a un estímulo específico y no implicaría un compromiso con la delimitación de objetos o eventos internos.

Ignoremos, no obstante, esta objeción y supongamos que un vehículo transmisor de información puede ser, al menos en algunos casos, una estructura neuronal delimitada (o su activación). Supongamos también que N es un vehículo transmisor de información. ¿Puede ahora ser relevante, para explicar la actividad cognitiva de O, caracterizar las propiedades de N como propiedades no representacionales de una representación? Para comprender el funcionamiento de N en cuanto vehículo transmisor de información, basta con captar, por un lado, que N es una estructura neuronal cuya activación depende nomológicamente de la instanciación de un acontecimiento A (interno o externo) y, por el otro, que la activación de N contribuye causalmente a la generación de una respuesta B que satisface, dada la instanciación de A, una función del sistema nervioso de O. Caracterizar N como una representación, y sus propiedades como propiedades no representacionales de una representación, no añade nada relevante a este funcionamiento. Por otro lado, asumir que el simple hecho de que N conecte causalmente A con B justifica la caracterización de N como una representación no es una buena opción, pues este tipo de funcionamiento es también relevante para la explicación de funciones no cognitivas, como la secreción de hormonas, por ejemplo. Admitir que N es una representación implicaría asumir la idea de que prácticamente toda actividad interna de un organismo, cognitiva o no cognitiva, es una forma de actividad representacional.

Pero ¿qué pasa si suponemos que N contribuye específicamente a la ejecución de tareas cognitivas y no a la satisfacción de funciones de otro tipo? ¿No es ahora relevante caracterizar N como una representación? Según Ramsey (2007), para captar el sentido de las explicaciones computacionales de la cognición, es necesario asumir que ciertos componentes internos que posibilitan la realización de tareas cognitivas están siendo empleados como representaciones. Considérese un ejemplo de



Ramsey: la realización de una operación de multiplicación. En un enfoque computacional se suele dividir la tarea a realizar en subtareas más simples, como, en este ejemplo en particular, la adición sucesiva de un número a sí mismo.

Estas subtareas son ejecutadas por módulos que procesan determinados *inputs* y producen determinados *outputs*. De acuerdo con Ramsey, es relevante, para la explicación computacional, asumir que estos *inputs* y *outputs* están siendo empleados, por el sistema mismo, como representaciones de sumandos y sumas, respectivamente, pues de otra manera no sería posible suponer que la operación efectuada es una multiplicación, ni comprender el éxito en su realización.

Sin embargo, nótese que, como lo señala el mismo Ramsey, la explicación computacional no cuestiona en ningún momento el hecho de que la tarea ejecutada por el sistema sea realmente una operación de multiplicación, sino que simplemente lo asume. Por lo tanto, la explicación computacional debe igualmente asumir que los procesos efectuados en la realización de esta tarea están regulados, de algún modo, por las normas que definen la multiplicación. De ahí que pueda parecer natural caracterizar los *inputs* y *outputs* del módulo interno como representaciones de sumandos y sumas, pues esta caracterización no es más que una manera de reconocer que debe haber una influencia de las normas de la multiplicación en los procesos del sistema (normas que podrían prescribir, en efecto, la adición sucesiva del multiplicando a sí mismo el número de veces indicado por el multiplicador). No por ello, sin embargo, tal influencia ha sido explicada. En otras palabras, la caracterización de los *inputs* y *outputs* como representaciones de sumandos y sumas no tiene una auténtica función explicativa, sino que se trata de un efecto de la inevitable proyección de las normas de la multiplicación en el comportamiento del sistema, una vez que se ha asumido que la operación efectuada es una multiplicación. Desde luego, un teórico podría intentar explicar cómo las normas de la multiplicación influyen efectivamente en los procesos del sistema nervioso de un organismo cuando una operación de este tipo es realizada. Pero abordar este problema no es el propósito de la explicación computacional, ni algo que una tal explicación pueda resolver.

Así, parece simplemente haber una incongruencia en el hecho de sostener que la postulación de representaciones internas tiene relevancia explicativa en virtud de la influencia causal que estas ejercen sobre la actividad cognitiva de los organismos y al mismo tiempo admitir que las propiedades representacionales de una representación interna no son causalmente relevantes. Tal planteamiento equivale a afirmar que las representaciones





internas son explicativamente relevantes, siempre y cuando se ignore el hecho de que se trata de representaciones. Dada esta incongruencia y dadas las dificultades que se han discutido en esta subsección, la conclusión que debe extraerse de la segunda parte del dilema es la misma que la de la primera parte: la noción de representación interna es incapaz de satisfacer la función explicativa que le ha asignado la propia TRM.

Recapitulemos los resultados del dilema. Si se admite que es gracias a sus propiedades representacionales que R (una representación de A formada en el sistema nervioso de un organismo O) ejerce influencia causal en la actividad cognitiva de O, debe postularse la existencia de procesos internos de interpretación y comprensión de R. Tal situación conduce a un *impasse*, pues las nociones de interpretación y comprensión no son más fáciles de integrar en el orden de la explicación causal que la noción misma de representación interna.

Pero admitir que es, más bien, en virtud de sus propiedades no representacionales que R ejerce dicha influencia causal no es una mejor opción. En efecto, si se acepta tal cosa, no es posible sostener que las propiedades no representacionales de R, en cuanto propiedades no representacionales de una representación, tienen una verdadera significación funcional en la actividad cognitiva de O.

En otras palabras, aun si las propiedades no representacionales de R tuvieran alguna función en esta actividad, no se podría sostener que dicha función tiene algo que ver con el hecho de que se trata de propiedades no representacionales de una representación. Así, las dos partes del dilema nos conducen a la misma conclusión: la noción de representación interna es incapaz de cumplir la función explicativa que le ha asignado la TRM.

## Dos ejemplos

Para reforzar la conclusión a la que se ha llegado en la sección anterior, veamos dos ejemplos del uso que se ha dado a la noción de representación interna por dos excelentes partidarias de la TRM: la filósofa Ruth Garrett Millikan y la psicóloga Susan Carey.

### *Primer ejemplo*

En sus primeros trabajos, Millikan (1984) establece una importante distinción entre íconos intencionales y representaciones. Un ícono intencional T es una estructura interna que colabora con un mecanismo biológico (no necesariamente cognitivo) con el fin de que dicho mecanismo



pueda realizar su función propia. La contribución de T a la realización de la función propia del mecanismo con el que colabora es posible, según Millikan, gracias al hecho de que T se corresponde con un estado de cosas cuya obtención es una condición normal para la adecuada satisfacción de dicha función.

La relación de correspondencia de la que habla Millikan es una relación de isomorfismo: la configuración de elementos constitutivos del estado de cosas externo del cual T es un ícono intencional se corresponde con la configuración de elementos constitutivos de la estructura de T y determinadas transformaciones de la configuración de este mismo estado de cosas se corresponden con determinadas transformaciones de la estructura de T.

Para Millikan, un ícono intencional, a pesar de ser una estructura interna con auténticas propiedades intencionales, no es aún, por dos razones fundamentales, una representación. La primera es que un ícono intencional contribuye a la realización de funciones biológicas de todo tipo, no necesariamente cognitivas. La segunda es que, a diferencia de un mero ícono intencional, una representación interna no debe simplemente corresponderse con un estado de cosas (interno o externo) con el fin de satisfacer una función, sino que debe también ser procesada de manera que permita al organismo representarse aquello que ella misma representa.

En efecto, de acuerdo con Millikan, una representación interna, además de ser un ícono intencional, debe permitir al organismo en cuyo sistema nervioso ha sido formada ‘identificar’ aquello que es representado por ella. Así, desde la perspectiva de Millikan, la cuestión cognitiva fundamental es: ¿cómo un ícono intencional puede transformarse en una auténtica representación?

En otras palabras: ¿cómo un organismo puede identificar el estado de cosas externo con el que se corresponde un ícono intencional que ha sido formado en su propio sistema nervioso?

Nótese que este problema no es otro que el problema ya mencionado (subsección 2.1) de cómo una representación de A formada en el sistema nervioso de un organismo O puede permitir a O, o a un componente interno de O, representarse A. ¿Cuál es la respuesta de Millikan?

Supongamos que T es un ícono intencional formado en el sistema nervioso de O y que un componente del estado de cosas externo con el que T se corresponde es un A. En la estructura de T debe haber, por consiguiente, un elemento que se corresponde con la presencia de un A en dicho estado de cosas. Supongamos que R es ese elemento. ¿Cómo puede



R permitir a O identificar el componente A del estado de cosas externo con el que T se corresponde?

Para Millikan, identificar es reidentificar. Más precisamente, ser capaz de identificar aquello con lo que R se corresponde no supone ser capaz de representarse R, el A con el que R se corresponde y la relación de correspondencia entre ambos, sino ser capaz de captar que las distintas representaciones de A formadas en el sistema nervioso representan lo mismo, es decir, ser capaz de captar cuándo un A es representado de nuevo (cuándo distintos componentes de distintos íconos intencionales se corresponden con lo mismo). Así, O es capaz de identificar aquello que R representa cuando es capaz de captar que las distintas representaciones de A formadas en su sistema nervioso, entre ellas R, representan lo mismo.

Tal captación no presupone la posesión de capacidades metarepresentacionales, sino que se reduce al uso normal —definido en términos evolucionistas— que los mecanismos cognitivos de O hacen de sus propias representaciones. Por ejemplo, en el caso de los mecanismos encargados de la realización de inferencias, las distintas representaciones de A que figuran en los contenidos que constituyen las premisas de un razonamiento deben poder ser utilizadas, cuando es requerido, como el término común que permite derivar la conclusión. Igualmente, en el caso de la acción, una representación de A que figura en el contenido de una intención y una representación de A que figura en el contenido de una experiencia perceptiva deben poder ser utilizadas como representaciones de lo mismo.

Un organismo cuyos mecanismos cognitivos emplean de esta forma sus distintas representaciones de A es, para Millikan (2000), un organismo capaz de identificar aquello que representan estas representaciones.<sup>6</sup>

Los planteamientos de Millikan son una respuesta al problema de la existencia de procesos internos de interpretación de representaciones. Los íconos intencionales que contribuyen a la realización de tareas cognitivas serían ‘consumidos’ de una manera tal que, normalmente, y cuando es requerido, sus distintos componentes que se corresponden con lo mismo serían procesados —en la percepción, el razonamiento o la acción— como componentes que se corresponden con lo mismo. Tales mecanismos de ‘consumo’ —o ‘lectores’— de íconos intencionales habrían sido forjados a lo largo del desarrollo evolutivo de las especies con capacidades cognitivas.

De acuerdo con Millikan (2000), la cuestión de qué tipo de indicador determina cuándo dos componentes de distintos íconos intencionales deben ser procesados como componentes que se corresponden con los mismos es una cuestión empírica, para la cual existen respuestas



plausibles como la duplicación de estructuras neuronales, la existencia de alguna forma de marcador o la activación sincronizada.

El procesamiento de componentes de distintos íconos intencionales como componentes que se corresponden con lo mismo se reflejaría en el comportamiento del organismo como tal, es decir, en las acciones, juicios e inferencias que el organismo efectúa y cuyo éxito depende, precisamente, de la identificación de individuos y propiedades como el mismo individuo o la misma propiedad anteriormente representados.

De este modo, los íconos intencionales serían transformados en auténticas representaciones internas que permitirían al organismo en cuyo sistema nervioso han sido formadas representarse aquello que ellas mismas, presuntamente, representan.

Pero ¿es posible realmente admitir que este procesamiento de componentes de distintos íconos intencionales como componentes que se corresponden con lo mismo es una forma interna de interpretación que mostraría la relevancia explicativa de la caracterización de estos componentes como representaciones?

Consideremos dos estructuras neuronales  $N$  y  $N'$ . Supongamos que  $N$  y  $N'$  son componentes de distintos íconos intencionales. Para ser procesadas como componentes que se corresponden con lo mismo,  $N$  y  $N'$  deben poseer las características físicas que los presuntos mecanismos 'lectores' de íconos intencionales utilizan como indicador de que se trata de componentes que se corresponden con lo mismo: ser del mismo tipo, tener cierto marcador o ser activadas de manera sincronizada.

Si  $N$  y  $N'$  poseen estas características físicas, serán procesadas como componentes que se corresponden con lo mismo. Sin embargo, dado que las propiedades relevantes para el funcionamiento de los mecanismos 'lectores' no son más que propiedades físicas, tal procesamiento puede simplemente ser visto como una secuencia de acontecimientos causalmente conectados que contribuye —puede suponerse— a la satisfacción de una determinada función.

Admitir que este tipo de procesamiento causal es una forma de interpretación interna de representaciones no añade nada esclarecedor, sino todo lo contrario, pues obliga a explicar por qué hablar de interpretación sería pertinente aquí, dado que los mecanismos 'lectores' no hacen más que reaccionar a ciertas propiedades físicas de  $N$  y  $N'$ . La noción de interpretación no es más fácil de integrar en el orden de la explicación causal que la noción de representación interna y esta dificultad no es más que una muestra de ello.



Una objeción posible a estos razonamientos sería la de señalar que los íconos intencionales de los que se habla son empleados en tareas cognitivas cuya realización implica la capacidad de identificar cuándo un individuo o una propiedad están siendo representados de nuevo.

Por consiguiente, una forma de interpretación de íconos intencionales que permita identificar cuándo sus componentes se corresponden con lo mismo sería requerida. Sin embargo, identificar cuándo un individuo o una propiedad están siendo representados de nuevo es, ante todo, algo que un organismo, como tal, es capaz de hacer. En nada ayuda, para explicar esta capacidad, suponer la existencia de procesos internos de interpretación, si con ello no se hace más que proyectar en la actividad del sistema nervioso la posesión de la capacidad misma que se pretendía explicar.

284



### *Segundo ejemplo*

Basándose en la teoría del factor dual, formulada por Block (1986), según la cual una representación interna es una estructura neuronal cuyo contenido está a la vez determinado por su conexión causal con propiedades u objetos externos (que, en virtud de tal conexión, constituyen su extensión) y por la manera en que es empleada por mecanismos cognitivos específicos, Carey (2009) distingue varios tipos de representaciones internas: sensibles, perceptivas, de cognición básica (*core cognition representations*) y conceptuales.

Todas estas, según Carey, cumplen con los requisitos de la teoría del factor dual, pues se trata de estructuras neuronales que, además de estar causalmente conectadas con propiedades u objetos externos, son procesadas por determinados mecanismos cognitivos.

Para Carey, una representación sensible es el *output* de un órgano sensorial; por ejemplo, en el caso del ojo, una imagen formada en la retina. Las representaciones perceptivas, en cambio, representan objetos concretos con propiedades estables, a pesar de la enorme variabilidad en la apariencia sensible de dichos objetos según las condiciones del entorno y la relación espacial entre el organismo que percibe y los objetos percibidos.

El tercer tipo de representación, la de cognición básica, es una forma de representación conceptual empleada en tareas especializadas que implican, según Carey, la utilización de información que no puede ser obtenida de simples representaciones sensibles o perceptivas.

Para ilustrar su funcionamiento, Carey (2009) presenta el ejemplo del azulejo índigo, un ave migratoria capaz de identificar, en las noches con cielo estrellado, el norte terrestre. Según los estudios experimentales

en los que se basa Carey, el azulejo índigo posee mecanismos innatos que le permiten detectar, cuando aún es un polluelo, el centro de rotación del cielo estrellado. Dado que la aparente rotación del cielo estrellado es un efecto de la rotación de la Tierra sobre su propio eje y que el eje de rotación de la Tierra atraviesa el planeta de norte a sur, detectar el centro de rotación del cielo estrellado permite identificar la dirección norte del planeta (el azulejo índigo habita el hemisferio norte de la Tierra).

Para Carey, el azulejo índigo es capaz de representarse perceptivamente el cielo estrellado con su centro de rotación, pero dada la manera en que sus representaciones son empleadas por sus mecanismos de navegación, dichas representaciones no pueden ser consideradas simples representaciones perceptivas. En particular, Carey (2009) afirma que los mecanismos de navegación del azulejo índigo son capaces de ‘inferir’, a partir de representaciones perceptivas del cielo estrellado, la dirección que esta ave debe seguir en su proceso de migración. La información extraída parece superar lo que es propiamente representado por el sistema perceptivo del azulejo índigo.

Así, de acuerdo con Carey, una representación perceptiva que es empleada de esta forma, es decir, en procesos cognitivos especializados que son efectuados por mecanismos innatos como los mecanismos de navegación del azulejo índigo, es una representación de cognición básica.

Finalmente, para Carey (2009), las representaciones de cognición básica, a pesar de ser representaciones conceptuales informativamente superiores a las representaciones perceptivas, deben ser distinguidas de las conceptuales que son empleadas en procesos flexibles de razonamiento y teorización, pues el uso dado a las de cognición básica se limita, como fue señalado, a la realización de tareas especializadas efectuadas por mecanismos innatos.

Para captar la irrelevancia explicativa de la noción de representación interna que Carey emplea, se debe notar, en primer lugar, que prácticamente cualquier estructura neuronal que podamos aislar en el sistema nervioso de un organismo tiene conexiones causales más o menos remotas con propiedades u objetos externos, así como efectos en el interior del sistema.

Tal afirmación no hace más que expresar el hecho de que, en el sistema nervioso de todo organismo, existen flujos causales conectados con el mundo exterior. Desde luego, la actividad del sistema nervioso no es un mero flujo causal, sino una actividad estructurada que explota las relaciones nomológicas que existen entre sus acontecimientos internos y el mundo exterior con el fin de satisfacer sus funciones. Un ejemplo de ello es, precisamente, la actividad que posibilita la navegación del azulejo índigo.





Steven Emlen (1975), en sus estudios experimentales, presenta a diferentes grupos de polluelos de azulejo índigo, criados en un planetario, vistas del cielo estrellado con centros de rotación distintos. Durante el periodo crítico de aprendizaje, como ya fue mencionado, los polluelos registran el centro de rotación observado de forma que, cuando el momento de la migración llega, la dirección de vuelo es seleccionada en función de la ubicación de dicho centro de rotación. El trabajo de Emlen muestra que el sistema nervioso del azulejo índigo es capaz de explotar las conexiones nomológicas que existen entre la estimulación de sus estructuras internas, el movimiento aparente de la bóveda celeste y la geografía terrestre, con el fin de satisfacer una función específica: la orientación en el proceso de migración. ¿En qué momento es necesario, o útil, para explicar esta forma de actividad estructurada, postular la existencia de representaciones internas?

Ciertamente, la observación del centro de rotación del cielo estrellado, durante el periodo crítico de aprendizaje, permite al azulejo índigo calibrar sus mecanismos de navegación. Pero la conexión funcional entre la observación del centro de rotación del cielo estrellado y la selección de la dirección de vuelo puede ser vista como la simple consolidación (gracias a mecanismos innatos) de un vínculo causal interno.

Tal consolidación no requiere la delimitación de estructuras neuronales que puedan ser caracterizadas como representaciones. Y aunque fuera posible delimitar ciertas estructuras (o eventos) en la serie de acontecimientos internos y asignarles la función de vehículos transmisores de información, no habría razón para suponer que tales estructuras (o eventos) son algo más que mediadores causales que, en virtud de sus conexiones nomológicas y sus efectos en el sistema, contribuyen a satisfacer la función de orientación del azulejo índigo.

Si se asumiera, a pesar de todo, que las estructuras en cuestión son auténticas representaciones, no por ello se mostraría, sin embargo, que dichas estructuras tienen relevancia explicativa en cuanto representaciones, pues el hecho de que estas estructuras delimitadas sean representaciones podría ser un efecto sin significación funcional en la actividad del sistema nervioso del azulejo índigo, es decir, un simple epifenómeno.

Para Carey, las representaciones formadas en el sistema nervioso de un organismo, en particular las representaciones de cognición básica, son procesadas por módulos cognitivos que producen, a partir de ellas, nuevas representaciones. En el caso del azulejo índigo, de acuerdo con Carey (2009), el módulo cognitivo innato gracias al cual esta ave logra orientarse en su proceso de migración admite como *inputs* representa-

ciones del cielo estrellado y produce como *outputs* representaciones de la dirección a seguir. Sin embargo, tanto la postulación de módulos cognitivos como la postulación de representaciones internas no son más que un puro efecto de la proyección, en la actividad del sistema nervioso, de las normas que regulan el ejercicio de la capacidad que, en cada ocasión particular, se pretende explicar.

Dado que se parte del hecho de que el sistema nervioso del azulejo índigo es capaz de orientar esta ave en su proceso de migración, debemos suponer que las normas que definen el correcto ejercicio de esta capacidad deben, de algún modo, regular la actividad del sistema. Hablar de representaciones del cielo estrellado procesadas por módulos internos que ‘inferen’ representaciones de la dirección de vuelo a seguir es simplemente una manera de reconocer que debe haber una influencia de dichas normas en la actividad del sistema nervioso del azulejo índigo, pero de ningún modo una explicación de cómo esta influencia es posible.



## Conclusión

Para concluir, precisemos que el hecho de rechazar la postulación de representaciones internas no obliga a rechazar la evidencia que Carey y sus colaboradores han obtenido en favor de la existencia de una forma básica e innata de cognición. La postulación de representaciones internas y módulos cognitivos encargados de procesarlas no es propiamente respaldada por dicha evidencia, sino que es el mero reflejo de la adopción de la TRM cuya tesis central es que la vida cognitiva de un organismo se reduce a la formación, almacenamiento y procesamiento de representaciones internas. El hecho de que el sistema nervioso de un organismo con capacidades cognitivas sea un sistema altamente estructurado y que tal estructuración sea en buena medida innata no conlleva, por sí mismo, la existencia de representaciones internas.

De forma general, el rechazo de la noción de representación interna no implica forzosamente el rechazo de los resultados obtenidos por los científicos de la cognición, aun por aquellos que emplean esta misma noción. La noción de representación interna es incapaz de satisfacer la función explicativa que le ha asignado la TRM, pero eso no significa que esta noción no pueda tener alguna otra función que pueda explicar, en alguna medida, su uso generalizado.

En particular, como fue mencionado, postular la existencia de representaciones internas es una manera (muchas veces implícita) de reco-



nocer que los procesos del sistema nervioso que permiten la realización de tareas cognitivas deben estar sometidos a las normas que definen la correcta ejecución de dichas tareas. Asumir tal cosa puede ser legítimo, aunque debe también admitirse que la postulación de representaciones internas no permite explicar la influencia de dichas normas, sino que es simplemente una manera de reconocerla. Explicar de qué naturaleza es la influencia en cuestión y cómo exactamente se produce no es necesariamente una obligación de los científicos cognitivos. Sin embargo, tampoco debe pensarse que la postulación de representaciones internas permite resolver estas dificultades.

## Notas

288



1. Estrictamente hablando, hay una tercera opción: el poder causal de una representación interna deriva de ambos tipos de propiedades. Esta opción, sin embargo, no es aquí relevante, pues no afecta los argumentos que se presentarán a continuación.
2. Fodor y Pylyshyn (2015) llegan incluso a afirmar que, tal vez, puede haber una mejor solución a este problema, pero que ellos no tienen conocimiento “de ninguna otra” (p. 7). Dicho sea de paso, el hecho de emplear la expresión “propiedades sintácticas”, o “propiedades de la forma de un símbolo”, para referirse a las propiedades no representacionales de una representación no modifica en nada su carácter en cuanto meras propiedades no representacionales de una representación.
3. Mi traducción.
4. Mi traducción.
5. En este sentido, diferimos de Dretske (1981) cuando afirma que “in the beginning there was information” (p. vii). En el inicio no había información, pues no había propiamente señales ni vehículos transmisores, sino meras relaciones nomológicas entre estructuras, eventos u objetos. Desde luego, había estructuras potencialmente transmisoras de información, es decir, estructuras que hubieran podido ser explotadas por sistemas u organismos con las capacidades requeridas (si los hubiera habido).
6. Millikan (2017) llega incluso a sostener que es, aparentemente, una consecuencia “necesaria” de la TRM que “una cosa es identificada cuando los signos de ella son coidentificados” (p. 51). Mi traducción.

## Bibliografía

- BLOCK, Ned  
1986 Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615-678. <http://dx.doi.org/10.1111/j.1475-4975.1987.tb00558.x>
- CAREY, Susan  
2009 *The origin of concepts*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195367638.001.0001>

- CUMMINS, Robert  
2010 *The world in the head*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:osobl/9780199548033.001.0001>
- DRETSKE, Fred  
1981 *Knowledge and the flow of information*. Cambridge, Massachusetts: M.I.T. Press.  
1988 *Explaining behavior*. Cambridge, Massachusetts: M.I.T. Press.
- DUMMETT, Michael  
1993 *The seas of language*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/0198236212.001.0001>
- EMLÉN, Steven  
1975 The stellar-orientation system of a migratory bird. *Scientific American*, 233, 102-111. <http://dx.doi.org/10.1038/scientificamerican0875-102>
- FODOR, Jerry  
1984 Semantics, Wisconsin style. *Synthese*, 59(3), 231-250. <http://dx.doi.org/10.1007/BF00869335>  
1990 *A theory of content and other essays*. Cambridge, Massachusetts: M.I.T. Press.  
1995 Concepts; a potboiler. *Philosophical Issues*, 6, 1-24. <http://dx.doi.org/10.2307/1523025>  
1998 *Concepts. Where cognitive science went wrong*. Oxford: Clarendon Press. <http://dx.doi.org/10.1093/0198236360.001.0001>
- FODOR, Jerry & LEPORE, Ernest  
1991 Why meaning (probably) isn't conceptual role. *Mind and Language*, 6(4), 328-343 <http://dx.doi.org/10.1111/j.1468-0017.1991.tb00260.x>
- FODOR, Jerry & PYLYSHYN, Zenon  
2015 *Minds without meanings*. Cambridge, Massachusetts: M.I.T. Press. <http://dx.doi.org/10.7551/mitpress/9780262027908.001.0001>
- GODFREY-SMITH, Peter  
2006 Mental representation, naturalism, and teleosemantics. En G. Macdonald & D. Papineau (Eds.), *Teleosemantics* (pp. 42-68). Oxford: Oxford University Press.
- HUTTO, Daniel & MYIN, Erik  
2013 *Radicalizing enactivism*. Cambridge, Massachusetts: M.I.T. Press. <http://dx.doi.org/10.7551/mitpress/9780262018548.001.0001>
- JOHNSON-LAIRD, Philip  
2006 *How we reason*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199551330.003.0028>
- MERCIER, Hugo & SPERBER, Daniel  
2017 *The enigma of reason*. Cambridge, Massachusetts: Harvard University Press. <http://dx.doi.org/10.4159/9780674977860>
- MILLIKAN, Ruth Garrett  
1984 *Language, thought and other biological categories*. Cambridge, Massachusetts, MIT Press.  
2000 *On clear and confused ideas*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511613296>  
2017 *Beyond concepts*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/oso/9780198717195.001.0001>



NEANDER, Karen

2017 *A mark of the mental*. Cambridge, Massachusetts: M.I.T. Press. <http://dx.doi.org/10.7551/mitpress/9780262036146.001.0001>

PRINZ, Jesse

2002 *Furnishing the mind*. Cambridge, Massachusetts: M.I.T. Press. <http://dx.doi.org/10.7551/mitpress/3169.001.0001>

RAMSEY, William

2007 *Representation reconsidered*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511597954>

STERELNY, Kim

1990 *The representational theory of mind*. Oxford: Basil Blackwell.

Fecha de recepción de documento: 23 de mayo de 2020

Fecha de revisión de documento: 15 de julio de 2020

Fecha de aprobación de documento: 15 de diciembre de 2020

Fecha de publicación de documento: 15 de julio de 2021

