

Un argumento epifenomenalista contra las teorías de tipo output del contenido mental.

An epiphenomenalist argument against output theories of mental content.



[Licencia Creative Commons Reconocimiento-NoComercial 4.0 Internacional.](#)

DOI: 10.32870/sincronia.axxix.n88.2.25b

Fabián Bernache Maldonado

Universidad de Guadalajara
(MÉXICO)

CE: fabian.bernache@academicos.udg.mx

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

Recibido: 03/02/2025 Revisado: 10/03/2025 Aprobado: 28/04/2025

Resumen.

Según las teorías de tipo output del contenido mental, el contenido semántico de las representaciones mentales queda fijado por los efectos de dichos estados internos, es decir, por su contribución causal a la actividad cognitiva. Desde este punto de vista, el uso antecede al contenido, y no al revés. Sin embargo, si es en virtud de su contribución causal a la actividad cognitiva que las representaciones mentales fijan su contenido semántico, el contenido semántico no puede ser considerado como uno de los factores que explican dicha contribución causal. Las representaciones mentales, como tales, podrían ser un mero subproducto de la actividad cognitiva. El objetivo de este artículo es desarrollar este argumento epifenomenalista y mostrar cómo se aplica a propuestas teóricas recientes.

Palabras clave: Cognición. Contenido. Explicación. Representacionalismo. Representaciones estructurales. Semántica Varitel.

Abstract:

According to output theories of mental content, the semantic content of mental representations is fixed by the effects of these internal states, that is, by their causal contribution to cognitive activity. On this view, use antedates content, and not the other way around. However, if it is by having a causal role in cognitive activity that mental representations get their semantic content fixed, semantic content cannot be taken as one of the factors that explain this causal role. Mental representations,

Cómo citar este artículo (APA):

En párrafo (cita parentética):
(Bernache, 2025, p. __)

En la lista de referencias:
Bernache, F. (2025). Un argumento epifenomenalista contra las teorías de la salida del contenido mental. *Revista Sincronía*. XXIX(88). 19-44
DOI: 10.32870/sincronia.axxix.n88.2.25b

as such, might be a mere by-product of cognitive activity. The aim of this paper is to develop this epiphenomenalist argument and to show how it applies to recent proposals.

Keywords: Cognition. Content. Explanation. Representationalism. Varitel Semantics. S-representations.

Introducción

El objetivo principal de este trabajo es presentar una objeción contra las teorías de output del contenido mental. Esta objeción puede caracterizarse como una especie de problema epifenomenalista. Argumento que la estrategia que adoptan los teóricos del output para explicar cómo se fija el contenido semántico de las representaciones mentales implica que estos estados internos no son más que un subproducto causalmente irrelevante de los procesos cognitivos. En la sección 1 se explican brevemente las teorías input y output del contenido mental. Enfoco mi atención en las diferentes formas en que estas teorías abordan las dos cuestiones centrales planteadas en el marco representacionista: el problema de la intencionalidad subyacente y el problema de cómo se realizan las manipulaciones sobre las representaciones mentales. En la sección 2, muestro cómo la forma particular en que las teorías de output abordan estas cuestiones implica que las representaciones mentales son epifenoménicas. Es decir, formulo lo que llamo el argumento epifenomenalista contra las teorías de output del contenido mental. Finalmente, en la sección 3, muestro cómo el argumento epifenomenalista se aplica a dos propuestas recientes y muy influyentes: La Semántica Varitel de Shea (2018) y el relato de las representaciones S de Gładziejewski y Miłkowski (2017).

Teorías de entrada y salida del contenido mental

Según la Teoría Representacional de la Mente (RTM), los procesos físicos que subyacen a las capacidades cognitivas implican manipulaciones realizadas sobre representaciones mentales. Las representaciones mentales se conciben generalmente como estados físicos que interactúan causalmente con otros estados físicos. Por supuesto, las representaciones mentales no son meros estados físicos. Son estados físicos que se supone que se refieren o tratan de otras cosas,

propiedades, o acontecimientos. En otras palabras, tienen intencionalidad. Sin embargo, en contraste con la intencionalidad de otros tipos de representaciones, la intencionalidad de las representaciones mentales se supone generalmente que es de un tipo muy especial. En efecto, un punto de vista común entre los partidarios de la RTM es que las representaciones mentales tienen, no sólo intencionalidad, sino intencionalidad subyacente u original. Es decir, los partidarios de la RTM generalmente asumen que las propiedades semánticas de otros tipos de representaciones (por ejemplo, las representaciones lingüísticas) derivan de las de las representaciones mentales, pero que las propias representaciones mentales (o al menos algunas de ellas) no derivan sus propiedades semánticas de las propiedades semánticas de representaciones de otros tipos. Por lo tanto, hay dos cuestiones centrales que cualquier persona interesada en desarrollar una versión particular de RTM debe abordar:

1. ¿Cómo adquieren contenido semántico las representaciones mentales (con intencionalidad no derivada)?
2. ¿Cómo se manipulan las representaciones mentales en los procesos subyacentes a las capacidades cognitivas?

Los filósofos han adoptado dos actitudes diferentes respecto a la relación entre estas cuestiones. La primera consiste en considerar que estas cuestiones son independientes entre sí: Podemos dar una respuesta relativamente completa a una de ellas y luego intentar averiguar qué respuesta podría darse a la otra. Por supuesto, las respuestas deben ser compatibles, pero se considera que las preguntas expresan dos problemas independientes. La segunda actitud consiste en considerar que las preguntas tienen cierto grado de dependencia mutua. Para los filósofos que adoptan esta actitud, para explicar cómo las representaciones mentales adquieren contenido semántico, debemos considerar los efectos que esos estados físicos tienen en el sistema en el que se producen. Al tener en cuenta estos efectos, también debemos tener en cuenta la forma en que las representaciones mentales interactúan causalmente con otros estados físicos dentro del sistema, y nuestra comprensión de esas interacciones causales

restringe parcialmente la respuesta que podemos dar a la pregunta de cómo se manipulan las representaciones mentales en la actividad cognitiva.

Estas actitudes están relacionadas con la preferencia por las teorías naturalistas de entrada o de salida del contenido mental. Tanto los teóricos de la entrada como los de la salida piensan que es, en parte, en virtud de ciertas relaciones que se obtienen entre los estados del cerebro y las cosas del mundo que algunos de estos estados representan esas cosas externas. Dependiendo de la teoría concreta que se considere, se afirma que las relaciones en cuestión son dependencias nómicas, covarianzas causales fiables o algún tipo de correspondencia estructural. El desacuerdo fundamental entre los teóricos del input y del output reside en el hecho de que, para los primeros, lo que fija el contenido de las representaciones mentales se encuentra en las causas de estos estados internos, mientras que para los segundos se encuentra más bien en sus efectos (Papineau, 1999). Así pues, los teóricos de la producción no pueden abordar la cuestión de cómo se fija el contenido semántico sin tener en cuenta los procesos por los que las representaciones mentales tienen sus efectos particulares. Desde esta perspectiva, al abordar la cuestión de la intencionalidad subyacente, ya estamos abordando la cuestión (al menos hasta cierto punto) de cómo se realizan las manipulaciones sobre las representaciones mentales. Los teóricos del input, en cambio, pueden abordar la cuestión de cómo se fija el contenido semántico sin molestarse en preguntarse por los efectos que tienen las representaciones mentales. Para los teóricos del input, al tratar de averiguar cómo se fija el contenido semántico, las únicas interacciones causales que deben tenerse en cuenta son las que pertenecen a la cadena causal que da lugar a las representaciones mentales. Qué efectos posteriores tienen las representaciones mentales y cómo los provocan son cuestiones que, según los teóricos del input, deben abordarse de forma independiente.

Un ejemplo destacado de teoría del input es la Semántica Informacional de Dretske. Según este punto de vista, "los estados internos derivan su contenido (al menos en primera instancia) de su origen informacional, no de sus efectos" (Dretske, 1981, p. 209). Al hablar del origen informacional de los estados internos, Drestke se refiere a las dependencias nómicas en virtud de las cuales esos estados portan información sobre los acontecimientos que covarían con ellos (principalmente debido a la causalidad). Esta información, cuando se transporta de forma

completamente digitalizada, constituye su contenido semántico. Dretske también afirma que un estado interno se califica como una auténtica estructura cognitiva (un concepto o una creencia) sólo en la medida en que su contenido semántico "es un determinante causal del output en el sistema en el que ocurre" (1981, p. 199). Pero insiste en que el propio contenido semántico está "determinado únicamente por el origen de la estructura, por su patrimonio informativo" (1981, p. 202). La influencia causal del contenido semántico se consigue, según Drestke, por las propiedades físicas de los estados que poseen este contenido, cuando se produce un "alineamiento" de sus papeles representacional y causal.

Un ejemplo diferente de teoría de la entrada es la teoría de la dependencia asimétrica de Fodor. La cuenta de dependencia asimétrica es la respuesta de Fodor al problema de la robustez del significado, es decir, el hecho de que, en contraste con la mera información, el significado o contenido semántico de un símbolo es "insensible a la heterogeneidad de las causas (reales y posibles) de sus tokens" (1990, p. 90). Según Fodor, los tokens de un símbolo significan vaca (por tomar prestado uno de sus ejemplos característicos) porque son causados de manera fiable por vacas, pero también porque la existencia de tokens del símbolo que no son causados por vacas depende de la existencia de tokens del símbolo que son causados por ellas, pero no viceversa (1990). Es decir, los símbolos no causados por vacas dependen asimétricamente de los causados por vacas. Una vez que un símbolo -que puede ser un estado cerebral- adquiere contenido semántico, se pueden realizar cálculos sobre sus símbolos. Las computaciones son, según Fodor, "relaciones causales entre símbolos que respetan de forma fiable las propiedades semánticas de los relatos" (1998, p. 10).

Podemos ver en la propuesta de Fodor la forma particular en que los teóricos del input abordan las dos cuestiones centrales planteadas en el marco de la RTM. La respuesta que Fodor da al problema de la intencionalidad subyacente es su explicación de la dependencia asimétrica del contenido mental; y su explicación de los procesos mentales como computaciones, es decir, como "relaciones causales entre símbolos que respetan el contenido" (1998, p. 11), es su respuesta al problema de cómo se realizan las manipulaciones sobre las representaciones mentales. Estas respuestas se dan a cuestiones que se consideran mutuamente independientes. Uno puede adoptar el relato de la dependencia asimétrica sin haber abordado todavía la cuestión

de cómo se realizan las manipulaciones sobre las representaciones mentales; y, a la inversa, uno puede estar convencido por el relato computacional sin tener todavía una respuesta en al problema de la intencionalidad subyacente. Con respecto a la teoría informacional de Dretske, se puede decir claramente algo similar.

Consideremos ahora las teorías de output del contenido mental. Un ejemplo muy influyente es la teoría teleosemántica de Millikan (1984, 1989). Según Millikan, las representaciones mentales son estados internos cuya estructura mapea la estructura de los estados de cosas en el mundo. Sin embargo, como señala la propia Millikan, las relaciones cartográficas por sí solas no pueden fijar el contenido semántico. Diferentes estados de cosas pueden ser mapeados por el mismo estado interno dependiendo de la función de mapeo particular considerada (Millikan, 1984; véase también Cummins, 1989; Godfrey-Smith, 1994; Shea, 2013). Desde el punto de vista de Millikan, el contenido semántico se fija en última instancia por la forma en que se "consumen" las representaciones mentales, es decir, por la forma en que contribuyen causalmente al cumplimiento de las funciones propias de los dispositivos que las utilizan. Estas funciones son "efectos de dispositivos que, por hablar estrictamente, han sido *retenidos* (no diseñados) a pesar de las presiones de selección y que continúan siendo duplicados o reproducidos porque están produciendo estos efectos" (Millikan, 2017, p. 6). Así, la función específica de mapeo que fija el contenido de un grupo particular de representaciones mentales está determinada por los efectos del dispositivo que utiliza estas representaciones o, más precisamente, por aquellos de sus efectos que explican su reproducción exitosa. Si uno de estos efectos es provocar el resultado *A* en ciertas condiciones *C*, y si la forma en que el dispositivo normalmente satisface esta función es desencadenando el comportamiento *B* cuando se obtiene el estado de cosas *S*, podemos decir entonces, según Millikan, que la función de mapeo que fija el contenido de las representaciones utilizadas por el dispositivo, si son representaciones descriptivas, sería la que mapea su estructura en la estructura de *S*, pero si las representaciones son directivas, la función de mapeo en cuestión sería la que mapea su estructura en la estructura de *B* (Millikan, 1984). Las representaciones más primitivas son a la vez descriptivas y directivas (Millikan, 1995).

Desde el punto de vista de Millikan, lo que fija el contenido de las representaciones mentales no se encuentra en las causas de estos estados internos, sino en sus efectos. Sean cuales sean sus orígenes causales, los estados internos que no contribuyen causalmente al cumplimiento de las funciones propias de los dispositivos del sistema no pueden ser representaciones mentales. Desde esta perspectiva, una vez que comprendemos cómo los estados internos se convierten en representaciones mentales, es decir, una vez que comprendemos cómo algunos de los efectos de los dispositivos que utilizan estos estados en la actividad cognitiva determinan las funciones de mapeo que fijan su contenido semántico, comprendemos también cómo se manipulan las representaciones mentales en la actividad cognitiva. Por supuesto, se podrían dar muchos más detalles sobre estas manipulaciones. Sin embargo, dado que es teniendo sus efectos particulares en la actividad cognitiva como las representaciones mentales adquieren contenido semántico, las dos cuestiones centrales planteadas en el marco de la RTM no pueden sino resolverse ambas a la vez. En otras palabras, la cuestión de la intencionalidad subyacente y la cuestión de cómo se realizan las manipulaciones sobre las representaciones mentales no son mutuamente independientes.

Para los teóricos del input, como hemos visto, el contenido de las representaciones mentales está fijado por sus causas, independientemente del uso que los sistemas cognitivos hagan de estos estados internos. Desde este punto de vista, una vez que los estados internos adquieren contenido semántico, deben establecerse relaciones causales "respetuosas con el contenido" entre ellos (o una "alineación" de sus funciones representacionales y causales) para que su contenido pueda ser explotado. En cambio, para los teóricos del output, el contenido de las representaciones mentales queda fijado por sus efectos, es decir, por el uso que los sistemas cognitivos hacen de esos estados internos. Desde esta perspectiva, el contenido de las representaciones mentales depende de la propia actividad cognitiva. Estos distintos puntos de vista plantean diferentes problemas específicos a cada uno de ellos.

Por ejemplo, en lo que respecta a las teorías del input, podemos preguntarnos: ¿Cómo se supone que se produce la "alineación" de los papeles representacionales y causales de los estados internos? ¿Cómo pueden las relaciones causales "respetar el contenido"? Se podría aceptar que existen relaciones causales que respetan el contenido entre las representaciones

mentales, pero entonces cabe preguntarse cómo es posible que se haya obtenido esta disposición causal concreta. La respuesta no puede ser que las relaciones causales entre representaciones mentales son respetuosas con el contenido porque son sensibles al contenido, ya que entonces habría que explicar cómo estas relaciones causales podrían tener esta forma de sensibilidad.¹ Sin embargo, si el contenido semántico en sí mismo no explica el establecimiento de relaciones causales respetuosas con el contenido entre representaciones mentales, ¿cómo podría considerarse entonces un ingrediente crucial en una explicación correcta de las capacidades cognitivas? No se trata del problema general de la eficacia causal del contenido semántico, sino de una preocupación más específica relacionada con la respuesta particular que las teorías del input dan a este problema. Uno podría estar de acuerdo en que el problema de la eficacia causal del contenido semántico podría "resolverse mediante la propuesta de que [las representaciones mentales] son entidades físicas con propiedades semánticas alineadas con sus propiedades sintácticas" (Smortchkova, Dołęga, & Schlicht, 2020, p. 12). Pero la pregunta sigue siendo: ¿Cómo se produce este "alineamiento"? Tiene algo que ver el contenido semántico de las representaciones mentales con su establecimiento? La mera existencia de una "alineación" no puede bastar si no hay una explicación clara de este hecho, una explicación en la que el contenido semántico de las representaciones mentales desempeñe un papel sustantivo.

Estas "preguntas de cómo es posible" (Hutto & Myin, 2020) son ejemplos de problemas típicos que amenazan a las teorías de entrada del contenido mental. Las teorías de output, en cambio, no sufren este tipo de dificultades. No hay necesidad de que los teóricos de la salida expliquen cómo podrían establecerse manipulaciones físicas que respeten de manera fiable el contenido semántico de las representaciones mentales; más bien, según los teóricos de la salida, es al ser manipulados causalmente en la actividad cognitiva que algunos estados internos adquieren contenido semántico y, por tanto, pueden considerarse representaciones mentales. Esto no significa, por supuesto, que las teorías del output no adolezcan de dificultades diferentes, pero igualmente graves. El objetivo de este artículo es explorar algunas de estas dificultades. En particular, en la siguiente sección formularé una objeción que puede caracterizarse como una

¹ Por eso Fodor dice que "cuando un cálculo 'mira' una representación en su dominio, lo que es capaz de 'ver', o de operar sobre ella, es la identidad y la disposición de sus constituyentes. Nada más". (2008, p. 107).

especie de problema epifenomenalista. En pocas palabras, el problema es que, dada la forma en que los teóricos de la salida abordan las dos cuestiones centrales planteadas en el marco de la RTM, las representaciones mentales deben ser vistas como subproductos causalmente irrelevantes de los procesos físicos que subyacen a la cognición.

El argumento epifenomenalista

Según los teóricos del output, como hemos visto, lo que fija el contenido de las representaciones mentales se encuentra, no en las causas de estos estados internos, sino en sus efectos: Es por tener sus efectos particulares en la actividad cognitiva que algunos estados internos adquieren contenido semántico (obtienen su contenido semántico fijo) y por lo tanto pueden ser considerados representaciones mentales. Sin embargo, dada esta forma de entender cómo se fija el contenido semántico, y por tanto cómo se produce la intencionalidad subyacente, podemos preguntar: ¿Por qué deberíamos asumir que las representaciones mentales, como estados representacionales, son causalmente relevantes para los procesos físicos que subyacen a la cognición, en lugar de decir que son sólo un resultado causalmente irrelevante de estos procesos? Si es por tener un papel causal particular en la actividad cognitiva por lo que algunos estados internos adquieren contenido semántico, que tengan tal contenido no puede tomarse como uno de los factores que explican que tengan ese papel causal particular. Su posesión de contenido semántico es más bien un efecto de su papel causal y puede explicarse por él. Pero si el papel causal de los estados internos es lo que explica su posesión de contenido semántico, y no al revés, ¿por qué deberíamos suponer que los estados internos con contenido semántico, como estados representacionales, son causalmente relevantes para la actividad cognitiva? Dado que el contenido semántico sólo se fija una vez que los estados internos ya están cumpliendo su papel causal en la actividad cognitiva, parece que las representaciones mentales como tales son epifenoménicas, un mero subproducto de los procesos físicos que subyacen a la cognición. Éste es el argumento epifenomenalista contra las teorías de output del contenido mental.

El problema puede ilustrarse con la teoría teleosemántica de Millikan. Hemos visto que, según esta teoría, las representaciones mentales son estados internos cuya estructura mapea la estructura de los estados de cosas en el mundo. Sin embargo, dependiendo de la función de

mapeo considerada, diferentes estados de cosas pueden ser mapeados por el mismo estado interno. Una función cartográfica específica que determine un único estado de cosas que debe ser cartografiado por un estado interno sólo puede elegirse en la medida en que este estado interno esté siendo utilizado por un dispositivo que tenga una función propia concreta. Una vez que los estados internos se utilizan de este modo, y sólo una vez que se utilizan de este modo, adquieren contenido semántico y pueden considerarse representaciones mentales. Pero entonces parece que las representaciones mentales como tales no pueden contribuir causalmente al funcionamiento de los dispositivos que las utilizan en los procesos cognitivos. Las representaciones mentales son más bien un resultado de la actividad de estos dispositivos y no pueden, por tanto, explicarla causalmente. Son epifenoménicas.

Son necesarias algunas aclaraciones. En primer lugar, el problema no es que las representaciones mentales sean epifenoménicas porque el contenido semántico sea abstracto y no pueda, como tal, ser causa de nada (Egan, 2020a). Podemos estar de acuerdo en que si las representaciones mentales se naturalizan en términos de ciertos fenómenos no intencionales, "la compra explicativa del posit representacional será en el caso más básico la compra explicativa de esos fenómenos no intencionales" (Neander, 2017, p. 85). Sin embargo, la idea central de las teorías de output es que es porque los fenómenos no intencionales implicados en la actividad cognitiva tienen su compra explicativa (su papel causal particular) que el contenido semántico se fija. Esos fenómenos no intencionales explican la actividad cognitiva y también explican cómo las representaciones mentales adquieren contenido semántico, pero las representaciones mentales como tales no tienen ningún papel explicativo que desempeñar. Así pues, la razón por la que las representaciones mentales son epifenoménicas no es que el contenido semántico sea abstracto. Según el argumento epifenomenalista, las representaciones mentales son epifenomenales porque son un mero subproducto de los fenómenos no intencionales que subyacen a la cognición.

El argumento epifenomenalista también debe dissociarse de una de las principales objeciones a los relatos teleosemánticos: el problema de la indeterminación del contenido (Martínez, 2013). Según este problema, los relatos teleosemánticos son incapaces de asignar contenidos únicos a las representaciones mentales. Se han propuesto estrategias para resolver

este problema, como apelar a agrupaciones homeostáticas de propiedades (Martínez, 2013) o a explicaciones de la co-ocurrencia de conjuntos de propiedades (Artiga, 2021). Otros autores han defendido la idea de que la indeterminación es una característica real y no problemática de las representaciones mentales (Bergman, 2023). Sin embargo, según el argumento epifenomenalista, el problema no es que las representaciones mentales sean epifenomenales porque su contenido sea indeterminado. Incluso si aceptamos cualquiera de las soluciones propuestas al problema de la indeterminación del contenido, o la idea de que la indeterminación no es un problema en absoluto, las representaciones mentales siguen siendo epifenoménicas. Son un mero subproducto de los procesos físicos que subyacen a la cognición.

El argumento epifenomenalista está estrechamente relacionado con el problema del contenido no explotado, señalado por Cummins y sus colaboradores (Cummins *et.al.*, 2006). Este problema se formula como una preocupación específica para los relatos teleosemánticos, pero concierne a las teorías de output del contenido mental en general. Como hemos visto, para los teóricos de la salida, el contenido de las representaciones mentales se fija por sus efectos, es decir, por la forma particular en que esos estados internos se utilizan en el sistema. Antes de ser utilizados o explotados, los estados internos no tienen contenido semántico (su contenido no está fijado) y, por tanto, no pueden considerarse representaciones mentales. Sin embargo, como señalan Cummins y sus colaboradores, todos los modelos de redes neuronales de aprendizaje presuponen que el cerebro puede aprender a explotar sus estructuras no explotadas anteriormente. Este aprendizaje consiste en el ajuste de los pesos sinápticos para responder adecuadamente a los patrones de activación de entrada. Pero ¿qué significa aquí responder adecuadamente a tales patrones? Parece significar, al menos según Cummins y sus colaboradores, que el ajuste responde al contenido específico de los patrones en cuestión. Así pues, contrariamente a lo que afirman los teóricos del output, los estados internos parecen tener contenidos determinados antes de ser explotados en los procesos cognitivos. Es decir, parece existir un contenido no explotado.

En respuesta a la réplica de los teóricos de la salida de que las redes neuronales crean el contenido de sus patrones de entrada a medida que progresa el aprendizaje, Cummins y sus colaboradores dicen que, si eso fuera cierto, no tendría sentido contar como errores las

respuestas tempranas en el proceso de ajuste de los pesos sinápticos. "Y si las respuestas tempranas no son errores, ¿por qué cambiar los pesos en una dirección concreta? De hecho, ¿por qué cambiarlas en absoluto?" (Cummins *et.al.*, 2006, p. 197, nota 2). Si existe un contenido no explotado, parece que las teorías de output no están abordando realmente el problema de la intencionalidad infravalorada. Este problema sería el de explicar cómo las representaciones mentales adquieren su contenido no explotado, un problema que las teorías de output no están diseñadas para resolver. Pero ¿por qué no rechazar simplemente la idea de que el contenido no explotado es realmente contenido? Cummins y sus colaboradores afirman que lo que ellos llaman contenido -incluido el contenido no explotado- es "lo que en última instancia hace el trabajo en la ciencia cognitiva representacionalista " (2006, pp. 205-206). Según ellos, el contenido no explotado se presupone simplemente en cualquier explicación de la cognición que asuma que las capacidades representacionales se aprenden o evolucionan.

Desde el punto de vista de los teóricos del output, no puede haber contenido no explotado. Los teóricos de la salida ciertamente deben afirmar que los estados internos con contenido semántico se crean a medida que las nuevas capacidades cognitivas evolucionan o se aprenden. Para estos teóricos, el uso es anterior al contenido, y no al revés. Esta idea implica, sin embargo, que los estados internos con contenido semántico, como estados representacionales, no pueden ser explicativamente relevantes para los procesos físicos que subyacen a la cognición. Estos procesos explican las capacidades cognitivas y también explican cómo las representaciones mentales adquieren contenido semántico, pero las representaciones mentales como tales no tienen ningún papel explicativo que desempeñar. Este es, por supuesto, el argumento epifenomenalista. Pero Cummins y sus colaboradores tienen una opinión diferente sobre esta cuestión.

Parece correcto afirmar, como hacen Cummins y sus colaboradores, que las respuestas tempranas en el proceso de ajuste de los pesos sinápticos deben contar como errores. De lo contrario, no tendría sentido llamar a este proceso un proceso de ajuste. Pero Cummins y sus colaboradores también parecen pensar que, una vez que vemos estas respuestas tempranas como errores, debemos admitir que a lo que el ajuste está respondiendo es al contenido semántico de los patrones de entrada, y por tanto que estos patrones son representaciones. De

ahí que concluyan que debe existir un contenido no explotado. Sin embargo, éste no es un paso inferencial obvio.

El objetivo de las teorías de la cognición, incluidos los modelos de redes neuronales, es explicar las capacidades cognitivas y, como señala Egan, "[la] caracterización de un fenómeno como *capacidad* o *competencia* es en sí misma normativa" (2014, p. 129). Es decir, vemos una capacidad como algo que se supone que produce un resultado concreto. Cuando el resultado se produce (de la forma en que se supone que debe producirse), consideramos que la capacidad se ha ejercido con éxito. En caso contrario, se considera que el ejercicio de la capacidad no ha tenido éxito, o que sus resultados son erróneos. Egan también señala que esta caracterización normativa "viene dada preteóricamente, con anterioridad a la caracterización por parte de la teoría de los mecanismos subyacentes a la competencia", por lo que los elementos normativos "están ahí desde el principio" (2014, p. 129). Así, lo que las teorías de la cognición pretenden explicar es algo que vemos preteóricamente como un fenómeno normativo. Pero las explicaciones causales puras parecen incapaces de dar cuenta de los fenómenos normativos. Hay que añadir algo más a esas explicaciones para que funcionen. Dado que la noción de representación tiene en sí misma características normativas, las representaciones mentales parecen especialmente adecuadas para complementar las explicaciones causales de las capacidades cognitivas.

Sin embargo, si el explanandum crítico es la normatividad, podrían explorarse distintas posibilidades. No hay ninguna razón para considerar la representación mental como la única posibilidad. Por ejemplo, se pueden combinar formas de explicar la normatividad con enfoques computacionales sin plantear representaciones mentales. Es el caso de las visiones teleomecanicistas de la computación (Coelho Mollo, 2021). Estas visiones pueden ser útiles, no para explicar la representación mental, sino para explicar directamente la normatividad de la cognición. En cualquier caso, para constituir una opción viable, los beneficios explicativos de plantear representaciones mentales deben superar sus costes. El argumento epifenomenalista trata de mostrar que puede no haber beneficios explicativos asociados a la postulación de representaciones mentales, al menos si entendemos las representaciones mentales como lo hacen las teorías del output. En la siguiente sección, intentaré mostrar cómo se aplica el

argumento epifenomenalista a las teorías de output más recientes del contenido mental. Discutiré dos propuestas: La Semántica Varitel de Shea (2018) y el relato de las representaciones S de Gładziejewski y Miłkowski (2017). Al elegir estas propuestas, no estoy sugiriendo en absoluto que sean particularmente débiles o poco convincentes. Muy al contrario, ambas propuestas son cuentas muy detalladas e influyentes de la representación mental, por lo que es extremadamente fructífero discutir las. Esta es la razón principal de mi elección. Mi objetivo es mostrar que, a pesar de sus innovaciones teóricas, son vulnerables al argumento epifenomenalista.

El argumento epifenomenalista aplicado a dos propuestas recientes

Semántica Varitel

Según Nicholas Shea, las representaciones mentales son "particulares físicos que interactúan causalmente en virtud de propiedades no semánticas (por ejemplo, su forma física) de maneras que son fieles a su contenido semántico" (2018, p. 31). Esta caracterización parece concordar con las teorías del input: El contenido semántico es anterior al uso de las representaciones mentales y los procesos causales en los que intervienen las representaciones mentales deben respetar de algún modo este contenido. Sin embargo, la propuesta de Shea es una teoría output del contenido mental. Para ver esto, debemos considerar cómo, según esta propuesta, se fija el contenido semántico.

Desde el punto de vista de Shea, hay dos tipos de relaciones que los estados internos pueden mantener con las cosas del mundo que representan: correlaciones y correspondencia estructural. En virtud del hecho de que las propiedades de los estados internos se correlacionan con las propiedades de las cosas del mundo, los estados internos contienen información correlacional sobre esas cosas externas. Shea define la información correlacional de la siguiente manera: "que a esté en el estado F conlleva *información correlacional* sobre que b esté en el estado G si $P(Gb|Fa) \neq P(Gb)$ " (2018, p. 76). La correspondencia estructural, por su parte, "es un mapeo bajo el cual se preservan las relaciones" (2018, p. 116). Según Shea, existe una correspondencia estructural entre la relación V sobre los vehículos representacionales putativos

v_m y la relación H sobre las entidades mundanas $x_{(n)}$ "si existe una función f que mapea la v_m sobre la x_n $\forall i, j V(v_i, v_j) \leftrightarrow H(f(v_i), f(v_j))$ (*mutatis mutandis* para otras poliadicidades)" (2018, p. 117).

Las correlaciones y las correspondencias estructurales son "relaciones explotables". Desde el punto de vista de Shea, las relaciones explotables son propiedades extrínsecas de los vehículos internos que explican cómo la implementación de un algoritmo que implica secuencias de operaciones en esos vehículos puede conducir al cumplimiento de funciones de tarea. Las funciones de tarea son resultados caracterizados distalmente de un sistema que son robustos y se han estabilizado. Para Shea, los resultados robustos son aquellos que se producen "en respuesta a un rango de entradas diferentes" y "en un rango de condiciones externas relevantes diferentes" (2018, p. 55); y los resultados estabilizados son aquellos que, debido a que conducen a buenas consecuencias, han sido retenidos por medio del éxito evolutivo, el aprendizaje o la contribución a la persistencia del organismo (2018, p. 64). Ejemplos de resultados robustos y estabilizados, y por tanto ejemplos de funciones de tarea, son la señalización animal, las habilidades de navegación, el alcance de objetos, el reconocimiento de caras, etc. Ahora tenemos todos los componentes principales del relato de Shea, que él llama "Semántica Varitel":

Hay componentes internos que están en relaciones explotables con aspectos del entorno que son relevantes para lograr un resultado (una función de tarea), donde un proceso interno realizado sobre vehículos con esas propiedades constituye un algoritmo para lograr con éxito el resultado caracterizado distalmente de una manera sensible al contexto. (2018, p. 51).

Pero ¿cómo se fija exactamente el contenido semántico en este relato? Las relaciones explotables por sí mismas no pueden fijar el contenido semántico. Un vehículo interno que se encuentra en un estado F conlleva información correlacional sobre un montón de cosas diferentes; y, dada una relación V sobre vehículos internos v_m y prácticamente cualquier relación H sobre cosas x_n en el mundo, siempre es posible encontrar una función de mapeo que nos permita establecer una correspondencia estructural entre V y H . Según Shea, "las relaciones explotables son el vínculo entre los componentes internos y la función de tarea caracterizada distalmente que el organismo está realizando" (2018, p. 36), pero las funciones de tarea

caracterizadas distalmente son el factor crucial que fija el contenido semántico. Sin funciones de tarea caracterizadas distalmente, no hay elementos particulares en el entorno que puedan constituir los objetivos de las representaciones mentales. Dado que las funciones de tarea pueden realizarse mediante diferentes algoritmos y que diferentes algoritmos pueden requerir diferentes contenidos, el algoritmo específico implementado por el sistema para realizar una función de tarea es también uno de los factores que fijan el contenido semántico. Pero, de nuevo, sin funciones de tarea caracterizadas distalmente, no tiene sentido determinar ningún algoritmo concreto. Aunque las relaciones explotables y los algoritmos contribuyen a fijar el contenido, ambos dependen crucialmente de las funciones de tarea, que constituyen el factor principal.

Según Shea, la fijación del contenido semántico se debe a la realización de funciones de tarea. Desde este punto de vista, (1) los resultados robustos y estabilizados (funciones de tarea) son producidos por (2) un algoritmo que implica operaciones sobre (3) componentes internos que están en relaciones explotables con características del entorno forman un grupo natural, y este grupo es lo que "constituye a los componentes internos como representaciones" (2018, p. 51). Sin ser utilizados de esta manera para el desempeño de las funciones de la tarea, los componentes internos que mantienen relaciones explotables con cosas del mundo no pueden fijar su contenido semántico. Una vez que se utilizan así, y solo una vez que se utilizan así, adquieren contenido semántico y pueden considerarse representaciones mentales. La propuesta de Shea es, pues, una teoría de salida del contenido mental. Desde este punto de vista, el uso es anterior al contenido, y no al revés. En consecuencia, el argumento epifenomenalista se aplica a la teoría de Shea: Las secuencias de operaciones sobre los componentes internos que se encuentran en relaciones explotables con las cosas del mundo explican la realización de las funciones de la tarea y también explican cómo las representaciones mentales adquieren contenido semántico, pero las representaciones mentales como tales no tienen ningún papel explicativo que desempeñar. Las representaciones mentales son más bien un efecto de esta actividad y, por tanto, no pueden explicarla causalmente.

El argumento epifenomenalista, tal como se aplica al relato de Shea, debe distinguirse de una importante objeción formulada por Frances Egan (2020b). La objeción se refiere al hecho de que la apelación a las representaciones mentales en la explicación de Shea del desempeño de las

funciones de la tarea parece ser prescindible. Según Egan, una explicación completa de la realización de funciones de tarea caracterizadas distalmente debe mencionar correlaciones entre estados del mundo y componentes internos del sistema, y entre esos componentes y resultados en el mundo. Sin embargo, Egan argumenta que no se añade nada esencial a esta explicación diciendo que los componentes internos son representaciones: Este discurso representacional es una "glosa" conveniente que nos permite llamar la atención sobre las correlaciones relevantes, "pero la atribución de contenido no es esencial para la explicación del éxito del organismo" (2020b, p. 374). Shea replica que caracterizar los componentes internos como representaciones es importante para entender "por qué algunos resultados producidos sólidamente y no otros deberían contar como éxitos" (2020, p. 406).

En mi opinión, Egan tiene razón, pero el problema que señala es diferente del problema destacado por el argumento epifenomenalista. Podemos admitir, en aras del argumento, que los procesos y propiedades relacionales descritos por Shea dan lugar a representaciones reales.² Es decir, podemos admitir que el discurso representacional de Shea no es una mera "glosa intencional" que no hace más que caracterizar "procesos computacionales de manera congruente con nuestra comprensión de sentido común de nosotros mismos" (Egan, 2019, p. 256), sino que las representaciones de las que habla son componentes internos reales con contenido semántico. Según el argumento epifenomenalista, sin embargo, el problema no es si las teorías de output logran identificar representaciones mentales reales y explicar cómo estas representaciones reales adquieren contenido semántico. Eso puede admitirse. El problema es que la forma en que se supone que lo hacen las teorías de output implica que las representaciones mentales son epifenoménicas. Las representaciones mentales aparecen sólo una vez (y sólo porque) los procesos físicos que subyacen a la realización de las funciones de la tarea ya están haciendo su trabajo. Por tanto, las representaciones mentales no pueden explicar causalmente estos procesos. Las representaciones de las que habla Shea bien podrían ser estados internos reales con contenido semántico. Pero entonces debemos concluir que, como tales, son un mero subproducto de los procesos físicos que subyacen a la cognición.

² Para serias dudas sobre la implementación real de estos procesos y propiedades relacionales en el cerebro, incluso en los casos que el propio Shea toma como ejemplo, véase (Burnston, 2021).

Como teórico del output, Shea puede responder que el contenido semántico se fija en cuanto se implementan los algoritmos que realizan las funciones de la tarea. Sin embargo, eso no ayuda. Cuando los vehículos internos adquieren contenido semántico, los procesos físicos que subyacen a la realización de las funciones de la tarea ya están en marcha. Por tanto, las representaciones mentales no son más que un resultado causalmente irrelevante de estos procesos.

Las representaciones S de Gładziejewski y Miłkowski

Según Gładziejewski y Miłkowski (2017), las representaciones mentales estructurales (o representaciones S) son componentes de los sistemas cognitivos que representan sus objetivos en virtud de la similitud estructural que guardan con ellos (véase también Gładziejewski, 2016a). La similitud estructural es un tipo de semejanza de segundo orden (O'Brien y Opie, 2004; véase también Gładziejewski, 2016b). A diferencia de la semejanza de primer orden, que implica compartir propiedades físicas, la semejanza de segundo orden consiste en relaciones que reflejan relaciones. Gładziejewski y Miłkowski adoptan la caracterización de O'Brien y Opie de la semejanza de segundo orden. Consideremos un sistema $S_V = (V, \mathfrak{R}_V)$ que comprende un conjunto V de objetos (que pueden ser conceptuales o concretos) y un conjunto \mathfrak{R} de relaciones (que pueden ser espaciales, causales, inferenciales, etc.) definidas sobre los miembros de V . Según O'Brien y Opie, existe una semejanza de segundo orden entre $S_V = (V, \mathfrak{R}_V)$ y otro sistema $S_O = (O, \mathfrak{R}_O)$ si se cumple la siguiente condición:

[...] para al menos *algunos* objetos en V y *algunas* relaciones en \mathfrak{R}_V existe un mapeo uno a uno de V a O y un mapeo uno a uno de \mathfrak{R}_V a \mathfrak{R}_O tal que cuando una relación en \mathfrak{R}_V es válida para objetos en V , la relación correspondiente en \mathfrak{R}_O es válida para los objetos correspondientes en O (2004, p. 11).

Como bien señalan Gładziejewski y Miłkowski, la similitud estructural no es suficiente para conferir a los componentes internos el estatus de representaciones mentales. La similitud estructural es "barata": Las representaciones S son estructuralmente similares a sus objetivos, pero también a un montón de cosas diferentes (e irrelevantes), y puede haber relaciones de

similitud estructural entre entidades externas y componentes del sistema que no sean representaciones S. Inspirándose en el trabajo de Shea (2014), Gładziejewski y Miłkowski sostienen que las representaciones-S solo entran en juego en la medida en que los compromisos de los sistemas cognitivos con las cosas del mundo dependen, en un sentido no trivial, de la similitud estructural entre los vehículos de las representaciones-S y esas cosas externas. La similitud estructural, afirman, "debe entenderse como una relación que es *explotable* para algún gran sistema que utiliza representaciones" (2017, p. 340).

Pero ¿qué significa exactamente que la similitud estructural sea *explotable*? Gładziejewski y Miłkowski ven las representaciones S como componentes de mecanismos cognitivos. Adoptan una explicación neomecanicista de la cognición (Bechtel, 2008; Boone y Piccinini, 2016; Craver, 2007; Miłkowski, 2013) según la cual el sistema cognitivo es una colección de mecanismos, y un mecanismo es un "conjunto de componentes organizados y operaciones de componentes que conjuntamente permiten al sistema más amplio exhibir un determinado fenómeno", que a menudo se entiende como una capacidad del sistema (Gładziejewski y Miłkowski, 2017, p. 341). Así pues, los mecanismos se individualizan, al menos en parte, por referencia a la función que desempeñan. Como dicen Gładziejewski y Miłkowski, "son esencialmente mecanismos *de* tal o cual función cognitiva (lectura mental, control motor, atención, categorización perceptiva, navegación espacial, etc.)" (2017, p. 341). Los componentes de los mecanismos tienen también funciones, que derivan de las funciones de sus mecanismos: Son el conjunto de operaciones que esos componentes realizan y que contribuyen al cumplimiento de la función de su mecanismo. Así, dado que las representaciones-S son componentes de mecanismos cognitivos, su función debe derivar de la función de sus mecanismos. Eso significa esencialmente, según Gładziejewski y Miłkowski, "que *la similitud* estructural entre la representación y lo que representa es lo que contribuye al buen funcionamiento del mecanismo" (2017, p. 341). Es en virtud de esta contribución a la actividad cognitiva que la similitud estructural puede considerarse una relación *explotable*.

Sin embargo, como hemos visto, la similitud estructural por sí misma no puede fijar el contenido semántico. ¿Cómo se determinan entonces los objetivos de las representaciones S en

la explicación de Gładziejewski y Miłkowski del contenido mental?³ Dado el marco neomecanicista adoptado por estos autores, la respuesta parece clara: los objetivos de las representaciones-S son aquellas cosas del mundo a las que las representaciones-S tendrían que ser estructuralmente similares para contribuir al cumplimiento de la función de su mecanismo. Por tanto, es la función de los mecanismos en los que están integradas las representaciones S la que determina los objetivos de dichas representaciones. En otras palabras, lo que fija el contenido de las representaciones S no se encuentra en las causas de estos componentes internos, sino en sus efectos, es decir, en su contribución causal al cumplimiento de la función de su mecanismo. En el relato (implícito) de Gładziejewski y Miłkowski, el uso es anterior al contenido: Al ser explotados en la actividad cognitiva, los vehículos de las representaciones S adquieren contenido semántico y, por tanto, pueden considerarse representaciones mentales. Una vez que las representaciones-S, como componentes de los mecanismos cognitivos, se utilizan en el desempeño de las funciones de estos mecanismos, y sólo una vez que se utilizan así, puede haber objetivos (cosas en el mundo) representaciones-S, no sólo son, sino que se supone que son estructuralmente similares. Así, como la Semántica Varitel de Shea, el relato de Gładziejewski y Miłkowski de las representaciones-S es una teoría de salida del contenido mental. El argumento epifenomenalista se aplica, por tanto, a este relato: Las operaciones causales sobre los componentes de los mecanismos cognitivos explican cómo estos mecanismos cumplen sus funciones, y también explican cómo algunos de estos componentes se convierten en representaciones-S (cómo puede haber objetivos a los que estos componentes, no sólo son, sino que se supone que son estructuralmente similares), pero las representaciones-S como tales no tienen ningún papel explicativo que desempeñar. Las representaciones-S sólo adquieren contenido semántico una vez que los mecanismos que las explotan en la actividad cognitiva ya están cumpliendo sus funciones, por lo que no pueden, como estados representacionales,

³ Para ser justos, esta no es una cuestión que Gładziejewski y Miłkowski aborden en su artículo. Sin embargo, me parece que las ideas que discuten y defienden presuponen una explicación de cómo se fija el contenido semántico en el caso específico de las representaciones S. Mi afirmación es que esta explicación no es correcta. Mi afirmación es que esta explicación es una especie de teoría de la salida del contenido mental y uno de mis propósitos en esta subsección es tratar de hacerla explícita.

contribuir causalmente a la actividad de estos mecanismos. Las representaciones-S son, pues, epifenoménicas.

Pero antes de concluir esto, debemos considerar el hecho de que el principal objetivo de Gładziejewski y Miłkowski en su artículo es mostrar cómo la similitud estructural como tal puede ser causalmente relevante para el éxito de los mecanismos que explotan las representaciones-S. Para demostrarlo, Gładziejewski y Miłkowski adoptan la teoría intervencionista de la relevancia causal de Woodward (2003, 2021). Según esta teoría, si una variable C (la causa putativa) es causalmente relevante para una variable E (el efecto), entonces es cierto que "si nosotros (o la naturaleza) fuéramos capaces de manipular C de la manera correcta, habría algún cambio asociado en E " (Woodward, 2021, p. 76). Más concretamente, C causa E en las circunstancias de fondo B si

- (i) hay alguna intervención posible que cambia el valor de C de tal manera que (ii) si esa intervención ocurriera en B , habría un cambio asociado en el valor de E o en la distribución de probabilidad $P(E)$ de esos valores (Woodward, 2021, p. 77).

Gładziejewski y Miłkowski utilizan el marco intervencionista para explicar la relevancia causal de la similitud estructural de la siguiente manera. Los valores de C corresponden a los grados de similitud estructural (dentro de un cierto rango) entre las representaciones S y sus objetivos, y los valores de E corresponden a los grados de éxito de los mecanismos que explotan estas representaciones en el desempeño de sus funciones. Por tanto, para que la similitud estructural sea causalmente relevante para el éxito de los mecanismos cognitivos basta con que las intervenciones que cambien el valor de C (en determinadas circunstancias de fondo) provoquen cambios sistemáticos en el valor de E . En concreto, al aumentar el valor de C debemos aumentar el valor de E , y al disminuir el valor de C debemos disminuir el valor de E .

Sin embargo, como señalan Gładziejewski y Miłkowski, hay un problema con esta forma de entender la relevancia causal de la similitud estructural. El problema es que, para realizar una intervención sobre la similitud estructural, necesariamente debemos intervenir en la estructura de al menos uno de sus relatos, es decir, debemos intervenir en la estructura de las representaciones de S o en la estructura de sus objetivos. Dada esta situación, parece mucho

más parsimonioso decir que lo causalmente relevante para el éxito de los mecanismos cognitivos no es la similitud estructural como tal, sino la estructura de los vehículos S-representacionales y/o la estructura de sus objetivos. Para replicar a esta objeción, Gładziejewski y Miłkowski proponen distinguir entre "intervenciones que cambian la forma en que *actúa* algún sistema cognitivo (conductual o cognitivamente) e intervenciones que cambian el éxito de sus acciones" (2017, p. 345). Señalan que los cambios en la estructura de los vehículos pueden cambiar la forma en que actúan los mecanismos cognitivos, pero no afectan necesariamente al éxito si los cambios resultantes en la acción van acompañados de cambios apropiados en el entorno. Según ellos, estos cambios apropiados en el entorno preservan el éxito al restaurar el ajuste estructural entre las representaciones S y sus objetivos. Gładziejewski y Miłkowski sostienen que es imposible decir cómo las intervenciones en la estructura de los vehículos afectarían al éxito de los mecanismos cognitivos independientemente de los hechos sobre los objetivos, "o más exactamente, independientemente de los hechos relativos a la similitud estructural entre el vehículo y el objetivo" (2017, p. 356). Concluyen de esto que "las intervenciones sobre la estructura del vehículo cambian el éxito *solo en la medida en que cambian el grado de similitud entre el vehículo y el objetivo*" (2017, p. 356), y por tanto que la similitud estructural como tal es causalmente relevante para el éxito.

Sin embargo, consideremos un componente desnudo de un mecanismo cognitivo, es decir, un componente que no es el vehículo de una representación-S. Suponemos que tales componentes existen, ya que suponemos que no todos los componentes de un mecanismo cognitivo son vehículos de representación. Si realizamos intervenciones en la estructura de este componente, podemos esperar observar cambios en las acciones del mecanismo. Es decir, podemos suponer que el componente tiene un papel causal en el funcionamiento del mecanismo y que las intervenciones en su estructura pueden afectar a este papel causal y producir cambios en la forma de actuar del mecanismo. Como hemos visto, los cambios en la acción pueden afectar al éxito, pero no necesariamente, si van acompañados de cambios apropiados en el entorno. Así, podemos identificar correlaciones sólidas y sistemáticas entre las modificaciones en la estructura del componente desnudo que afectan al éxito y las modificaciones en el entorno que lo restablecen, y podemos establecer una correspondencia estructural entre el componente

desnudo y esas condiciones externas. Esta correspondencia estructural puede considerarse una semejanza de segundo orden y, por tanto, una especie de similitud estructural. Sin embargo, al identificar estas correlaciones y establecer esta semejanza de segundo orden, no estamos demostrando que la semejanza estructural entre el componente desnudo y las condiciones externas en cuestión sea causalmente relevante para el éxito. Lo que hemos demostrado es algo mucho más modesto: que el éxito de los mecanismos cognitivos en el desempeño de sus funciones depende sistemáticamente del mundo externo. Esto es cierto para cualquier mecanismo, incluso para los mecanismos que realizan funciones no cognitivas.

Ahora bien, ¿hay alguna diferencia fundamental entre ese componente desnudo y un vehículo S-representacional en cuanto a su significado causal? No podemos decir simplemente que el vehículo de representación S tiene realmente propiedades de representación, porque de lo que se trata es de dilucidar a qué equivale la posesión de estas propiedades en términos de influencia causal. Tanto el componente desnudo como el vehículo representacional S son estructuralmente similares a las condiciones externas relevantes para el éxito del mecanismo. Es decir, en ambos casos, los cambios en la similitud estructural afectan al éxito. La diferencia crucial entre el componente desnudo y el vehículo de representación S es que, mientras que el componente desnudo sólo es estructuralmente similar a las condiciones externas relevantes para el éxito (la similitud estructural es barata), *se supone que* el vehículo de representación S es, además, estructuralmente similar a estas condiciones. Sin embargo, ¿tiene esta diferencia un significado causal real? La propuesta de Gładziejewski y Miłkowski no responde a esta pregunta.

Sólo una vez que un mecanismo cognitivo ya está cumpliendo su función existen realmente condiciones externas relevantes para su éxito. Por tanto, sólo cuando un mecanismo cognitivo ya está cumpliendo su función puede establecerse una semejanza de segundo orden entre esas condiciones y los componentes del mecanismo. Podemos ver estos componentes como representaciones S y las condiciones externas como sus objetivos. Sin embargo, al hacer esto no estamos demostrando que la similitud estructural entre las representaciones-S y los objetivos sea causalmente relevante para el éxito del mecanismo, aunque observemos que el aumento de la similitud estructural aumenta el éxito y su disminución lo reduce. Este tipo de correlación sólo refleja el hecho de que el éxito de los mecanismos cognitivos depende

sistemáticamente del mundo externo. Dado que la similitud estructural en cuestión sólo puede establecerse en la medida en que el mecanismo ya está cumpliendo su función, esta similitud estructural no puede contribuir causalmente al éxito. Es más bien el hecho de que ya exista un mecanismo cognitivo cumpliendo su función lo que permite establecer una similitud estructural entre los componentes del mecanismo y las condiciones externas relevantes para el éxito, y no esta similitud estructural la que explica causalmente el éxito del mecanismo al realizar su trabajo. Como vemos, se trata sólo de una forma diferente de formular el argumento epifenomenalista. Por lo tanto, podemos concluir finalmente que el argumento epifenomenalista se aplica a la propuesta de Gładziejewski y Miłkowski: Las representaciones-S como tales son sólo un subproducto causalmente irrelevante de la actividad de los mecanismos cognitivos.

Conclusión

Si la forma en que se utilizan los estados internos en la actividad cognitiva explica el contenido, entonces el contenido no puede explicar el uso de esos estados internos, y por tanto no puede explicar su papel causal en la actividad cognitiva. Esta es la idea central que subyace al argumento epifenomenalista contra las teorías output del contenido mental. Sin embargo, el argumento no tiene un mero propósito negativo, ya que mi objetivo es hacer una pequeña contribución a la consolidación de un espacio lógico para nuevas posibilidades en nuestra comprensión de la cognición. Estas nuevas posibilidades ya están siendo exploradas y mi esperanza es que la discusión en este artículo pueda ayudarnos a verlas como opciones realmente importantes.

Referencias

- Artiga, M. (2021). Más allá de los Puntos Negros y las Cosas Nutritivas: Una Solución al Problema de la Indeterminación. *Mind & Language*, 36(3), 471-490.
- Bechtel, W. (2008). *Mental Mechanisms*. Routledge.
- Bergman, K. (2023). ¿Debería temer el teleosemanticista la indeterminación semántica? *Mind & Language*, 38(1), 296-314.
- Boone, W. & Gualtieri P. (2016). La revolución de la neurociencia cognitiva. *Synthese*, (193) 1509-1534.

- Burnston, D. (2021). Contenidos, vehículos y análisis de datos complejos en neurociencia. *Synthese*. (199) 1617-1639.
- Coelho, D. (2021). Why go for a computational-based approach to cognitive representation. *Synthese*. (199) 6875-6895.
- Craver, C. (2007). *Explaining the Brain*. Oxford University Press.
- Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge, Massachusetts: MIT Press.
- Cummins, R., Blackmon, J., Byrd, D., Lee, A. y Roth, M. (2006). Representation and Unexploited Content. En G. Macdonald & D. Papineau (Eds.), *Teleosemantics* (pp. 195-207). Oxford University Press.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.
- Egan, F. (2014). Cómo pensar sobre el contenido mental. *Philosophical Studies*. 170(1), 115-135.
- Egan, F. (2019). La naturaleza y función del contenido en los modelos computacionales. En M. Sprevak & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 247-258). Routledge.
- Egan, F. (2020a). A Deflationary Account of Mental Representation. En J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *What are Mental Representations?* (pp. 26-53). Oxford University Press.
- Egan, F. (2020b). Content is Pragmatic: Comments on Nicholas Shea's *Representation in Cognitive Science*. *Mind & Language*, 35(3), 368-376.
- Fodor, J. (1990). *A Theory of Content and Other Essays*. MIT Press.
- Fodor, J. (1998). *Concepts*. Oxford: Oxford University Press.
- Fodor, J. (2008). *LOTE 2. El lenguaje del pensamiento revisitado*. Oxford University Press.
- Gładziejewski, P. (2016a). Action Guidance is not Enough, Representations Need Correspondence too: Un alegato a favor de una teoría bifactorial de las representaciones. *Nuevas ideas en psicología*, (40). 13-25.
- Gładziejewski, P. (2016b). Codificación predictiva y representacionalismo. *Synthese*. (193) 559-582.
- Gładziejewski, P., & Miłkowski, M. (2017). Representaciones estructurales: Causalmente relevantes y diferentes de los detectores. *Biología y Filosofía*. (32) 337-355.

- Godfrey-Smith, P. (1994). A Continuum of Semantic Optimism. En S. Stich & T. Warfield (Eds.), *Mental Representation* (pp. 259-277). Blackwell.
- Hutto, D. D., & Myin, E. (2020). Deflating Deflationism about Mental Representation. En J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *¿Qué son las representaciones mentales?* (pp. 79-100). Oxford University Press.
- Martínez, M. (2013). Teleosemántica e indeterminación. *Dialéctica*, 67(4), 427-453.
- Miłkowski, M. (2013). *Explicando la mente computacional*. MIT Press.
- Millikan, R. G. (1984). *Lenguaje, pensamiento y otras categorías biológicas*. MIT Press.
- Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, 86(6), 281-297.
- Millikan, R. G. (1995). Representaciones Pushmi-Pullyu. *Perspectivas filosóficas*. (9) 185-200.
- Millikan, R. G. (2017). *Más allá de los conceptos*. Oxford University Press.
- Neander, K. (2017). *Una marca de lo mental*. MIT Press.
- O'Brien, G., y Opie, J. (2004). Notas hacia una teoría estructuralista de las representaciones mentales. En H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in Mind* (pp. 1-20). Elsevier.
- Papineau, D. (1999). Normatividad y juicio. *Actas de la Sociedad Aristotélica, Volúmenes Suplementarios*. (3) 17-43.
- Shea, N. (2013). El requisito de isomorfismo de Millikan. En D. Ryder, J. Kingsbury & K. Williford (Eds.), *Millikan and Her Critics* (pp. 63-80). Wiley-Blackwell.
- Shea, N. (2014). Isomorfismo explotable y representación estructural. *Actas de la Sociedad Aristotélica*. (114) 123-144.
- Shea, N. (2018). *Representación en ciencia cognitiva*. Oxford University Press.
- Shea, N. (2020). Representación en ciencias cognitivas: Replies. *Mind & Language*. 35(3), 402-412.
- Smortchkova, J., Dołęga, K., & Schlicht, T. (2020). Introduction. En J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *¿Qué son las representaciones mentales?* (pp. 1-25). Oxford University Press.
- Woodward, J. (2003). *Making Things Happen*. Oxford University Press.
- Woodward, J. (2021). *Causation with a Human Face*. Oxford University Press.