ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88

Revista Sincronía

Biannual Journal of Philosophy, Literary Studies and Humanities

Philosophy

An epiphenomenalist argument against output theories of mental content.

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

 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

Received: 03/02/2025 **Revised**: 10/03/2025 **Approved**: 28/04/2025

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, 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.

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

Cómo citar este artículo (APA):

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

En la lista de referencias:
Bernache, F. (2025). An epiphenomenalist argument against output theories of mental content. *Revista Sincronía*. *XXIX*(88). 19-44
DOI: 10.32870/sincronia.axxix.n88.2.25b

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

objetivo de este artículo es desarrollar este argumento epifenomenalista y mostrar cómo se aplica a propuestas teóricas recientes.

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

Introduction

The main purpose of this paper is to present an objection against output theories of mental content. This objection can be characterized as a kind of epiphenomenalist problem. I argue that the strategy output theorists adopt for explaining how the semantic content of mental representations gets fixed entails that these internal states are no more than a causally irrelevant by-product of cognitive processes. In section 1, input and output theories of mental content are briefly explained. I focus my atention on the different ways these theories address the two central issues raised in the representationalist framework: the problem of underived intentionality and the problem of how manipulations on mental representations are performed. In section 2, I show how the particular way output theories address these issues entails that mental representations are epiphenomenal. That is, I formulate what I call the epiphenomenalist argument against output theories of mental content. Finally, in section 3, I show how the epiphenomenalist argument applies to two recent and very influential proposals: Shea's Varitel Semantics (2018) and Gładziejewski and Miłkowski's account of S-representations (2017).

Input and output theories of mental content

According to the Representational Theory of Mind (RTM), the physical processes that underlie cognitive capacities involve manipulations performed on mental representations. Mental representations are generally conceived of as physical states that causally interact with other physical states. Of course, mental representations are not merely physical states. They are physical states that are supposed to refer to or to be about other things, properties, or events. In other words, they have intentionality. However, in contrast to the intentionality

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

of other types of representations, the intentionality of mental representations is generally assumed to be of a very special kind. In effect, one common view among partisans of RTM is that mental representations have, not only intentionality, but underived or original intentionality. That is, partisans of RTM generally assume that the semantic properties of other types of representations (e.g., linguistic representations) derive from those of mental representations, but that mental representations themselves (or at least some of them) do not derive their semantic properties from the semantic properties of representations of other types. So, there are two central questions that anyone interested in developing a particular version of RTM must address:

1. How do mental representations (with underived intentionality) acquire semantic content?
2. How are mental representations manipulated in the processes underlying cognitive capacities?

Philosophers have adopted two different attitudes concerning the relation between these questions. The first attitude consists in taking the questions to be mutually independent: We can give a relatively complete answer to one of them and then try to figure out what answer could be given to the other. The answers must, of course, be compatible, but the questions are seen as expressing two independent problems. The second attitude consists in taking the questions to have some degree of mutual dependence. For philosophers adopting this attitude, in order to explain how mental representations acquire semantic content, we must consider the effects those physical states have in the system in which they occur. In taking into account these effects, we must also take into account the way in which mental representations causally interact with other physical states inside the system, and our understanding of those causal interactions partially constrains the answer we can give to the question of how mental representations are manipulated in cognitive activity.

Revista Sincronía
ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Biannual Journal of Philosophy, Literary Studies and Humanities

These attitudes are related to preference for input or output naturalist theories of mental content. Both input and output theorists think that it is, partly, in virtue of certain relations that obtain between states in the brain and things in the world that some of these states represent those external things. Depending on the particular theory considered, the relations in question are claimed to be nomic dependencies, reliable causal covariances, or some kind of structural correspondence. The fundamental disagreement between input and output theorists lies in the fact that, for the former, what fixes the content of mental representations is to be found in the causes of these internal states, while for the latter it is rather to be found in their effects (Papineau, 1999). So, output theorists cannot address the issue of how semantic content is fixed without taking into account the processes by which mental representations have their particular effects. From this perspective, in addressing the question of underived intentionality, we are already addressing the question (at least to some extent) of how manipulations on mental representations are performed. Input theorists, in contrast, can address the issue of how semantic content is fixed without bothering to ask about the effects mental representations have. For input theorists, in trying to figure out how semantic content is fixed, the only causal interactions that must be taken into account are those belonging to the causal chain that gives rise to mental representations. What downstream effects mental representations have and how they bring about them are questions that, according to input theorists, must be independently addressed.

A prominent example of an input theory is Dretske's Informational Semantics. According to this view, "internal states derive their content (in the first instance at least) from their informational origin, not from their effects" (Dretske, 1981, p. 209). In talking about the informational origin of internal states, Drestke is referring to the nomic dependencies in virtue of which those states carry information about the events that covary with them (mainly due to causation). This information, when it is carried in a completely digitalized form, constitutes their semantic content. Dretske also claims that an internal state qualifies as an authentic cognitive structure (a concept or a belief) only insofar as its semantic content

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

"is a causal determinant of output in the system in which it occurs" (1981, p. 199). But he insists that semantic content itself is "determined solely by the structure's origin—by its information heritage" (1981, p. 202). The causal influence of semantic content is achieved, according to Drestke, by the physical properties of the states that possess this content, when an "alignment" of their representational and causal roles is produced.

A different example of an input theory is Fodor's asymmetric dependence account. The asymmetric dependence account is Fodor's answer to the problem of the robustness of meaning, that is, the fact that, in contrast to mere information, the meaning or semantic content of a symbol is "insensitive to the heterogeneity of the (actual and possible) causes of its tokens" (1990, p. 90). According to Fodor, tokens of a symbol mean cow (to borrow one of his characteristic examples) because they are reliably caused by cows, but also because the existence of tokens of the symbol that are not caused by cows depends on the existence of tokens of the symbol that are caused by them, but not vice versa (1990). That is, non-cow caused tokens of the symbol asymmetrically depend on cow caused ones. Once a symbol—which can be a brain state—acquires semantic content, computations can be performed on its tokens. Computations are, according to Fodor, "causal relations among symbols which reliably respect semantic properties of the relata" (1998, p. 10).

We can see in Fodor's proposal the particular way in which input theorists address the two central issues raised in the RTM framework. The answer Fodor gives to the problem of underived intentionality is his asymmetric dependence account of mental content; and his account of mental processes as computations, that is, as "content-respecting causal relations among symbols" (1998, p. 11), is his answer to the problem of how manipulations on mental representations are performed. These answers are given to questions that are seen as mutually independent. One can adopt the asymmetric dependence account without having yet addressed the issue of how manipulations on mental representations are performed; and, conversely, one can be convinced by the computational account while still having no answer to the problem of underived intentionality. Concerning Dretske's informational account, something similar can clearly be said.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

Consider now output theories of mental content. One very influential example is Millikan's teleosemantic theory (1984, 1989). According to Millikan, mental representations are internal states whose structure maps the structure of states of affairs in the world. As Millikan herself points out, however, mapping relations by themselves cannot fix semantic content. Different states of affairs can be mapped by the same internal state depending on the particular mapping function considered (Millikan, 1984; see also Cummins, 1989; Godfrey-Smith, 1994; Shea, 2013). On Millikan's view, semantic content is ultimately fixed by the way in which mental representations are "consumed", that is, by the way in which they causally contribute to the fulfillment of the proper functions of the devices that use them. These functions are "effects of devices that have, to speak strictly, been *retained* (not designed) despite selection pressures and that continue to be duplicated or reproduced because they are producing these effects" (Millikan, 2017, p. 6). So, the specific mapping function that fixes the content of a particular group of mental representations is determined by the effects of the device that uses these representations or, more precisely, by those of its effects that explain its successful reproduction. If one of these effects is to bring about outcome *A* in certain conditions *C*, and if the way the device normally satisfies this function is by triggering behavior *B* when state of affairs *S* obtains, we can then say, according to Millikan, that the mapping function that fixes the content of the representations used by the device, if they are descriptive representations, would be the one that maps their structure onto the structure of *S*, but if the representations are directive, the mapping function in question would be the one that maps their structure onto the structure of *B* (Millikan, 1984). More primitive kinds of representations are at once descriptive and directive (Millikan, 1995).

On Millikan's view, then, what fixes the content of mental representations is to be found, not in the causes of these internal states, but in their effects. Whatever their causal origins, internal states that do not causally contribute to the fulfillment of the proper functions of the system's devices cannot be mental representations. From this perspective, once we understand how internal states become mental representations, that is, once we

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

understand how some of the effects of the devices that use these states in cognitive activity determine the mapping functions that fix their semantic content, we also understand how mental representations are manipulated in cognitive activity. Of course, much more details about these manipulations could be provided. However, given that it is by having their particular effects in cognitive activity that mental representations acquire semantic content, the two central issues raised in the RTM framework cannot but be solved both at once. In other words, the question of underived intentionality and the question of how manipulations on mental representations are performed are not mutually independent.

For input theorists, as we have seen, the content of mental representations is fixed by their causes, independently of the use made of these internal states by cognitive systems. On this view, once internal states acquire semantic content, "content-respecting" causal relations between them (or an "alignment" of their representational and causal roles) must be established for their content to get exploited. In contrast, for output theorists, the content of mental representations is fixed by their effects, that is, by the use made of these internal states by cognitive systems. From this perspective, the content of mental representations depends on cognitive activity itself. These different views raise different problems specific to each of them.

For instance, concerning input theories, we can ask: How is the "alignment" of the representational and causal roles of internal states supposed to be produced? How can causal relations be "content-respecting"? One might accept that there are content-respecting causal relations between mental representations, but then one can ask how this particular causal arrangement could possibly have obtained. The answer cannot be that causal relations between mental representations are content-respecting because they are content-sensitive, for then it should be explained how these causal relations could have this form of sensitivity.[1] However, if semantic content itself does not explain the establishment of content-respecting causal relations between mental representations, how could it then

[1] That is why Fodor says that "when a computation 'looks at' a representation in its domain, what it is able to 'see', or to operate upon, is the identity and arrangement of its constituents. Nothing else." (2008, p. 107).

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

be seen as a crucial ingredient in a correct account of cognitive capacities? This is not the general problem of the causal efficacy of semantic content, but a more specific worry related to the particular answer input theories give to this problem. One might agree that the problem of the causal efficacy of semantic content could be "solved by the proposal that [mental representations] are physical entities with semantic properties lining up with their syntactic properties" (Smortchkova, Dołęga, & Schlicht, 2020, p. 12). But the question remains: How is this "lining up" produced? Does the semantic content of mental representations have something to do with its establishment? The mere existence of an "alignment" cannot suffice if there is not a clear explanation of this fact, an explanation in which the semantic content of mental representations plays a substantive role.

These "how possible questions" (Hutto & Myin, 2020) are examples of typical problems threatening input theories of mental content. Output theories, in contrast, do not suffer from these kinds of difficulties. There is no need for output theorists to explain how physical manipulations that reliably respect the semantic content of mental representations could possibly be established; rather, according to output theorists, it is by being causally manipulated in cognitive activity that some internal states acquire semantic content, and thus can be considered mental representations. That does not mean, of course, that output theories do not suffer from different, but equally serious, difficulties. The aim of this paper is to explore some of these difficulties. In particular, in the next section, I will formulate an objection that can be characterized as a kind of epiphenomenalist problem. Put simply, the problem is that, given the way output theorists address the two central issues raised in the RTM framework, mental representations must be seen as causally irrelevant by-products of the physical processes that underlie cognition.

The epiphenomenalist argument

According to output theorists, as we have seen, what fixes the content of mental representations is to be found, not in the causes of these internal states, but in their effects: It is by having their particular effects in cognitive activity that some internal states acquire

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

semantic content (get their semantic content fixed) and thus can be considered mental representations. However, given this way of understanding how semantic content is fixed, and so how underived intentionality is produced, we can ask: Why should we assume that mental representations, as representational states, are causally relevant to the physical processes that underlie cognition, instead of saying that they are only a causally irrelevant result of these processes? If it is by having a particular causal role in cognitive activity that some internal states acquire semantic content, their having such content cannot be taken as one of the factors that explain their having that particular causal role. Their possession of semantic content is rather an effect of their causal role and can be explained by it. But if the causal role of internal states is what explains their possession of semantic content, and not the other way around, why should we suppose that internal states with semantic content, as representational states, are causally relevant to cognitive activity? Given that semantic content gets fixed only once internal states are already fulfilling their causal role in cognitive activity, it seems that mental representations as such are epiphenomenal, a mere by-product of the physical processes that underlie cognition. This is the epiphenomenalist argument against output theories of mental content.

The problem can be illustrated with Millikan's teleosemantic theory. We have seen that, according to this theory, mental representations are internal states whose structure maps the structure of states of affairs in the world. Depending on the mapping function considered, however, different states of affairs can be mapped by the same internal state. A specific mapping function determining a unique state of affairs to be mapped by an internal state can be picked out only insofar as this internal state is being used by a device having a particular proper function. Once internal states are so used, and only once they are so used, they acquire semantic content and can be considered mental representations. But then it seems that mental representations as such cannot be a causal contributor to the workings of the devices that use them in cognitive processes. Mental representations are rather a result of the activity of these devices and cannot, therefore, causally explain it. They are epiphenomenal.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

Some clarifications are necessary. First, the problem is not that mental representations are epiphenomenal because semantic content is abstract and cannot, as such, be a cause of anything (Egan, 2020a). We may agree that if mental representations are naturalized in terms of certain nonintentional phenomena, "the explanatory purchase of the representational posit will in the most basic case be the explanatory purchase of those nonintentional phenomena" (Neander, 2017, p. 85). However, the central idea of output theories is that it is because the nonintentional phenomena involved in cognitive activity have their explanatory purchase (their particular causal role) that semantic content gets fixed. Those nonintentional phenomena explain cognitive activity and also explain how mental representations acquire semantic content, but mental representations as such have no explanatory role to play. So, the reason why mental representations are epiphenomenal is not because semantic content is abstract. According to the epiphenomenalist argument, mental representations are epiphenomenal because they are a mere by-product of the nonintentional phenomena that underlie cognition.

The epiphenomenalist argument must also be dissociated from one of the major objections to teleosemantic accounts: the content-indeterminacy problem (Martínez, 2013). According to this problem, teleosemantic accounts are unable to assign unique contents to mental representations. Strategies to solve this problem have been put forward, such as appealing to homeostatic property clusters (Martínez, 2013) or to explanations of the co-occurrence of sets of properties (Artiga, 2021). Other authors have defended the idea that indeterminacy is a real and unproblematic feature of mental representations (Bergman, 2023). However, according to the epiphenomenalist argument, the problem is not that mental representations are epiphenomenal because their content is indeterminate. Even if we accept any of the proposed solutions to the content-indeterminacy problem, or the idea that indeterminacy is not a problem at all, mental representations remain epiphenomenal. They are a mere by-product of the physical processes that underlie cognition.

The epiphenomenalist argument is closely related to the problem of unexploited content, pointed out by Cummins and his collaborators (Cummins *et.al*., 2006). This problem

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

is formulated as a specific worry for teleosemantic accounts, but it concerns output theories of mental content in general. As we have seen, for output theorists, the content of mental representations is fixed by their effects, that is, by the particular way those internal states are used in the system. Before being used or exploited, internal states have no semantic content (their content is not fixed), and thus cannot be considered mental representations. As Cummins and his collaborators note, however, it is presupposed by all neural network models of learning that the brain can learn to exploit its previously unexploited structures. This learning consists in the adjustment of synaptic weights so as to properly respond to patterns of input activation. But what does it mean here to properly respond to such patterns? This seems to mean, at least according to Cummins and his collaborators, that the adjustment responds to the specific content of the patterns in question. So, contrary to what output theorists claim, internal states seem to have determinate contents before they are exploited in cognitive processes. That is, there seems to be unexploited content.

In answering the output theorists' riposte that neural networks create the content of their input patterns as learning progresses, Cummins and his collaborators say that, if that were true, there would be no point in counting early responses in the process of adjustment of synaptic weights as errors. "And if early responses are not errors, why change the weights in any particular direction? Indeed, why change them at all?" (Cummins *et.al*., 2006, p. 197, footnote 2). If there is unexploited content, it seems that output theories are not really addressing the problem of underived intentionality. This problem would be that of explaining how mental representations acquire their unexploited content, a problem that output theories are not designed to solve. But why not simply reject the idea that unexploited content is really content? Cummins and his collaborators say that what they call content—including unexploited content—is "what ultimately does the work in representationalist cognitive science" (2006, pp. 205-206). According to them, unexploited content is simply presupposed by any account of cognition that assumes that representational capacities are learned or evolved.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

From the point of view of output theorists, there cannot be unexploited content. Output theorists must certainly claim that internal states with semantic content are created as new cognitive capacities evolve or are learned. For these theorists, use antedates content, and not the other way around. This idea implies, however, that internal states with semantic content, as representational states, cannot be explanatory relevant to the physical processes that underlie cognition. These processes explain cognitive capacities and also explain how mental representations acquire semantic content, but mental representations as such have no explanatory role to play. This is, of course, the epiphenomenalist argument. But Cummins and his collaborators have a different view on this issue.

It seems right to claim, as Cummins and his collaborators do, that early responses in the process of adjustment of synaptic weights must count as errors. Otherwise, there would be no point in calling this process an adjustment process. But Cummins and his collaborators also seem to think that, once we see these early responses as errors, we must admit that what the adjustment is responding to is the semantic content of input patterns, and so that these patterns are representations. Hence, they conclude that there must be unexploited content. This is not, however, an obvious inferential step.

The aim of theories of cognition, including neural network models, is to explain cognitive capacities and, as Egan points out, "[the] characterization of a phenomenon as a *capacity* or *competence* is itself normative" (2014, p. 129). That is, we see a capacity as something that is supposed to bring about a particular outcome. When the outcome is produced (in the way it is supposed to be produced), we consider that the capacity has been successfully exercised. Otherwise, the exercise of the capacity is considered unsuccessful, or its outputs mistaken. Egan also points out that this normative characterization is "given pre-theoretically, prior to the theory's characterization of the mechanisms underlying the competence", and so that normative elements "are there from the beginning" (2014, p. 129). Thus, what theories of cognition aim to explain is something that we pre-theoretically see as a normative phenomenon. But pure causal explanations seem unable to account for normative phenomena. Something else must be added to those explanations to make them

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

work. Given that the notion of representation has itself normative characteristics, mental representations seem especially well suited to complementing causal explanations of cognitive capacities.

However, if the critical explanandum is normativity, different possibilities could be explored. There is no reason to see mental representation as the unique possibility. For instance, ways of explaining normativity can be combined with computational approaches without positing mental representations. That is the case of teleomechanistic views of computation (Coelho Mollo, 2021). These views can be useful, not for explaining mental representation, but for directly explaining the normativity of cognition. In any case, in order to constitute a viable option, the explanatory benefits of positing mental representations must outweigh its costs. The epiphenomenalist argument tries to show that there might be no explanatory benefits associated with the postulation of mental representations, at least if we understand mental representations as output theories do. In the next section, I will try to show how the epiphenomenalist argument applies to more recent output theories of mental content. I will discuss two proposals: Shea's Varitel Semantics (2018) and Gładziejewski and Miłkowski's account of S-representations (2017). In choosing these proposals, I am not at all suggesting that they are particularly weak or unconvincing. Quite the contrary, both proposals are very detailed and very influential accounts of mental representation, and so it is extremely fruitful to discuss them. This is the main reason for my choice. My aim is to show that, despite their theoretical innovations, they are vulnerable to the epiphenomenalist argument.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

The epiphenomenalist argument applied to two recent proposals

Varitel Semantics

According to Nicholas Shea, mental representations are "physical particulars which interact causally in virtue of non-semantic properties (e.g., their physical form) in ways that are faithful to their semantic content" (2018, p. 31). This characterization seems to agree with input theories: Semantic content antedates the use of mental representations and the causal processes in which mental representations are involved must somehow be respectful of this content. However, Shea's proposal is an output theory of mental content. To see this, we must consider how, according to this proposal, semantic content is fixed.

On Shea's view, there are two kinds of relations internal states can bear to those things in the world they represent: correlations and structural correspondence. In virtue of the fact that properties of internal states correlate with properties of things in the world, internal states carry correlational information about those external things. Shea defines correlational information in the following way: "a being in state F carries *correlational information* about b being in state G iff $P(Gb|Fa) \neq P(Gb)$" (2018, p. 76). Structural correspondence, on the other hand, "is a mapping under which relations are preserved" (2018, p. 116). According to Shea, there is a structural correspondence between relation V on putative representational vehicles v_m and relation H on worldly entities x_n "iff there is a function f which maps the v_m onto the x_n and $\forall i,j\ V(v_i, v_j) \leftrightarrow H(f(v_i), f(v_j))$ (*mutatis mutandis* for other polyadicities)" (2018, p. 117).

Correlations and structural correspondence are "exploitable relations". On Shea's view, exploitable relations are extrinsic properties of internal vehicles that explain how the implementation of an algorithm involving sequences of operations on those vehicles can lead to the fulfillment of task functions. Task functions are distally characterized outcomes of a system that are robust and have been stabilized. For Shea, robust outcomes are those produced "in response to a range of different inputs" and "in a range of different relevant external conditions" (2018, p. 55); and stabilized outcomes are those that, because they lead to good consequences, have been retained by means of evolutionary success, learning, or

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

contribution to the persistence of the organism (2018, p. 64). Examples of robust and stabilized outcomes, and so examples of task functions, are animal signaling, navigational skills, object reaching, face recognition, etc. We have now all the main components of Shea's account, which he calls "Varitel Semantics":

> There are internal components which stand in exploitable relations to aspects of the environment that are relevant to achieving an outcome (a task function), where an internal process performed over vehicles with those properties constitutes an algorithm for achieving the distally characterized outcome successfully in a context-sensitive way. (2018, p. 51).

But how does exactly semantic content get fixed on this account? Exploitable relations by themselves cannot fix semantic content. An internal vehicle being in a state F carries correlational information about a lot of different things; and, given a relation V on internal vehicles v_m and practically any relation H on things x_n in the world, it is always possible to find a mapping function that enables us to establish a structural correspondence between V and H. On Shea's account, "exploitable relations are the link between internal components and the distally characterized task function which the organism is performing" (2018, p. 36), but distally characterized task functions are the crucial factor that fixes semantic content. Without distally characterized task functions, there are no particular elements in the environment that can constitute the targets of mental representations. Given that task functions can be performed by different algorithms and that different algorithms may call for different contents, the specific algorithm implemented by the system for performing a task function is also one of the factors that fix semantic content. But again, without distally characterized task functions, there is no point in determining any particular algorithm. Even if exploitable relations and algorithms contribute to fixing content, both of them crucially depend on task functions, which constitute the main factor.

According to Shea's account, it is because task functions are performed that semantic content gets fixed. On this view, (1) robust and stabilized outcomes (task functions) being produced by (2) an algorithm involving operations on (3) internal components which stand

Revista Sincronía
ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Biannual Journal of Philosophy, Literary Studies and Humanities

in exploitable relations to features of the environment form a natural cluster, and this cluster is what "constitutes the internal components as being representations" (2018, p. 51). Without being used in this way for the performance of task functions, internal components bearing exploitable relations to things in the world cannot get their semantic content fixed. Once they are so used, and only once they are so used, they acquire semantic content and can be considered mental representations. Shea's proposal is, then, an output theory of mental content. On this view, use antedates content, and not the other way around. Consequently, the epiphenomenalist argument applies to Shea's account: Sequences of operations on internal components which stand in exploitable relations to things in the world explain the performance of task functions and also explain how mental representations acquire semantic content, but mental representations as such have no explanatory role to play. Mental representations are rather an effect of this activity and cannot, therefore, causally explain it.

The epiphenomenalist argument, as it applies to Shea's account, must be distinguished from an important objection made by Frances Egan (2020b). The objection concerns the fact that the appeal to mental representations in Shea's explanation of the performance of task functions seems to be dispensable. According to Egan, a full explanation of the performance of distally characterized task functions must mention correlations between world states and internal components of the system, and between those components and outcomes in the world. Egan argues, however, that nothing essential is added to this explanation by saying that the internal components are representations: This representational talk is a convenient "gloss" allowing us to draw attention to the relevant correlations, "but content attribution is not essential to the explanation of the organism's success" (2020b, p. 374). Shea replies that characterizing internal components as representations is important to understand "why some robustly produced outcomes rather than others should count as successes" (2020, p. 406).

In my view, Egan is right, but the problem she is pointing out is different from the problem highlighted by the epiphenomenalist argument. We can admit, for the sake of

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

argument, that the processes and relational properties described by Shea give rise to real representations.[2] That is, we can admit that Shea's representational talk is not a mere "intentional gloss" doing no more than characterizing "computational processes in ways congruent with our commonsense understanding of ourselves" (Egan, 2019, p. 256), but that the representations he is talking about are real internal components with semantic content. According to the epiphenomenalist argument, however, the problem is not whether output theories succeed in identifying real mental representations and in explaining how these real representations acquire semantic content. That can be admitted. The problem is that the way in which output theories are supposed to do this implies that mental representations are epiphenomenal. Mental representations appear only once (and only because) the physical processes that underlie the performance of task functions are already doing their job. Mental representations cannot, therefore, causally explain these processes. The representations Shea is talking about might well be real internal states with semantic content. But then we must conclude that, as such, they are a mere by-product of the physical processes that underlie cognition.

As an output theorist, Shea can reply that semantic content gets fixed as soon as the algorithms performing task functions are implemented. However, that does not help. When internal vehicles acquire semantic content, the physical processes that underlie the performance of task functions are already at work. Mental representations are, thus, no more than a causally irrelevant result of these processes.

Gładziejewski and Miłkowski's S-representations

According to Gładziejewski and Miłkowski (2017), structural mental representations (or S-representations) are components of cognitive systems that represent their targets in virtue of the structural similarity they bear to them (see also Gładziejewski, 2016a). Structural similarity is a second-order kind of resemblance (O'Brien & Opie, 2004; see also

[2] For serious doubts about the actual implementation of these processes and relational properties in the brain, even in the cases Shea himself takes as examples, see (Burnston, 2021).

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

Gładziejewski, 2016b). In contrast to first-order resemblance, which implies the sharing of physical properties, second-order resemblance is about relations mirroring relations. Gładziejewski and Miłkowski adopt O'Brien and Opie's characterization of second-order resemblance. Consider a system $S_V = (V, \mathfrak{R}_V)$ comprising a set V of objects (which can be conceptual or concrete) and a set $\mathfrak{R}$ of relations (which can be spatial, causal, inferential, etc.) defined on the members of V. According to O'Brien and Opie, there is a second-order resemblance between $S_V = (V, \mathfrak{R}_V)$ and another system $S_O = (O, \mathfrak{R}_O)$ if the following condition obtains:

> […] for at least *some* objects in V and *some* relations in $\mathfrak{R}_V$ there is a one-to-one mapping from V to O and a one-to-one mapping from $\mathfrak{R}_V$ to $\mathfrak{R}_O$ such that when a relation in $\mathfrak{R}_V$ holds of objects in V, the corresponding relation in $\mathfrak{R}_O$ holds of the corresponding objects in O (2004, p. 11).

As Gładziejewski and Miłkowski rightly point out, structural similarity is not sufficient to confer on internal components the status of mental representations. Structural similarity is "cheap": S-representations are structurally similar to their targets, but also to a lot of different (and irrelevant) things, and there can be relations of structural similarity between external entities and components of the system that are not S-representations. Inspired by Shea's work (2014), Gładziejewski and Miłkowski hold that S-representations come into play only insofar as the engagements of cognitive systems with things in the world depend, in a nontrivial sense, on the structural similarity between vehicles of S-representations and those external things. Structural similarity, they claim, "should be understood as a relation that is *exploitable* for some large representation-using system" (2017, p. 340).

But what does it mean exactly for structural similarity to be exploitable? Gładziejewski and Miłkowski see S-representations as components of cognitive mechanisms. They adopt a neomechanistic explanation of cognition (Bechtel, 2008; Boone & Piccinini, 2016; Craver, 2007; Miłkowski, 2013) according to which the cognitive system is a collection of mechanisms, and a mechanism is a "set of organized components and component

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

operations which jointly enable the larger system to exhibit a certain phenomenon", which is often understood as a capacity of the system (Gładziejewski and Miłkowski, 2017, p. 341). Mechanisms are thus individuated, at least partly, by reference to the function they perform. As Gładziejewski and Miłkowski put it, "they are essentially mechanisms *of* this or that cognitive function (mindreading, motor control, attention, perceptual categorization, spatial navigation, etc.)" (2017, p. 341). Components of mechanisms have also functions, which derive from the functions of their mechanisms: They are the set of operations those components perform that contribute to the fulfilment of their mechanism's function. So, given that S-representations are components of cognitive mechanisms, their function must derive from the function of their mechanisms. That essentially means, according to Gładziejewski and Miłkowski, "that *structural similarity* between the representation and what it represents is what contributes toward the mechanism's proper functioning" (2017, p. 341). It is in virtue of this contribution to cognitive activity that structural similarity can be considered an exploitable relation.

As we have seen, however, structural similarity by itself cannot fix semantic content. How are then determined the targets of S-representations on Gładziejewski and Miłkowski's account of mental content?[3] Given the neomechanistic framework adopted by these authors, the answer seems clear: The targets of S-representations are those things in the world S-representations would have to be structurally similar to in order to contribute to the fulfilment of their mechanism's function. It is thus the function of the mechanisms S-representations are embedded in that determines the targets of these representations. In other words, what fixes the content of S-representations is to be found, not in the causes of these internal components, but in their effects, that is, in their causal contribution to the fulfilment of their mechanism's function. On Gładziejewski and Miłkowski's (implicit) account, use antedates content: It is by being exploited in cognitive activity that vehicles of

[3] To be fair, this is not a question that Gładziejewski and Miłkowski address in their paper. However, it seems to me that the ideas they discuss and defend presuppose an account of how semantic content gets fixed in the specific case of S-representations. My claim is that this account is a kind of output theory of mental content and one of my purposes in this subsection is to try to make it explicit.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

S-representations acquire semantic content, and thus can be considered mental representations. Once S-representations, as components of cognitive mechanisms, are used in the performance of these mechanisms' functions, and only once they are so used, there can be targets (things in the world) S-representations, not only are, but are supposed to be structurally similar to. So, as Shea's Varitel Semantics, Gładziejewski and Miłkowski's account of S-representations is an output theory of mental content. The epiphenomenalist argument applies, therefore, to this account: Causal operations on components of cognitive mechanisms explain how these mechanisms fulfil their functions, and also explain how some of these components become S-representations (how there can be targets these components, not only are, but are supposed to be structurally similar to), but S-representations as such have no explanatory role to play. S-representations acquire semantic content only once the mechanisms that exploit them in cognitive activity are already fulfilling their functions and so they cannot, as representational states, causally contribute to the activity of these mechanisms. S-representations are, thus, epiphenomenal.

But before concluding this, we must consider the fact that Gładziejewski and Miłkowski's main purpose in their paper is to show how structural similarity as such can be causally relevant to the success of the mechanisms that exploit S-representations. In order to show this, Gładziejewski and Miłkowski adopt Woodward's interventionist theory of causal relevance (2003, 2021). According to this theory, if a variable *C* (the putative cause) is causally relevant to a variable *E* (the effect), then it is true that "if we (or nature) were able to manipulate *C* in the right way, there would be some associated change in *E*" (Woodward, 2021, p. 76). More precisely, *C* causes *E* in background circumstances B iff:

> (i) there is some possible intervention that changes the value of *C* such that (ii) if that intervention were to occur in B, there would be an associated change in the value of *E* or in the probability distribution *P(E)* of those values (Woodward, 2021, p. 77).

Gładziejewski and Miłkowski use the interventionist framework to explain the causal relevance of structural similarity in the following way. The values of *C* correspond to degrees

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

of structural similarity (within a certain range) between S-representations and their targets, and the values of *E* correspond to degrees of success of the mechanisms that exploit these representations in performing their functions. For structural similarity to be causally relevant to the success of cognitive mechanisms is thus simply for interventions that change the value of *C* (in certain background circumstances) to result in systematic changes in the value of *E*. In particular, by increasing the value of *C* we must increase the value of *E*, and by decreasing the value of *C* we must decrease the value of *E*.

As Gładziejewski and Miłkowski point out, however, there is a problem with this way of understanding the causal relevance of structural similarity. The problem is that, in order to perform an intervention on structural similarity, we must necessarily intervene in the structure of at least one of its relata, that is, we must intervene in the structure of S-representations or in the structure of their targets. Given this situation, it seems much more parsimonious to say that what is causally relevant to the success of cognitive mechanisms is not structural similarity as such, but rather the structure of S-representational vehicles and/or the structure of their targets. To replay to this objection, Gładziejewski and Miłkowski propose to distinguish between "interventions that change the way some cognitive system *acts* (behaviorally or cognitively) and interventions that change the *success* of its actions" (2017, p. 345). They note that changes in the structure of vehicles can change the way cognitive mechanisms act, but do not necessarily affect success if the resultant changes in action are accompanied by appropriate changes in the environment. According to them, these appropriate changes in the environment preserve success by restoring the structural fit between S-representations and their targets. Gładziejewski and Miłkowski hold that it is impossible to say how interventions in the structure of vehicles would affect the success of cognitive mechanisms independently of facts about the targets, "or more precisely, independently of the facts regarding structural similarity between the vehicle and the target" (2017, p. 356). They conclude from this that "interventions on the vehicle's structure change the success *only insofar as they change the degree of similarity between the vehicle and the target*" (2017, p. 356), and so that structural similarity as such is causally relevant to success.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

However, consider a bare component of a cognitive mechanism, that is, a component that is not the vehicle of an S-representation. We suppose that such components exist, for we assume that not every component of a cognitive mechanism is a representational vehicle. If we perform interventions on the structure of this component, we can expect to observe changes in the actions of the mechanism. That is, we can assume that the component has a causal role in the workings of the mechanism and that interventions in its structure can affect this causal role and produce changes in the way the mechanism acts. As we have seen, changes in action can affect success, but not necessarily, if they are accompanied by appropriate changes in the environment. We can thus identify robust and systematic correlations between modifications in the bare component's structure that affect success and modifications in the environment that restore it, and we can establish a structural correspondence between the bare component and those external conditions. This structural correspondence can be seen as a second-order resemblance, and so as sort of structural similarity. By identifying these correlations and establishing this second-order resemblance, however, we are not showing that the structural similarity between the bare component and the external conditions in question is causally relevant to success. What we have shown is a much more modest thing: that the success of cognitive mechanisms in performing their functions systematically depends on the external world. This is true of any mechanism, even of mechanisms performing non-cognitive functions.

Now, is there a fundamental difference between such a bare component and an S-representational vehicle concerning their causal significance? We cannot simply say that the S-representational vehicle really has representational properties, for the point is to try to elucidate what the possession of these properties amounts to in terms of causal influence. Both the bare component and the S-representational vehicle are structurally similar to external conditions relevant to the success of the mechanism. That is, in both cases, changes in structural similarity affect success. The crucial difference between the bare component and the S-representational vehicle is that, while the bare component is only structurally similar to external conditions relevant to success (structural similarity is cheap), the S-

Revista Sincronía
ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Biannual Journal of Philosophy, Literary Studies and Humanities

representational vehicle is, additionally, *supposed* to be structurally similar to these conditions. However, does this difference have a real causal significance? Gładziejewski and Miłkowski's proposal does not answer this question.

It is only once a cognitive mechanism is already fulfilling its function that external conditions relevant to its success actually exist. It is thus only once a cognitive mechanism is already doing its job that a second-order resemblance between those conditions and components of the mechanism can be established. We can see these components as S-representations and the external conditions as their targets. By doing so, however, we are not showing that the structural similarity between S-representations and targets is causally relevant to the mechanism's success, even if we observe that increasing structural similarity increases success, and decreasing it decreases success. This kind of correlation only reflects the fact that the success of cognitive mechanisms systematically depends on the external world. Given that the structural similarity in question can be established only insofar as the mechanism is already fulfilling its function, this structural similarity cannot be a causal contributor to success. It is rather the fact that there already is a cognitive mechanism fulfilling its function that makes it possible to establish a structural similarity between components of the mechanism and external conditions relevant to success, and not this structural similarity that causally explains the success of the mechanism in doing its work. As we can see, this is only a different way of formulating the epiphenomenalist argument. Therefore, we can finally conclude that the epiphenomenalist argument applies to Gładziejewski and Miłkowski's proposal: S-representations as such are only a causally irrelevant by-product of the activity of cognitive mechanisms.

Conclusion

If the way internal states are used in cognitive activity explains content, then content cannot explain the use of those internal states, and so it cannot explain their causal role in cognitive activity. This is the core idea behind the epiphenomenalist argument against output theories of mental content. The argument does not have a mere negative purpose though, for my aim

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

is to make a little contribution to the consolidation of a logical space for new possibilities in our understanding of cognition. These new possibilities are already being explored and my hope is that the discussion in this paper could help us see them as really important options.

Referencias

Artiga, M. (2021). Beyond Black Dots and Nutritious Things: A Solution to the Indeterminacy Problem. *Mind & Language*, *36*(3), 471-490.

Bechtel, W. (2008). *Mental Mechanisms.* New York: Routledge.

Bergman, K. (2023). Should the Teleosemanticist Be Afraid of Semantic Indeterminacy? *Mind & Language*, *38*(1), 296-314.

Boone, W. & Gualtiero P. (2016). The Cognitive Neuroscience Revolution. *Synthese*. (193) 1509-1534.

Burnston, D. (2021). Contents, Vehicles, and Complex Data Analysis in Neuroscience. *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*. MIT Press.

Cummins, R., Blackmon, J., Byrd, D., Lee, A., & Roth, M. (2006). Representation and Unexploited Content. In 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). How to Think about Mental Content. *Philosophical Studies*. *170*(1), 115-135.

Egan, F. (2019). The Nature and Function of Content in Computational Models. In 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. In J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *What are Mental Representations?* (pp. 26-53). Oxford University Press.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

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 University Press.

Fodor, J. (2008). *LOT 2. The Language of Thought Revisited*. Oxford University Press.

Gładziejewski, P. (2016a). Action Guidance is not Enough, Representations Need Correspondence too: A Plea for a Two-factor Theory of Representations. *New Ideas in Psychology*. (40), 13-25.

Gładziejewski, P. (2016b). Predictive Coding and Representationalism. *Synthese*. (193) 559-582.

Gładziejewski, P., & Miłkowski, M. (2017). Structural Representations: Causally Relevant and Different from Detectors. *Biology and Philosophy*. (32) 337-355.

Godfrey-Smith, P. (1994). A Continuum of Semantic Optimism. In S. Stich & T. Warfield (Eds.), *Mental Representation* (pp. 259-277). Blackwell.

Hutto, D. D., & Myin, E. (2020). Deflating Deflationism about Mental Representation. In J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *What are Mental Representations?* (pp. 79-100). Oxford University Press.

Martínez, M. (2013). Teleosemantics and Indeterminacy. *Dialectica*, *67*(4), 427-453.

Miłkowski, M. (2013). *Explaining the Computational Mind*. MIT Press.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, *86*(6), 281-297.

Millikan, R. G. (1995). Pushmi-Pullyu Representations. *Philosophical Perspectives*. (9) 185-200.

Millikan, R. G. (2017). *Beyond Concepts*. Oxford University Press.

Neander, K. (2017). *A Mark of the Mental.* MIT Press.

O'Brien, G., & Opie, J. (2004). Notes Towards a Structuralist Theory of Mental Representations. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in Mind* (pp. 1-20). Elsevier.

Papineau, D. (1999). Normativity and judgment. *Proceedings of the Aristotelian Society, Supplementary Volumes*. (73) 17-43.

ISSN: 1562-384X
Year XXIX / Issue 88. July-December 2025
DOI: 10.32870/sincronia.axxix.n88
Revista Sincronía
Biannual Journal of Philosophy, Literary Studies and Humanities

Shea, N. (2013). Millikan's Isomorphism Requirement. In D. Ryder, J. Kingsbury & K. Williford (Eds.), *Millikan and Her Critics* (pp. 63-80). Wiley-Blackwell.

Shea, N. (2014). Exploitable Isomorphism and Structural Representation. *Proceedings of the Aristotelian Society*. (114) 123-144.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.

Shea, N. (2020). Representation in Cognitive Sciences: Replies. *Mind & Language*, *35*(3), 402-412.

Smortchkova, J., Dołęga, K., & Schlicht, T. (2020). Introduction. In J. Smortchkova, K. Dołęga, & T. Schlicht (Eds.), *What are Mental Representations?* (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