https://doi.org/10.17163/soph.n31.2021.10

The explanatory function of the notion of internal representation

La función explicativa de la noción de representación interna

Fabián Bernache Maldonado[*]
Universidad de Guadalajara, Guadalajara, México
fabian.bernache@academicos.udg.mx
Orcid Number: https://orcid.org/0000-0001-7158-892X

Abstract

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

Keywords

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

Suggested citation: Bernache, Fabián (2021). The explanatory function of the notion of internal representation. *Sophia, colección de Filosofía de la Educación, 31*, pp. 247-269.

[*] Professor in the Department of Philosophy of Universidad de Guadalajara. Has a PhD in Philosophy and Cognitive Science from Jean Nicod Institute, Paris, France. Member of the National System of Researchers of the National Council of Science and Technology of Mexico. His research interests focus on the notion of internal representation and philosophical and empirical theories about reasoning and argumentation.

Resumen

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

Palabras clave

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

Introduction

According to the Representational Theory of Mind (RTM), the cognitive life of an organism consists essentially on the formation and storage of representations and the application of operations to such representations, according to authors such as Sterelny (1990), Fodor (1998) and Fodor and Phylysyn (2015). Given an individual or an A category, the ability to represent A consists on the ability to form—in the mind or in the brain— representations of A. Thinking of A in a *t* moment, is to form a representation of A in a *t* moment, in working memory. To retain information about A is to preserve, in long-term memory, certain propositional representations that include representations of A. To believe that there is an A in my pocket is to be willing to employing, in specific circumstances and processes, an internal representation whose content is ‘there is an A in my pocket’ in a certain way that systematically differs from the way in which this same representation (or another with the same content) would be employed if, instead of believing, I would like an A in my pocket.

Beyond the difficulties presented when these ideas are discussed in detail, one of the great attractions of RTM is the possibility that it offers to reduce the philosophical cognition problem to a single question: what ultimately explains something to be a representation? As noted above, for RTM supporters, representational capabilities derive essentially from the power to form and process certain representations (internal or mental). Therefore, mechanisms or properties that ultimately explain that something is a representation cannot, from this perspective, depend on the prior possession of representational capabilities. How are these mechanisms or properties based on? According to a classical conception of representation, the similarity between ideas and objects is what explains something as a representation; on the one hand, that similarity connects ideas with their objects and, on the other, ideas are original representations, i.e., representations from which the other types of representations derive, particularly linguistic representations, as Fodor and Lepore said (1991).

The similarity between ideas and objects seems to satisfy the requirements of RTM, because it is not assumed that the existence of such similarity implies that the organisms already possess representational capabilities. Hence, it is still possible to find in some contemporary authors, such as Cummins (2010) and Johnson-Laird (2006), some adherence to sophisticated forms of the classical conception of representation. Classical conception, however, has its difficulties. For example, given that similarity allows degrees, what degree of similarity must exist between an idea and an object so that the idea can be considered a representation of the object? In addition, the similarity ratio is symmetrical: If A is similar to B, B is similar to A. Therefore, as Fodor (1984) has pointed out, if an idea is a representation of an object by virtue of the similarity that exists between both, the object would also have to be considered a representation of the idea.

But there are other influential proposals that do not use the notion of similarity. One of them is the informational semantics of Fred Dreske (1981). According to informational semantics, nomological relationships between the instances of R and the instances of A explain that a type R event represents a type A event; therefore, due to such relationships, R instances are capable of taking information from A. Imagine, for example, that R is the activation of a neuronal structure that cannot (nominologically) occur without a type A event, through some sensory stimulation causing it. Given this situation, we can state that R is an indicator of A, i.e., that R takes the information according to which A has occurred. According to informational semantics, under this relationship between R and A, R can be considered a representation of A.

Informational semantics face important problems like the classical conception of representation. The pivotal is the problem of error, pointed out by Fodor (1984): while representations can be true or wrong, nomological relationships may or may not exist, but they cannot be wrong.

Therefore, it does not seem to be true that by mere nomological relations between R and A, R can be considered a representation of A.

Several complementary notions have been introduced to reinforce informational semantics and give a solution to the problem of error, structure (Dretske, 1988), asymmetric causal dependence (Fodor, 1990) or incipient cause notions (Prinz, 2002). However, the aim of this work is not to discuss these or different proposals, but to make a general objection to the RTM. This objection, presented as a dilemma, is directed at its central thesis: the idea that the explanatory cognition foundation is the notion of internal representation.

Various objections have already been raised to RTM, some of which have been considered major challenges by representationalist theorists, such as Godfrey-Smith (2006). Among the most discussed recent objections is the so-called Hard Problem of Content raised by Hutto and Myin (2013). This problem is also presented as a dilemma. Hutto and Myin argue, on the one hand, that the conceptions of information acceptable from a naturalistic perspective are insufficient to support a true notion of internal representation and, on the other, that the information concepts that would allow a true notion of internal representation to be founded are unacceptable from a naturalistic perspective. Thus, according to these authors, in the current state of research on the notion of internal representation, either naturalism is preserved and RTM is abandoned, or RTM is preserved and naturalism is abandoned. The problem here is different and even more radical, because it does not depend on the acceptance of naturalism, nor is it based on the details of existing conceptions of internal representation. What is intended is to show that the notion of internal representation, regardless the way it is founded, is not able to fulfill the explanatory function attributed to it by the supporters of the RTM. Section 1 of this paper will present the dilemma posed to the supporters of the RTM and will discuss the consequences of each of the two responses presented. Two examples of key RTM representatives will be presented in section 2 to illustrate these difficulties. These discussions will be the basis of the conclusion of the article, namely: the notion of internal representation is uncapable of satisfying the explanatory function assigned to it by RTM itself.

The dilemma of internal representation

As the supporters of the RTM mention, the power to form and manipulate internal representations is what, fundamentally, explains the possession of representational capabilities. How do such representations fulfill their various functions in cognitive activity? In the framework of RTM, explanatory value from the notion of internal representation is not reduced to the mere intelligibility that results from the description of the cognitive processes of an organism in terms of manipulation of internal representations, but it is based on the capacity these representations would possess, as truly existing entities to influence causally on such processes

While the notion of internal representation can be seen as a component of an explanatory cognition model inspired by the functioning of public representations, an internal representation for RTM supporters is not merely an explanatory instrument without ontological implications, but an entity placed in the order of causal relations that is able to participate in them, as Ramsey (2007), Fodor and Phylysyn (2015), Neander (2017) have pointed out.

Given this conception of internal representations, one wonders: what properties exert on an internal representation its causal influence on cognitive activity? Two answers for this question seem possible: either an internal representation has causal power by virtue of its representational properties, or its non-representational properties derive the power[1]. Both options lead to serious difficulties for RTM. The explanation of these properties will be presented below.

First part of the dilemma

Suppose that R is a representation of A formed in the nervous system of an O organism and that the causal influence of R on the cognitive activity of O is possible thanks to the representational properties of R. Thus, in accordance with this approach, it is fundamentally from the fact, as such, that R represents A which derives the causal power of R in the cognitive activity of O.

But how can such a causal power derive from a semantic fact? In other words, how can a semantic fact be causally related, as a semantic fact, to other kinds of facts?

A simple answer is: through a process of interpretation and understanding. In order to be able to admit that that R causally influences the cognitive activity of O due to its representational properties, and not to other properties, it seems necessary to assume that the nervous system of O somehow ‘understands’ R and that this understanding of R determines, at least a bit, the cognitive processes performed in O. Since R represents A, understanding R is nothing more than representing A thanks to R.

But how can the O nervous system do such a thing? With regard to the problem of understanding language representations, the supporters of RTM often have a clear answer: to understand an E statement belonging to a L language is to create an internal representation whose content is the same as the content conventionally associated with E in L and to associate such representation with E. However, what about the understanding of an internal representation? When the problem is to explain the understanding of internal representations, it is obvious that it is not possible to appeal to the formation of new internal representations, since it would only be postponed.

The point discussed is related to Dummett's famous warning (1993) that any theory of meaning (or representation) that claims to be satisfactory must be accompanied by a theory of understanding. To assume that R is a representation of A, regardless of its spatial-temporal location and material composition, is to assume that R is destined to be understood; i.e., that as a representation of A, R must be able to allow any interpreter to represent A. Hence, if the existence of internal representations is postulated, the existence of internal processes of understanding must also be postulated, otherwise it would not be coherent to assume that the entities whose existence is being postulated are authentic representations. But since such internal processes of understanding cannot consist of processes of producing new internal representations, how are they to be understood?

Understanding representations is something that certain organisms can do: For example, humans can understand the language they have learned in childhood. However, when the supporters of RTM talk about internal representations, they refer to representations to which no one has proper access, because they are neural structures, i.e., structures literally located within organisms and not within the inner space (metaphorically speaking) of their minds. And although an organism may have sensitive access to some of its internal processes, such as digestive processes, what happens in the nervous system is not part of internal processes to which an organism can sensitively access. Thus, if there is an understanding of internal representations, such an understanding must be carried out by an internal component of the organism and not by the organism as such. But how can a simple component of an organism possess the ability to understand representations?

An objection to these ideas would be to point out that talking about understanding representations makes sense only if it refers to the capabilities that an organism can possess and not to the capabilities of mere components of an organism. Assuming that an internal representation must be understood would be an error. However, if mentioning that understanding is inadequate in the case of internal representations, it can then be concluded that the use of the notion of internal representation by RTM supporters is a metaphorical use that would allow, perhaps, to form a simple and intelligible image of cognitive activity, but not to explain such activity under the criteria required by the RTM itself. The intelligibility of this image would derive from the fact that it uses linguistic communication as a model, the realization of which effectively implies the use of representations, but also of interpretation and comprehension capabilities. In order to contribute to the clarification of cognitive activity, as a notion of representation in a strict sense, the notion of internal representation should therefore be accompanied by a plausible notion of internal understanding. This latter notion, however, is no easy to understand because is intelligible independently of the other. The application of internal representations does not appear to be useful for the understanding of cognitive activity if it is assumed that their causal power derives from their representational properties.

However, an internal structure may have some kind of functioning that can legitimately be characterized as a representational functioning, even if such functioning does not presuppose the realization of internal understanding processes such as those mentioned. As Ramsey (2007) points out, it would be legitimate to assume that certain structures in an organism's nervous system are representations if it is somehow achieved to show that these structures function as representations, and therefore if it can be established that characterizing them is essential to understanding the processes in which they participate. But what role must an internal structure perform so that it can legitimately be assumed that such a structure functions as a representation? Strictly speaking, a R structure, event or object function as a representation of A, not only if it represents A, but also if it allows a system or body to represent A. But, in order for an A system or organism to be represented by R, it is essential that the system or organism be able to interpret and understand R. Thus, if this reasoning is correct, any application of internal structures that function as representations commits us to the existence of internal processes of interpretation and understanding of representations.

Likewise, Godfrey-Smith (2006) recognizes that RTM supporters cannot settle by internal representations, but must also admit the need to apply for internal 'readers' or 'interpreters' of such representations. For Godfrey-Smith, the activity of these 'readers' mechanisms must be recognized as an authentic interpretation activity, without being so sophisticated that can jeopardize the explanatory value of the application of internal representations (presupposing, in some way, the prior possession of representational capabilities).

But what kind of activity can satisfy such criteria? This problem is a central concern in the proposal of Garrett Millikan (1984, 2000, 2017), who bases his conception of internal representation not only on the specification of the relationship that must exist between R and A for R to represent A, but also in the type of processes to which R must be submitted and which would allow the organism in which R has been formed —or to an internal 'reader' mechanism—identify what R represents.

Millikan's proposal is complex and will be discussed in more detail in subsection 3.1, where attempts will be made to show its inadequacy. From the perspective of this article, it is wrong to assume that it is possible to understand the explanatory value of the notion of internal representation from the notion of interpretation, since neither of these notions is easier to integrate in causal explanation.

Moving from the notion of internal representation to the notion of interpretation is not a step forward, but rather it shows that it has reached a dead end. Thus, the conclusion that can be drawn from this first part of the dilemma (which will be reinforced later) is that the notion of internal representation seems incapable of fulfilling the explanatory function assigned to it by the RTM.

The second part of the dilemma

But it may be possible to avoid these problems if admitting that the causal power of an internal representation derives not from its representational properties, but from its non-representational properties. Such is the position explicitly assumed by Fodor (1995, 2008), but implicitly assumed by many authors who, like Mercier and Sperber (2017), argue that an internal representation is a concrete object that possesses, like any concrete object, causal powers[2]. Similarly, Neander (2017) adopts this position when stating that "if we naturalize mental representations in terms of certain unintended phenomena, the explanatory force of postulated representations will be the explanatory force of these unintentional phenomena in the most basic case" (p. 85)[3].

Therefore, assuming that R is a representation of A formed in the nervous system of an O organism and that it is due to its non-representational properties that R causally influences the cognitive activity of O. In this case, it does not seem necessary to postulate the existence of internal processes of interpretation and understanding of R, since it is not from the fact, as such, that R represents A from which derives the causal power of R, but from the properties that R possesses as mere neuronal structure. In other words, the question of how a semantic fact relates causally to other kinds of facts does not originate here, since causally related facts that constitute the cognitive activity of O would all be purely material facts. However, given such an approach, it is possible to ask: to what extent can it be argued that the non-representational properties of R exert their causal influence as non-representational properties of a representation and not as properties of a simple neuronal structure? In other words, why should it be assumed that the fact that non-representational properties of R are properties of a representation that must have some relevance to the causal influence they exert on the cognitive activity of O?

As mentioned, when RTM supporters talk about internal representations, they often assume, like Fodor and Pylyshyn (2015), that these are concrete objects or events, as "chalk marks on a chalkboard, ink marks on paper, uttered sentences, neuronal events, etc." (p. 7)[4]. Another common way of assuming this idea is observed in the distinction between mental representations (in the brain or in the mind) and public representations (statements or images); it is usually assumed that both groups of representations are composed of discrete objects that differ basically in their location and material composition. Note, therefore, that R is a discrete object, i.e., in this particular case, a neuronal N structure (or, if preferred, the activation of N), located as a more or less well delimited component (or event) on O nervous system. What relevance does, for the causal explanation of the cognitive activity of O, characterization of certain properties of N have as non-representational properties of a representation? The answer is: None.

The brain mass fragment identified as N may certainly possess causally relevant properties for the cognitive activity of O. However, characterizing these properties as non-representational properties of a representation does not allow understanding its role in the cognitive activity of O, for the fact that N is a representation (i.e., the fact that N is R) it has no explanatory value in that activity, since it has been rejected that the representational properties of N have a causal influence on the cognitive processes of O. This explanatory incapacity is a consequence of the fact that, on the one hand, the neuronal N structure has been delimited assuming that it is a representation, but, on the other, it has been assumed that the representational properties of N (i.e., the properties of

N as a representation) are not causally relevant. The neuronal N structure does not have authentic functional significance in the cognitive activity of O, nor its properties, at least to the extent that they are characterized as non-representational properties of a representation. Generally, acknowledging that a representation is a particular object (or event) with causal powers does not force us to assume that all the effects that this object (or event) produces are necessarily related to the fact that it is a representation. Thus, even if N is a representation, it cannot be concluded, without additional reasons, that the effects N produces are related to it.

The nervous system of O, like any space-temporal object, is subjected to various forces that can affect it nomologically in different ways. But also, in addition to being a physical object subjected to causal influences, like any physical object, the O nervous system has channels through which it can capture specific causal flows that affect its internal structure in a more precise and controlled way. However, these causal flows are not, by themselves, information flows. In order for internal causal flows to be transformed into authentic information flows, they need to be properly exploited by the O. nervous system to clarify this idea, to illustrate, tree's growth rings.

As known, the growth rings that form the trunk of trees are related with the age of trees, so that a system or organism with the required capacities could collect information about the age of a tree from these rings. This does not mean, however, that growth rings should be seen, by themselves, as information-transmitting transporters, as growth rings are simple physical structures resulting from a certain series of causal events. Another example is an animal footprint printed in the mud: by itself, a footprint does not transmit information, but is a simple physical structure that is also the product of a series of causal events. In order for such a structure to be transformed into a genuine transporter for transmitting information, there must be an organism or system capable of exploiting the nomological connections that exist between the structure and the events of the environment (past, present or future).

Evidently, such a form of exploitation does not necessarily imply understanding the information, as it is simply a process through which the information is used in order to perform a certain function of the operating body or system. In summary, there is no real flow of information, no transmitting transporter, no exploitation processes of nomological relationships; and there are no exploitation processes of nomological relationships without using these relationships to satisfy certain functions[5].

Thus, the causal flows in the nervous system of O, by themselves, are nothing more than causally connected and nomologically determined events. Assuming that these internal causal flows are information flows is to believe that the O nervous system, in addition, is structured in such a way that it is able to exploit the nomological relationships that exist between its internal causal flows and various external events in order to satisfy certain functions. These functions can be cognitive, such as visual or auditory perception, but can also have other functions, such as regulator of body temperature or heart rate.

Going back to R, i.e., a representation of A formed in the nervous system of O. R, as a concrete object (or event), is a neuronal N structure (or an activation of N). Is it relevant for the explanation of the cognitive activity of O to characterize the properties of N as non-representational properties of a representation? If the activity of the nervous system of O is seen as a simple causal flow, it is not possible to characterize N as a representation, nor its properties as non-representational properties of a representation, unless it is willing to accept the idea that the pure natural nomological order is sufficient to produce representations. But what if considering the activity of the nervous system of O, not as a simple internal causal flow, but as a true flow of information? Is it relevant, in this case, to characterize the properties of N as non-representational properties of a representation?

What precise function does N perform in the nervous system of O as a structured system? It has been assumed that N is a representation of A. The question is: Can this characterization lead to understand the function of N in the structured activity of the O nervous system? It is reasonable to assume that the properties of N play some, perhaps crucial, role in this activity. However, as noted above, it cannot simply be assumed that the importance of these properties in the performance of such a function — whatever it may be — is related to the fact that they are, as has been admitted, non-representational properties of a representation. For this, it is worth mentioning that it is relevant to the functioning of the nervous system of O to characterize certain neuronal structures as representations, even if the causal power of these structures derives exclusively from their non-representational properties. The fact of admitting that O nervous system is a structured system does not allow us to establish such relevance. Thus, the question is: How to show the explanatory relevance of the characterization of certain neural structures or events as representations?

One option is to argue that characterizing certain neural structures or events as representations is essential, or perhaps just useful, to explain the structuring process of the nervous system. Such an option is not advisable, since it leads back to pansemanticism, in so far as it assumes the idea that, in the internal causal flow prior to the structuring of the nervous system, there are representations that contribute to this structuring by virtue of their non-representational properties.

A second option is to assume that the process of structuring the nervous system, or at least some of its parts, is a process in which internal representations are formed. The problem with this option is that it does not really clarify the function of the formed representations, since it could be assumed that simple structuring of the nervous system is sufficient to explain the cognitive activity of an organism, regardless whether there are representations that result, in some way, from the structuring process. In other words, the formation of representations, in this case, may be nothing more than an epiphenomenon.

A third option and perhaps the most convincing is to deny that simple structuring of the nervous system is sufficient to explain the cognitive activity of organisms and admit that it is not possible, or at least difficult, to explain such activity without characterizing as representations certain components or events of the structured system. Note that this statement is not an obvious approach, but a substantive compromise. Let us admit, however, its truth. The first question to ask is: Which components or events of the structured system should (or can) be characterized as representations? The most common response has been to characterize vehicles transmitting information as representations. This response is not, however, satisfactory. We have assumed, in accordance with RTM, that an internal representation is a more or less well-delimited object or event: a neuronal N structure (or an activation of N). But a vehicle that is transmitting information should not necessarily be viewed in this way. The existence of a flow of information, as noted above, implies nomological relations in order to fulfill certain functions. But the exploitation of nomological relations does not necessarily imply the delimitation of objects or events that can serve as vehicles transmitting information. The exploitation of nomological relations is carried out when, in the face of a relevant external stimulus, a response that contributes to the satisfaction of a function of the operating system or organism is generated. The structuring that enables such a response can be seen as consolidating an internal causal link between stimulation and response, without assuming that a particular object or event must be delimited in the events that constitute such relation. From this perspective, talking about information-transmitting vehicles would simply be a way to point out that the system or organism reacts to a specific stimulus and would not imply a commitment to the delimitation of internal objects or events.

If ignoring this objection, it must be assumed that an information-transmitting vehicle may, at least in some cases, be a delimited neuronal structure (or its activation). If assuming that N is a vehicle that is transmitting information, would it be relevant to explain the cognitive activity of O to characterize the properties of N as non-representational properties of a representation? To understand the operation of N as a vehicle transmitting information, it is sufficient to grasp, on the one hand, that N is a neuronal structure whose activation depends nomologically on an A event (internal or external) and, on the other, that the activation of N contributes causally to the generation of a B response that satisfies, given the instantiation of A, a function of the nervous system of O. Characterizing N as a representation, and its properties as non-representational properties of a representation, is not relevant to this operation. On the other hand, assuming that the simple fact that N causally connects A to B justifies the characterization of N as a representation, which is not a good option, since this type of functioning is also relevant for the explanation of non-cognitive functions, such as hormone secretion, for example. Admitting that N is a representation would imply assuming the idea that every internal activity of an organism, cognitive or non-cognitive, is a form of representational activity.

But what if we assume that N specifically contributes to the execution of cognitive tasks and not to the satisfaction of other functions? Is it not now relevant to characterize N as a representation? According to Ramsey (2007), to capture the sense of computational explanations of cognition, it is necessary to assume that certain internal components that make possible to perform cognitive tasks are being used as representations. Consider an example of Ramsey: performing a multiplication operation. In a computational approach, it is typically divided the task to be performed into simpler subtasks, such as, in this particular example, the successive addition of a number.

These subtasks are executed by modules that process certain inputs and produce certain outputs. According to Ramsey, it is relevant to assume that these inputs and outputs are being used for the computational explanation, by the system itself, as representations of sums, respectively, otherwise it would not be possible to assume that the operation carried out is a multiplication, nor understand its success.

However, as Ramsey points out, the computational explanation does not question the fact that the task performed by the system is actually a multiplication operation, but simply assumes it. Therefore, the computational explanation must also assume that the processes conducted in carrying out this task are regulated, in some way, by the rules that define multiplication. Hence it may seem natural to characterize the inputs and outputs of the internal module as representations of sums, as this characterization is only a way of recognizing that there must be an influence of the rules of multiplication on the processes of the system (rules that could prescribe the successive addition by multiplying the number of times indicated by the multiplier). However, such influence has been explained. In other words, the characterization of inputs and outputs as representations of sums does not have a real explanatory function, but is an effect of the inevitable projection of multiplication rules on the behavior of the system, once assumed that the operation carried out is a multiplication. Of course, a theorist might try to explain how multiplication rules effectively influence the processes of an organism's nervous system when such an operation is performed. But addressing this problem is not the purpose of computational explanation, nor anything that such an explanation can solve.

Hence, it seems to be inconsistent by the fact that the application of internal representations is explanatory relevant because of the causal influence they exert on the cognitive activity of organisms, and at the same time it admits that the representational properties of an internal representation are not causally relevant. Such an approach is equivalent to saying that internal representations are relevant, as long as it is ignored that they are representations. Given this incongruity and the difficulties discussed in this subsection, the conclusion to be drawn from the second part of the dilemma is the same as that of the first part: The notion of internal representation is incapable of satisfying the explanatory function assigned to it by the RTM.

Doing a recap of the results of the dilemma, if it is admitted that it is due to its representational properties that R (a representation of A formed in the nervous system of an O organism) exerts causal influence on the cognitive activity of O, the existence of internal processes of interpretation and understanding of R must be postulated. Such a situation leads to an impasse, as notions of interpretation and understanding are not easy to integrate into the order of causal explanation.

But admitting that R exerts such causal influence by virtue of its non-representational properties is not a better option. Indeed, if such a thing is accepted, it is not possible to argue that the non-representational properties of R, as non-representational properties of a representation, have a true functional significance in the cognitive activity of O.

In other words, even if the non-representational properties of R had any role in this activity, it could not be argued that such a function has something to do with the fact that they are non-representational properties of a representation. Thus, the two parts of the dilemma lead us to the same conclusion: the notion of internal representation is incapable of fulfilling the explanatory function assigned to it by the RTM.

Two Examples

To reinforce the conclusion obtained in the previous section, two examples of the use given to the notion of internal representation are presented by two excellent RTM supporters: philosopher Ruth Garrett Millikan and the psychologist Susan Carey.

First example

In her early works, Millikan (1984) makes an important distinction between intentional icons and representations. An intentional T icon is an internal structure that collaborates with a biological (not necessarily cognitive) mechanism so that such a mechanism can perform its own function. T's contribution to the performance of the function of the mechanism with which it collaborates is possible, according to Millikan, thanks to the fact that T corresponds to a state of things which is a normal condition for the proper satisfaction of that function.

The correspondence relationship Millikan speaks of is an isomorphism relationship: the configuration of constituent elements of the external things of which T is an intentional icon corresponds to the configuration of constituent elements of the T structure, and certain transformations of the configuration of this same thing correspond to certain transformations of the structure from T.

For Millikan, an intentional icon, despite being an internal structure with true intentional properties, is not yet one because of two reasons, a representation. The first is that an intentional icon contributes to the conduction of biological functions of all kinds, not necessarily cognitive. The second is that, unlike a mere intentional icon, an internal representation should not simply correspond to a state of things (internal or external) in order to satisfy a function, but it must also be processed in a way that it allows the organism to represent what it represents itself.

Indeed, according to Millikan, an internal representation, besides being an intentional icon, must allow the organism in whose nervous system it has been formed to 'identify' what is represented by it. Thus, from Millikan's perspective, the fundamental cognitive question is: How can an intentional icon become a true representation?

In other words: How can an organism identify the state of external things that corresponds to an intentional icon that has been formed in its own nervous system?

Note that this problem is nothing other than the above-mentioned problem (subsection 2.1) of how a representation of A formed in the nervous system of an O organism may allow O, or an internal component of O to represent A. What is Millikan's response?

Imagining that T is an intentional icon formed in the nervous system of O and that a component of the external state of things to which T corresponds is an A. The structure of T must therefore have an element that corresponds to the presence of an A in such a state of affairs. Suppose R is that element. How can R allow O identify an A component of the external things to which T corresponds?

For Millikan, identifying is re-identifying. More precisely, being able to identify what R corresponds to does not mean to be able to represent R, A to which R corresponds and the correspondence relationship between both, but to be able to grasp that the different representations of A formed in the nervous system represent the same thing, i.e., to be able to capture when A is represented again (when different components of different intentional icons correspond to the same). Thus, O is able to identify what R represents when it is able to grasp that the different representations of A formed in its nervous system, including R, represent the same thing.

It does not presuppose the possession of metarepresentational capabilities, but those cognitive mechanisms of O make their own representations. For example, in the case of mechanisms responsible for inferences, the different representations of A contained in the contents that constitute the premises of a reasoning must be able to be used when required, such as the common term that allows the conclusion to be derived. Similarly, in the case of the action, a representation of A in the content of an intention, and a representation of A in the content of a perceptive experience must be used as its representations.

An organism whose cognitive mechanisms use its different representations of A is, for Millikan (2000), an organism capable of identifying what these representations represent[6].

Millikan's approaches respond the problem of the existence of internal interpretation processes of representations. Intentional icons that contribute to cognitive tasks would be 'consumed' in such a way that, normally, and when required, its different components would be processed—in perception, reasoning, or action—as components that correspond to it. Such mechanisms of 'consumption' – or 'readers' – of intentional icons would have been forged throughout the evolutionary development of species with cognitive abilities.

According to Millikan (2000), the question of what type of indicator determines when two components of different intentional icons should be processed as components that correspond to them is an empirical issue, for which there are plausible responses such as duplication of neural structures, the existence of some form of marker or synchronized activation.

The processing of components of different intentional icons as components that correspond to it would be observed in the behavior of the organism, i.e., in the actions, judgments and inferences that the organism carries out and whose success depends precisely on the identification of individuals and properties as the same individual or property previously represented.

In this way, intentional icons would be transformed into authentic internal representations that would allow the organism, in whose nervous system they have been formed, to represent what they presumably represent.

But is it really possible to admit that this processing of components of different intentional icons is an internal interpretation that would show the explanatory relevance of characterizing these components as representations?

Considering two neuronal structures N and N', and supposing that N and N' are components of different intentional icons, to be processed as components that correspond to it, N and N' must possess the physical characteristics that the alleged 'readers' mechanisms of intentional icons use as an indicator that these are components that correspond to it: have the same type, have a certain marker or be activated in a synchronized way.

If N and N' have these physical characteristics, they will be processed as components that correspond to it. However, since the properties relevant to the functioning of the 'readers' mechanisms are nothing more than physical properties, such processing can simply be seen as a sequence of causally connected events that contribute to the satisfaction of a particular function.

Admitting that this type of causal processing is a form of internal interpretation of representations is not enlightening, but quite the opposite, since it forces to explain why speaking of interpretation would be relevant, since the 'readers' mechanisms only react to certain physical properties of N and N'. The notion of interpretation is no easy to integrate into the order of causal explanation, and this difficulty is but one example.

One possible objection to these reasoning would be to point out that the intentional icons are used in cognitive tasks whose conduction implies the ability to identify when an individual or property is being represented.

Therefore, a way of interpreting intentional icons that allows to identify when components correspond to the same thing would be required. However, identifying when an individual or property is being represented again is something that an organism, as such, is capable of doing. In no way does it help to assume the existence of internal processes of interpretation, if this is done only to project in the activity of the nervous system the possession of the capacity itself that was intended to explain.

Second example

Based on the dual factor theory, formulated by Block(1986), that states that an internal representation is a neuronal structure whose content is at once determined by its causal connection to external properties or objects (which, by virtue of such connection it constitutes its extension) and because of the way it is used by specific cognitive mechanisms, Carey (2009) distinguishes several types of internal representations: sensitive, perceptual, basic cognition (core cognition representations) and conceptual.

All these, according to Carey, meet the requirements of two factor theory, as they are neural structures that, in addition to being causally connected with external properties or objects, are processed by certain cognitive mechanisms.

For Carey, a sensitive representation is the output of a sensory organ; for example, in the case of the eye, an image formed in the retina. Perceptual representations, on the other hand, represent concrete objects with stable properties, despite the enormous variability in the sensitive appearance of such objects, according to the conditions of the environment and the spatial relationship between the organism that perceives and the objects perceived.

The third type of representation, basic cognition, is a form of conceptual representation used in specialized tasks that imply, according to Carey, the use of information that cannot be obtained from simple sensitive or perceptive representations.

To illustrate its functioning, Carey (2009) presents the example of Indigo Tile, a migratory bird capable of identifying, in starry nights, the Earth's north. According to the experimental studies on which Carey is based, the indigo tile has innate mechanisms that allow it to detect, when still a chick, the rotation center of the starry sky. Since the apparent rotation of the starry sky is an effect of Earth's rotation on its own axis and the Earth's axis of rotation crosses the planet from north to south, detecting the rotation center of the starry sky allows identifying the northern direction of the planet (the indigo tile inhabits the northern hemisphere of the Earth).

For Carey, the indigo tile is capable of perceptively representing the starry sky with its center of rotation, but given the way its representations are employed by its navigation mechanisms, such representations cannot be considered simple perceptual representations. In particular, Carey (2009) states that the navigation mechanisms of the indigo tile are capable of 'inferring', from perceptual representations of the starry sky, the direction that this bird must follow in its migration process. The information extracted seems to exceed what is properly represented by the perceptual system of the indigo tile.

Thus, according to Carey, a perceptual representation that is used in specialized cognitive processes that are carried out by innate mechanisms such as the navigation mechanisms of the indigo tile, is a representation of basic cognition.

Finally, for Carey (2009), the representations of basic cognition, although they are conceptual representations informally superior to the perceptual representations, must be distinguished from the concepts employed in flexible processes of reasoning and theorizing, since the use given to those of basic cognition is limited, as noted above, to carry out specialized tasks carried out by innate mechanisms.

To get the explanatory irrelevance of the notion of internal representation that Carey uses, it should be noted that any neuronal structure that can be isolated in an organism's nervous system has more or less remote causal connections with external properties or objects, as well as effects inside the system.

Such a statement simply expresses the fact that there are connected causal flows in the nervous system of every organism with the outside world. Of course, the activity of the nervous system is not a mere causal flow, but a structured activity that explodes the nomological relationships between their internal events and the outside world in order to fulfill their functions. An example of this is the activity that makes possible the indigo tile to navigate.

Steven Emlen (1975), in his experimental studies, presents different groups of indigo tile chicks, raised in a planetarium, viewing starry sky with different rotation centers. During the critical learning period, as mentioned above, chicks record the observed rotation center so that, when migration arrives, the direction of flight is selected based on the location of that rotation center. Emlen's work shows that the nervous system of the indigo tile is able to exploit the nomological connections that exist between the stimulation of its internal structures, the apparent movement of the celestial vault and the terrestrial geography in order to satisfy a specific function: orientation in the migration process. At what point is it necessary, or useful, to explain this form of structured activity to postulate the existence of internal representations?

Certainly, the observation of the rotation center of the starry sky, during the critical period of learning, allows the indigo tile to calibrate its navigation mechanisms. But the functional connection between observation of the rotation center of the starry sky and selection of the direction of flight can be seen as a simple consolidation (thanks to innate mechanisms) of an internal causal connection.

Such consolidation does not require the delimitation of neural structures that can be characterized as representations. And although it was possible to delimit certain structures (or events) in the series of internal events and assign them the function of vehicles transmitting information, there would be no reason to assume that such structures (or events) are more than causal mediators that, by virtue of its nomological connections and its effects on the system, they contribute to satisfying the orientation function of the indigo tile.

If assuming, however, that the structures are genuine representations, these structures would not have explanatory relevance as representations, because the fact that these delimited structures are representations could be an effect without functional significance on the activity of the nervous system of the indigo tile, i.e., a simple epiphenomenon.

For Carey, the representations formed in the nervous system of an organism, particularly the representations of basic cognition, are processed by cognitive modules that produce new representations. In the case of indigo tile, according to Carey (2009), the innate cognitive module by which this bird manages to orient in its migration process admits representations of the starry sky as inputs and produces representations of the direction to follow as outputs. However, both the application of cognitive modules and the application of internal representations are nothing more than a pure effect of the projection, in the activity of the nervous system, of the norms that regulate the capacity that is intended to explain.

Based on the fact that the indigo tile nervous system is capable of guiding this bird in its migration process, we must assume that the rules defining the proper exercise of this capacity must, in some way, regulate the activity of the system. Talking about representations of the starry sky processed by internal modules that 'infer' representations of the direction of flight to be followed is simply a way of recognizing that there must be an influence of such rules on the activity of the nervous system of the indigo tile, but by no means it is an explanation of how this influence is possible.

Conclusion

To conclude, the rejection of the application of internal representations does not oblige us to reject the evidence that Carey and her collaborators have obtained in favor of the existence of a basic and innate form of cognition. The application of internal representations and cognitive modules responsible for processing them is not properly supported by such evidence. It is merely a reflection of the use of RTM, whose central thesis is that the cognitive life of an organism is reduced to the formation, storage and processing of internal representations. The fact that the nervous system of an organism with cognitive abilities is a highly structured system and that such structuring is largely innate does not imply the existence of internal representations.

In general, the rejection of the notion of internal representation does not necessarily imply the rejection of the results obtained by the scientists of cognition, even by those who use this same notion. The notion of internal representation is incapable of satisfying the explanatory function assigned to it by RTM, but that does not mean that this notion cannot have any other function that can explain, to some extent, its widespread use.

As mentioned above, postulating the existence of internal representations is a (often implicit) way of recognizing that the processes of the nervous system that allow the conduction of cognitive tasks must be subjected to the rules that define the correct execution of such tasks. Assuming such a thing may be legitimate, although it must also be admitted that the application of internal representations does not explain the influence of such rules, but it is simply a way to recognize it. Explaining the nature of its influence and how exactly it occurs is not necessarily an obligation of cognitive scientists. Nor should it be thought, however, that the application of internal representations can solve these difficulties.

Notes

1. Strictly speaking, there is a third option: The causal power of an internal representation derives from both types of properties. This option, however, is not relevant here, as it does not affect the arguments that will be presented below.
2. Fodor and Pylyshyn (2015) even claim that there may be a better solution to this problem, but they do not know "any other" (p. 7). In short, the use of the expression "syntactic properties", or "properties of the shape of a symbol", to refer to the non-representational properties of a representation does not modify in any way its character as mere non-representational properties of a representation.
3. My translation.
4. My translation.
5. In this respect, we differ from Dretske (1981) when he states that "in the beginning there was information" (p. vii). In the beginning there was no information, as there were no actual signals or transmitting vehicles, but mere nomological relationships between structures, events or objects. Of course, there were potentially transmitting structures of information, i.e. structures that could have been exploited by systems or agencies with the required capabilities (if any).
6. Millikan (2017) even argues that it is apparently a "necessary" consequence of RTM that "one thing is identified when its signs are co-identified" (p. 51).

References

BLOCK, Ned
1986 Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615-678. http://dx.doi.org/10.1111/j.1475-4975.1987.tb00558.x

CAREY, Susan
2009 *The origin of concepts*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195367638.001.0001

CUMMINS, Robert
2010 *The world in the head*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:osobl/9780199548033.001.0001

DRETSKE, Fred
1981 *Knowledge and the flow of information*. Cambridge, Massachusetts: M.I.T. Press.
1988 *Explaining behavior*. Cambridge, Massachusetts: M.I.T. Press.

DUMMETT, Michael
1993 *The seas of language*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/0198236212.001.0001

EMLEN, Steven
1975 The stellar-orientation system of a migratory bird. *Scientific American*, *233*, 102-111. http://dx.doi.org/10.1038/scientificamerican0875-102

FODOR, Jerry
1984 Semantics, Wisconsin style. *Synthese*, *59*(3), 231-250. http://dx.doi.org/10.1007/BF00869335
1990 *A theory of content and other essays*. Cambridge, Massachusetts: M.I.T. Press.

1995 Concepts; a potboiler. *Philosophical Issues*, *6*, 1-24. http://dx.doi.org/10.2307/1523025

1998 *Concepts. Where cognitive science went wrong*. Oxford: Clarendon Press. http://dx.doi.org/10.1093/0198236360.001.0001

FODOR, Jerry & LEPORE, Ernest

1991 Why meaning (probably) isn't conceptual role. *Mind and Language*, *6*(4), 328-343 http://dx.doi.org/10.1111/j.1468-0017.1991.tb00260.x

FODOR, Jerry & PYLYSHYN, Zenon

2015 *Minds without meanings.* Cambridge, Massachusetts: M.I.T. Press. http://dx.doi.org/10.7551/mitpress/9780262027908.001.0001

GODFREY-SMITH, Peter

2006 Mental representation, naturalism, and teleosemantics. En G. Macdonald & D. Papineau (Eds.), *Teleosemantics* (pp. 42-68). Oxford: Oxford University Press.

HUTTO, Daniel & MYIN, Erik

2013 *Radicalizing enactivism.* Cambridge, Massachusetts: M.I.T. Press. http://dx.doi.org/10.7551/mitpress/9780262018548.001.0001

JOHNSON-LAIRD, Philip

2006 *How we reason.* Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199551330.003.0028

MERCIER, Hugo & SPERBER, Daniel

2017 *The enigma of reason.* Cambridge, Massachusetts: Harvard University Press. http://dx.doi.org/10.4159/9780674977860

MILLIKAN, Ruth Garrett

1984 *Language, thought and other biological categories*, Cambridge, Massachusetts, MIT Press.

2000 *On clear and confused ideas.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511613296

2017 *Beyond concepts.* Oxford: Oxford University Press. http://dx.doi.org/10.1093/oso/9780198717195.001.0001

NEANDER, Karen

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

PRINZ, Jesse

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

RAMSEY, William

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

STERELNY, Kim

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

Document reception date: May 23, 2020
Document review date: July 15, 2020
Document approval date: December 15, 2020
Document publication date: July 15, 2021