Aufklärung Revista de Filosofia

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86
DOI: https://doi.org/10.18012/arf.v8iesp.60021
Recebido: 30/01/2021 | Aceito: 30/03/2021

THE REPRESENTATIONAL THEORY OF MIND AND COMMON SENSE PSYCHOLOGY[1]

[A TEORIA REPRESENTACIONAL DA MENTE E A PSICOLOGIA DO SENSO-COMUM]

Raquel Krempel *
Universidade Federal de São Paulo, Brasil

ABSTRACT: The goal of this paper is to present some advantages of the representational and computational theories of mind when compared to other views, especially behaviorism. The idea is that representational and computational theories allow us to conceive propositional attitudes (mental states such as beliefs and desires) in a way that preserves two essential features we take them to have in common sense psychological explanations: semantic evaluability and causal efficacy. Behaviorism reconceives mental states in a way that doesn't preserve these essential features. In so doing, it makes a mystery of the success of common sense psychology. I illustrate some of the difficulties that behaviorism faces by considering and criticizing Wittgenstein's approach to linguistic understanding. The upshot is that representational and computational theories of mind do a better job at vindicating common sense psychology, and so are to be preferred when compared to behaviorism.

KEYWORDS: Mental representations; psychological explanations; behaviorism; Wittgenstein

RESUMO: O objetivo deste artigo é apresentar algumas vantagens das teorias representacional e computacional da mente quando comparadas a outras visões, especialmente o behaviorismo. A ideia é que as teorias representacionais e computacionais nos permitem conceber atitudes proposicionais (estados mentais, como crenças e desejos) de uma forma que preserva duas características essenciais que consideramos que elas têm nas explicações psicológicas do senso comum: avaliabilidade semântica e eficácia causal. O Behaviorismo reconcebe os estados mentais de uma forma que não preserva essas características essenciais. Ao fazer isso, o sucesso da psicologia do senso comum torna-se um mistério. Ilustro algumas das dificuldades que o behaviorismo enfrenta ao considerar e criticar a abordagem de Wittgenstein da compreensão linguística. O resultado é que as teorias representacionais e computacionais da mente fazem um trabalho melhor em defender a psicologia do senso comum e, portanto, devem ser preferidas em comparação com o behaviorismo.

PALAVRAS-CHAVE: representações mentais; explicações psicológicas; behaviorismo; Wittgenstein

* *Pesquisadora de Pós-doutorado em Filosofia Unifesp/FAPESP. Email: raquelak@gmail.com*

"nothing ever reduces to anything, however hard philosophers may try"
(FODOR, 1998, p. 66)

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

INTRODUCTION

Lara went to the bakery because she wanted to buy bread and cookies. There she was partly successful: she got the bread but she found out that, contrary to what she believed, they didn't sell cookies. She was disappointed when she learned that, because she really wanted cookies. But she moved on with her life. This simple story serves to illustrate some aspects of our common sense psychology. We ordinarily attribute, to others and to ourselves, mental states such as beliefs and desires, and this allows us to predict and explain behavior.

Beliefs, desires and other propositional attitudes[2] have two central features: causal efficacy and semantic evaluability (FODOR, 1987). As Fodor notes, we ordinarily assume the existence of three types of mental causation: mental states cause other mental states, mental states cause behaviors and external events cause mental states. Lara had the desire to buy bread and cookies and the belief that she could find both things at the bakery, which made her decide to go there. This illustrates our conception that mental states can cause other mental states. But mental states can also cause behavior: Lara's decision to go to the bakery caused her to go to the bakery. And external events can cause mental states: her being told that they didn't sell cookies caused Lara to be disappointed. In short, in our daily psychological explanations of behavior we attribute causal powers to mental states.

Moreover, beliefs, desires and other propositional attitudes have a relation to the world that makes them true or false (in the case of beliefs) and fulfilled or frustrated (in the case of desires) (FODOR, 1987, p. 10). That is, these mental states, besides having causal efficacy, also seem to have semantic or representational content, which makes them semantically evaluable, that is, evaluable by how they relate to the world. Beliefs, for instance, represent the world as being a certain way (or express a given proposition), and this makes them semantically evaluable, that is, capable of being true or false.[3] Lara's belief that they sell cookies at the bakery represented the world as being a certain way, which ended up being false. Her desire to eat cookies was about (or represented) cookies and herself eating them, but it ended up not being fulfilled.

At the same time that beliefs, desires and other propositional attitudes can cause things to happen, like material things usually can (e.g. the moon can cause tides, or hurricanes can cause destruction), they can also be about things, in a way that most material things cannot. In order to explain and predict the behavior of humans and many non-human animals, but not the behavior of the moon, or of hurricanes, we resort to states that are representational and causally efficacious. What best explains Lara's behavior of leaving the house and walking to the bakery is her desire to buy bread and cookies (and not, say the laws of physics or even neurology). Everyday intentional psychological explanations are in fact so pervasive in our lives that, were we to find that it is not literally true that intentional mental states have causal powers, that would be, as Fodor famously puts it, "the greatest intellectual catastrophe in the history of our species; if we're that wrong about the mind, then that's the wrongest we've ever been about anything." (1987, p. xii).

Attributions of propositional attitudes, conceived as being both causally efficacious and semantically evaluable, are then central to our understanding of ourselves and others, in that they allow us to predict and explain behavior. They are

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

generally, even if not always, successful (DEVITT 2006, CRANE 2003, FODOR 1987). This suggests that a theory of intentional mental states that preserves their two central features, namely, causal efficacy and semantic evaluability (as well as representational content), is to be preferred when compared to a theory that doesn't accommodate those features. A theory that denied that intentional mental states are causally efficacious and semantically evaluable would force us to reconceive our common sense psychological explanations, possibly implying their falsity.

How are we then to understand the nature of mental states? Different views have been proposed. In section 1, I will present the representational and the computational theories of the mind, which preserve and explain the common sense understanding of propositional attitudes. In section 2, I will briefly present the identity theory and behaviorism, to give the reader a sense of how they differ from the representational and computational views, when it comes to accommodating common sense psychology. My focus, however, will be on behaviorism and some of its shortcomings (section 3), as it is unable to explain mental causation. In section 4 I will illustrate this by considering and criticizing Wittgenstein's approach to linguistic understanding, as it bears some similarities to behaviorism, showing that it suffers from similar difficulties. The conclusion is that representational and computational theories, unlike behaviorism, are compatible not only with common sense but also with cognitive psychology, and so are to be preferred.

1. Representational and computational theories of mind

One of the difficulties for a theory of mind that is intended to vindicate common-sense psychology is to explain how states with semantic content can have causal efficacy. As Crane puts it, "if we are going to explain thought, then we have to explain how there can be states which can at the same time be representations of the world and causes of behavior." (2003, p. 83). How can beliefs and desires, which seem to be essentially intentional states (in the sense of being *about* things), cause and be caused by other mental states, external events, and behaviors? How can, say, my belief that it is raining cause me to get an umbrella? The representational and the computational theories of the mind are intended to be theories that vindicate common-sense psychology, giving a naturalistic account of mental states that are at once causally efficacious and semantically evaluable, without eliminating or reducing them to entities that lack those features.

How do representational and computational theories do that? Let us briefly look first at what the Representational theory of mind (RTM) says about intentional mental states, and then at what the Computational theory of mind (CTM) says about mental processes. According to RTM, propositional attitudes such as beliefs and desires are functional relations that organisms have to mental representations. If I have a belief or a desire, what happens is that I have a mental representation to which I am related in a certain way, which varies depending on whether my mental state is a belief, a desire, etc. Having the belief that it will rain, for example, is to be in a functional relation to a mental representation or symbol that means that it will rain. To wish or hope for rain is to be in a different functional relation, with different causal roles, to the same mental symbol. Beliefs and desires, as attitudes, are different because they have different functions in a system, that is, they each have typical causes and effects, typical relations with inputs, outputs, and other mental states. What makes my belief that it will rain a belief, and not a desire for rain, are the typical causes and effects it has. It can be caused

by my perception of dark clouds in the sky, and may, along with my desire not to get wet, cause me to take an umbrella before leaving the house, etc. My desire for rain, on the other hand, does not have the same causes and effects as my belief that it will rain. This desire can be caused, for example, by my desire to stay at home doing nothing, and probably will not, by itself, cause me to take an umbrella before leaving the house. What both of these mental states have in common is that both are relations that individuals have to tokens of the same (or approximately the same) type of mental symbol, a symbol which represents something like that there will be rain here in the near future.

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

The computational theory of mind is introduced as an account of mental processes, that is, causal sequences of thoughts. If I believe it is raining, and I think that if it is raining, I will not go to the beach, and from that I decide to cancel my plan to go to the beach, what happened was a causal sequence of symbols or mental representations. According to CTM, sequences of thoughts are to be treated as transformations of symbols or mental representations that occur mechanically in virtue of their syntactic forms (and not their semantic contents) and that follow certain rules or algorithms. The comparison of the mind to a computer, the computational theory of the mind, serves mainly to explain mental processes such as reasoning, i.e. to explain how, when we reason, one thought can cause another while preserving semantic properties such as truth.[4]

A computer can be understood as "a device which processes representations in a systematic way" (CRANE, 2003, p. 85; see also p. 121). In other words, computers process information through the algorithmic transformation of symbols (or representations). The transformation of symbols by a computer is guided by their syntax (or by their formal properties), and not by what these symbols mean. As Fodor says, "computations just *are* processes in which representations have their causal consequences in virtue of their form." (1981a, p. 241). Fodor takes the syntactic form of a symbol to be "one of its higher-order physical properties. To a metaphorical first approximation, we can think of the syntactic structure of a symbol as an abstract feature of its shape." (1987, p. 18).

In adopting RTM and CTM, one usually also accepts that mental representations are, at least in biological organisms, realized in the brain. The usual way to understand this is to think of the brain as the hardware, and of the mind as the program that the brain instantiates. The idea is that there is a legitimate level of explanation about the functioning of the mind that refers to symbols and their transformations, just as in a computer we can talk of the programming, which is a description that falls in between the fully physical level of the electrical circuits, and the level of the outputs. And if symbol manipulation is not mysterious in a computer (at least not for programmers), neither should it be in the mind and brain.

In sum, the computational theory of mind helps us understand how at least some mental processes can occur mechanically. Computers show us how symbols can cause other symbols and preserve semantic relations at the same time. Analogously, we explain the causal powers of intentional mental states by the assumption that they have causal powers in virtue of the formal aspects of mental representations. Thinking ceases to be conceived as an abstract and immaterial activity, whose causal power is mysterious. It is taken to be transformations of symbols, which are ultimately realized in the brain. RTM and CTM are intended to show us how the causal powers of intentional states can be made a little less mysterious. To the extent that beliefs (and other propositional attitudes) are relations to symbols, and that mental processes are computations, we can explain from a naturalistic point of view how there can be causal

sequences of intentional states, such as in reasoning, that preserve semantic properties (such as truth).

2. RTM AND OPPOSING VIEWS

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

In the last section, we saw some central aspects of the representational and computational theories of mind. Now we can ask what their rivals are and the advantages of accepting RTM and CTM. Let us first look at some central aspects of the representational theory of mind, and then briefly at what some competing theories say, and why they do not seem so advantageous. My intention is not to give an exhaustive treatment of the opposing views, but merely to indicate that RTM and CTM have some advantages over some opposing views (in particular behaviorism, as we will see in the next section), when it comes to preserving common sense psychology.

First, being a *realist* theory about the existence of propositional attitudes, RTM contrasts with eliminativism, which denies the existence of intentional states. The paradigmatic example here is the eliminative materialism of the Churchlands (which I will not discuss here).[5]

In addition to being a realist theory, RTM can be classified as a materialist theory of the mind. In principle, as Fodor (1981b) admits, functionalism – and possibly also RTM and CTM – is compatible with substance dualism (the Cartesian idea that mind and matter are two different substances) and even with idealism (such as Berkeley's, according to which there is no matter, only minds and ideas). But to the extent that we are interested in a naturalist theory of the mind, there is no need to assume any of these views, for RTM and CTM are also compatible with a materialist view of the mind.

In addition, RTM and CTM are non-reductive about mental states. In taking propositional attitudes to be functional relations that we have to mental representations, and in taking mental processes to be computations, RTM and CTM preserve both the semantic and the causal aspects of propositional attitudes. Hence, they do not reduce mental states and processes, as they are conceived in common sense psychology, to entities without these properties. They seem to vindicate intentional common-sense psychology.

As a non-reductive materialist theory of mind, Fodor (1975, 1981b) notes that RTM is opposed to materialist theories that can be considered reductionist, such as behaviorism and the identity theory (also known as type-physicalism). Behaviorism can be characterized as the view according to which mental states can be reduced to behavioral dispositions. The identity theory can be characterized as the view that types of mental states are identical to types of states in the brain. As Fodor notes, these theories can be considered forms of materialism, since mental states are nothing more than material entities: behaviors (or behavioral dispositions), in the case of behaviorism, and brain states in the case of the identity theory. They do not, unlike dualism, take mental states to belong to a domain other than the domain of material entities. But unlike RTM, they can be considered reductionist, since for them mental states do not have an independent status, but are instead reduced to other types of entities: behavioral dispositions and brain states.[6]

One question that arises is whether the entities to which behaviorism and the identity theory reduce beliefs and desires preserve the two characteristics that, as we've seen, seem to be essential to them: causal efficacy and semantic evaluability. If not, they cannot be said to vindicate common sense intentional psychology.

As for causal efficacy, it seems that it can be preserved by brain states, insofar as these are material entities. An identity theorist could explain the causation between mental states by saying that it is nothing more than a brain state causing another brain state. A mental state causing a behavior would similarly be nothing more than a physical brain state causing a behavior. And an external event that causes a mental state would be an external event causing a brain state. The identity theory passes this first condition for the acceptance of propositional attitudes, since brain states are causally efficacious.

Behaviorism could try to account for some cases of mental causation by saying that a mental state causing a behavior is nothing more than the disposition to produce a response given a certain stimulus. For example, instead of saying that it was thirst that caused John to drink water, Fodor (1981b) notes that the behaviorist might say that being thirsty is the same as being disposed to drink water if water were available, and that in this case there was water available. There was a stimulus, the presence of water, which led John to respond by drinking the water. However, it is questionable that this behaviorist approach actually preserves the common sense notion of mental causation. Behaviors are not mere responses to stimuli, but are often the product of interactions of mental states. It is questionable that stimuli and behavioral responses can be regarded as causally efficacious in the same sense as mental states, since they do not account for causation between mental states, and doubtfully preserve the idea that mental states cause behavior (I will come back to this in the next section). The entities accepted by behaviorism do not seem to be sufficient to account for all kinds of mental causation admitted by common sense psychology.

As for the second characteristic we commonly attribute to mental states, namely, semantic evaluability, it is not clear that it is preserved by either the identity theory or behaviorism. The problem is that it seems like a category mistake to say that a brain state, or a behavioral disposition, is false (like a belief is false), or that it has been frustrated (like a desire has). On the face of it, it would be like attributing blueness to my brain state when I have a perceptual experience of the blue sky. A case would have to be made for the idea that semantic content and evaluability can be attributed to brain states and behavioral dispositions, just as we can say that a belief or a desire is about something and evaluable by how it relates to the world.[7]

Proponents of the identity theory and behaviorism could embrace the view that the entities they reduce mental states to aren't really semantically evaluable and/or causally efficacious. If so, then these theories are probably not entirely realist about the states presupposed by intentional common-sense psychology, since they reduce such states to entities which do not preserve all their essential properties. In this respect, they would differ from the representational and the computational theories of mind. They do not accept entities with the same essential properties as the ones propositional attitudes have according to common sense (i.e. causal efficacy and semantic evaluability), and so they do not vindicate common-sense psychology. But as we've seen, our psychological explanations of behavior are pervasive and generally successful. Knowing what someone believes and desires helps us explain why she behaves the way she does, as well as predict what she will or would do in other circumstances. In denying that intentional mental states are really causally efficacious and/or semantically evaluable, behaviorism and the identity theory would force us to reconceive common sense psychological explanations, implying that they are not literally true. They would make a mystery of the success of common sense psychological explanations. RTM, on the contrary, in accepting entities that are semantically evaluable and causally efficacious, accommodates the success of common sense psychology really well. Let us now see in

a little more detail some of the problems with behaviorism.

3. Behaviorism and its shortcomings

"ontology is one thing, epistemology is quite another."
(FODOR, 2003)

Hilary Putnam (1963; 1967) was one of the main philosophers to oppose behaviorism. He raised serious problems for the behaviorist hypothesis that mental states, like pain, are dispositions to behave in certain ways. Putnam acknowledges that the behaviorist's conception of pain has the advantage of being more in line with how we verify that someone is in pain. When we attribute pain to someone, we usually do so based on the person's behavior, and not by observing her functional organization, or her brain states. However, while the way we verify whether someone is in pain may tell us something relevant about the concept "pain", Putnam rightly observes that it does not tell us what pain is. In general, the way we verify that a thing *x* is *A* may in fact be irrelevant for knowing what property *A* is. In everyday life, I don't verify that a substance is water by analyzing its chemical composition, but that does not mean that water is not H_2O. Similarly, I certainly do not check that someone is in pain by opening her skull, or by analyzing her functional organization, but that does not mean that pain is not a cerebral or a functional state.[8] To think that verification criteria determine what a thing is would be, to use one of Fodor's expressions, to "put the epistemological cart before the ontological horse."[9]

In "Brains and behavior" (1963), a paper in which Putnam, in his own words, attempts to bury logical behaviorism (conceived as a theory about the meaning of mental terms), he offers more reasons for not wanting to identify pain with a disposition to behave in a certain way. He argues that from the fact that we can conceive of worlds in which pain does not correspond to any kind of behavioral disposition we can conclude that there is no necessary connection between pain and typical pain behavior, and therefore that pains (and presumably other mental states) cannot be reduced to behavior dispositions. Moreover, according to Putnam, the word "pain" does not mean a certain group of behavioral responses to certain stimuli, but rather "the presence of an event or condition that normally causes these responses." Thus, against logical behaviorism, he concludes that sentences about pain cannot be translated into sentences about behaviors without loss of meaning, because by "pain" we mean what *causes* certain typical behaviors, not the behaviors themselves.

Fodor also notes that we not only take behaviors to be effects of mental states, but also that behaviors are often the result of mental processes, that is, of causal sequences of mental states. So it doesn't seem possible to associate a single mental state with a behavioral disposition, because it is usually the *interaction* of different mental states that causes behavior. According to him,

> Mental causes typically give rise to behavioral effects by virtue of their interaction with other mental causes. For example, having a headache causes a disposition to take aspirin only if one also has the desire to get rid of the headache, the belief that aspirin exists, the belief that taking aspirin reduces headaches and so on. Since mental states interact in generating behavior, it will be necessary to find a construal of psychological explanations that posits mental processes: causal sequences of mental events. It is this construal that logical behaviorism fails to provide. (FODOR, 1981b, p. 116).

The behaviorist does not seem to be able to explain, in purely dispositional terms, without mentioning mental states, how one can, for instance, be in pain and not manifest it *because* one believes that it is shameful to express pain. If mental states have to be mentioned to characterize behavioral dispositions, then mental states are not really reducible to behavioral dispositions.

We have been considering the case of sensations like pain and thirst, but the reduction of propositional attitudes to behavioral dispositions is also problematic. After all, to what kind of behavioral disposition can we reduce, say, the belief that Mercury is the smallest planet in the solar system, or the desire to eat chocolate? One could suggest that there are several behavioral dispositions associated with the belief that Mercury is the smallest planet in the solar system, one of them being the disposition to answering "yes" if someone asks you if you believe that Mercury is the smallest planet in the solar system. But Mary might have that belief while not being disposed to answering the question because, e.g., she doesn't like the sound of her own voice, or because she thinks the answer is too obvious.[10] Similar observations apply to the desire to eat chocolate. This desire will only be associated with being disposed to eat chocolate if one doesn't also have the desire not to harm one's teeth, or the desire to lose weight (coupled with the belief that chocolates make you gain weight, and the determination to resist the temptation of chocolate). In both cases, the circumstances that need to be in place for the dispositions to hold will inevitably make reference to other mental states of the individual. It does not seem possible, then, to characterize the belief that Mercury is the smallest planet in the solar system, or the desire to eat chocolate, in terms of dispositions to behave in certain ways without introducing other mental states in the characterization of the circumstances presupposed by the dispositions. The behaviorist's attempt to reduce mental states to behavioral dispositions is, then, unsuccessful.

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

Even if the behaviorist were successful in reducing mental states to behavioral dispositions, leaving the mental state out is leaving the cause of the behavior out, as Putnam observes. That is, even if there were a clear correspondence between mental states and behavioral dispositions, not mediated by several other mental states – which there doesn't seem to be – it is not clear that we should be satisfied with this characterization of mental states, since in so characterizing them we lose the explanatory power we ordinarily attribute to mental states, but not to stimuli and behavioral responses. Behaviorism is therefore not successful in preserving the causal efficacy that common sense psychology attributes to mental states. As Fodor notes, behaviorism invites us to deny the undeniable: the contribution of internal states to the causation of behavior (FODOR, 1975, pp. 1-2). It is unclear how we could predict and explain behavior with no reference to causally efficacious intentional mental states.

4. The case of Wittgenstein

It is worth considering one example of a philosophical approach, found in Wittgenstein, that is in certain respects similar to behaviorism. Wittgenstein didn't hold the view that all mental states can be reduced to behavioral dispositions, but, in the spirit of behaviorism, he takes that, at least when it comes to explaining certain phenomena (such as linguistic understanding), mental states and processes are unnecessary. In *The blue book*, for instance, Wittgenstein criticizes the idea that mental processes give life or meaning to linguistic signs, which he appears to take to be the same as the idea that we must form mental *images* that interpret linguistic signs, in order to understand a sentence. For example, if I ask someone to bring me a red flower,

according to Wittgenstein, it is not necessary to suppose that a mental image of a red flower passes through one's mind, serving as an interpretation of the request and allowing one to execute it. Wittgenstein's point, I believe, is that a mental state that functions as an interpretation of linguistic signs does not always occur, and need not occur, for words to be meaningful, or for us to be able to, e.g., understand and follow orders. Even if a mental image of a red flower does occur when I hear the order to pick up a red flower, for Wittgenstein this image cannot be what explains my understanding of the order. Were we to suppose that what explains the comprehension of a linguistic sign is a mental sign (such as a mental image), we would then be required to say what in turn explains the understanding of that mental sign. If it were another mental sign, the problem would go on ad infinitum. Wittgenstein then suggests that the meaning of a word is not something mental accompanying it, but rather its use. As he says,

> The signs of our language seem dead without these mental processes [of understanding and meaning]; and it might seem that the only function of the signs is to induce such processes, and that these are the things we ought really to be interested in. (...) But if we had to name anything which is the life of the sign, we should have to say that it was its *use*. (WITTGENSTEIN, 1958, pp. 3-4).[11]

For Wittgenstein, the linguistic sign is not something dead that gains meaning through a psychological process of understanding, which takes place in a mysterious mental medium; it gains meaning or life by being used in a certain way. We are tempted to think that mental processes corresponding to meaning and understanding are necessary to explain linguistic behavior but, for Wittgenstein, they are not.

I think it is possible to raise against Wittgenstein similar problems to those that we have raised against the behaviorist characterization of pain. We can apply to Wittgenstein's flower example the distinction drawn by Putnam between what a thing is and the ways or criteria we use to verify it. In that example, since Wittgenstein denies that understanding the order of fetching a red flower is a mental process, he would probably say that it is rather something that is manifest in one's action of bringing a red flower. Now, there is no need to deny that we could *verify* that someone has understood an order to bring a red flower by e.g. observing his or her flower-picking behavior, but that does not mean that the understanding of the order *is* the same thing as being disposed to bring a flower. Similarly, we typically learn that someone understands a sentence by observing whether the person is able to use it correctly, or to behave appropriately. But that doesn't mean that using a sentence correctly, or behaving appropriately, is all there is to understanding it, or even that using a sentence correctly, or behaving appropriately, is necessary for understanding. On the contrary, using a sentence correctly, or obeying an order, are *consequences* of the understanding of a sentence or an order. Understanding is part of the explanation of the behavior of obeying the order, and not something to be identified with it.

Mental processes are necessary e.g. to explain the differences in behavior between an English speaker and a monolingual Chinese speaker when receiving an order in English to bring a red flower. It seems reasonable to say that the English speaker brings the red flower in part *because* she understood the order, while the Chinese speaker remained static *because* she did not understand the order. To the extent that the understanding of an order is part of the causal chain that leads one to bringing a red flower when asked to, it seems false that no underlying mental processes corresponding to the understanding are required to explain linguistic behavior.

It is true that the evidence we use to attribute the understanding of an order to someone is usually behavioral, just like it is in the case of pain. But that does not mean

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

that understanding an order, pain, etc., *just are* behavioral rather than mental events. When I say that someone is in pain, I do not mean to be saying that she is behaving in a certain way. What I mean is that she is *feeling* pain, and that this feeling is the cause of her pain behavior, even if the behavior (and not the feeling) is the evidence I use to attribute pain to her. Likewise, that someone uses the words of a language correctly, and that she brings me a red flower when requested, may be the evidence I use to determine that this person understands what I say. But understanding a request is not the same thing as being disposed to obey it.

Understanding an order cannot be the same thing as being disposed to obey it because there is no necessary connection between one thing and the other, just as there is no necessary connection between feeling pain and being disposed to exhibit pain behavior. One can understand an order and not be willing to obey it, or obey it by chance, without understanding it. We can conceive a situation, similar to Searle's Chinese room (1980), in which someone who does not speak English systematically brings me a red flower whenever I ask for one in English, because of an incredible coincidence, without her action being causally related to my request, and without her having *understood* my request. If being disposed to obey the order were the same thing as understanding it, we should say in this case that the person understood the order, which is absurd, since by assumption the person does not speak English. If this is so, then there is a difference between linguistic understanding and behavior (or behavioral dispositions), the latter being generally a consequence of the former.[12] And if linguistic understanding is, as it seems to be, a mental event, then mental processes corresponding to the interpretation of linguistic signs are not unnecessary or irrelevant to explain behavior, as Wittgenstein seems to believe. Mental processes typically cause behavior, including language-related behavior.

4.1 Mental processes and mental images

Now Wittgenstein seems to think that, if mental states or processes were involved in the understanding of a sentence, they would take the form of conscious mental *images*. That is, he seems to be equating the idea that mental processes are responsible for linguistic understanding with the idea that mental images are responsible for linguistic understanding. As he says, "it may seem essential that, at least in certain cases, when I hear the word 'red' with understanding, a red image should be before my mind's eye." (WITTGENSTEIN, 1958, p. 4). He then opposes this idea. He argues that mental images are not necessary to explain linguistic understanding. They are not what gives life to linguistic signs. To illustrate this, he imagines a situation in which one has a chart with names for colors corresponding to colored squares. To carry out the order of bringing a red flower, instead of using a mental image, the person searches for the color in the chart and tries to find a flower with a color that matches it. In that situation, Wittgenstein says, we would not be inclined to posit a mental image corresponding to the understanding of the word "red", in addition to the physical color chart. But then why should the image ever be necessary? "As soon as you think of replacing the mental image by, say, a painted one, and as soon as the image thereby loses its occult character, it ceases to seem to impart any life to the sentence at all." (Idem, p. 5). Mental images, even if they occur, are not necessary and do not explain linguistic understanding.[13]

Wittgenstein is right to say that mental images are not necessary for linguistic understanding. However, we should not equate having mental processes responsible for

linguistic understanding with forming conscious mental images. In saying that linguistic understanding involves mental processes, and that it is not just a matter of how one behaves given certain linguistic stimuli, I don't mean to be saying that understanding a sentence requires forming conscious mental *images*. Most cognitive scientists, including Fodor, would deny that, along with Wittgenstein. One thing that suggests that this view is false is that there are individuals, with a condition called aphantasia, who are incapable of forming conscious mental images (ZEMAN et al., 2015, DAWES et al. 2020). If understanding a sentence required the forming of conscious mental images, then aphantasics should be incapable of understanding language. But nothing indicates that they are. Given that they are capable of understanding language, then forming conscious mental images cannot be necessary for linguistic understanding.

It would be wrong, however, to conclude from the fact that there need be no conscious mental images accompanying linguistic understanding, that no mental states or processes are involved in linguistic understanding. The science of psycholinguistics would not have come so far without explanations of linguistic understanding that appealed to mental states and processes involving mental representations. Mental representations are likely involved in linguistic understanding even if conscious mental images are not.

Of course Wittgenstein might rightly demand an explanation of the mental processes responsible for understanding and that give life to linguistic signs, especially if they involve other signs, this time mental signs (i.e. mental representations), which in turn need to be interpreted. That is, if the signs of a language acquire life (or meaning) because they express mental signs, we can ask what gives life to these mental signs. Do we need to postulate other signs that give life to them, leading to an infinite regress? I take that this is one of Wittgenstein's main points against mentalist accounts of linguistic understanding. This may in fact be a problem for proponents of the representational and the computational theories of mind who accept that there are mental (though not necessarily imagistic) symbols. But, as Cain notes, "it is far from clear that cognitive scientific explanations must be either circular or lead to infinite regress" (CAIN, 2002, p. 40). He argues that the regress can be eliminated if we accept that CTM shows us how the manipulation of symbols in the mind can happen in a purely mechanical way. Natural language symbols can derive their meanings from mental symbols, which in turn are assumed to be physically realized in the brain. Mental symbols would end the symbolic chain, with no need to assume more symbols to explain our understanding of them. One could then simply say that the symbols in the language of thought don't require any interpretation. We interpret natural language sentences, but we simply have thoughts, we don't interpret our thoughts.[14]

This is not to say that the semantics of thought is unproblematic. Here one could still demand some explanation for how thoughts manage to be about things (or, to put the issue in Wittgensteinian terms, an explanation for what gives life to mental signs), and for what makes a mental representation mean what it does. Several attempts have been made. Fodor, for instance, argues in several places for a causal theory of meaning (1987, 1990), but it is not my intention to review it here. But from the fact that we need to account for the semantics of mental representations, and that that is not an easy task, it doesn't follow that mental representations do not exist, nor that assuming their existence is not explanatorily useful. Language-related behaviors would be mysterious if we did not attribute mental causes to them. There does not seem to be anything absurd with the assumption that mental events cause behavior, since we daily assign mental causes to explain and predict people's behavior.

4.2 Wittgenstein and Fodor

As distant as Wittgenstein's and Fodor's ideas may be, there being perhaps nothing (the second) Wittgenstein would more strongly object to than the idea of a language of thought, it is interesting to note that their philosophical projects seem to have similar motivations. Wittgenstein wants to reject the notion of mind as an immaterial entity whose mechanisms are incomprehensible, as well as the notion of thought as a mysterious attribute belonging to this "queer kind of medium" (1958, p. 3). Ryle (1949), another philosopher often associated with behaviorism, has similar concerns. He criticizes Cartesian dualism, which has the undesirable consequence that we cannot know the mental states of other people, since the body is different from the mind and the mind is supposed to be something internal, not accessible to external observers. What bothered Ryle and Wittgenstein, and what they wanted to deny, was primarily the idea that mental states are private entities, only knowable to the person who has them.[15]

But from what we have seen so far, Fodor would be in perfect agreement with the rejection of Cartesian dualism. When Fodor compares the mind to a computer, what he wants is precisely to bring the mind back to earth. Mental states, like anything else, must be explainable from a naturalistic point of view, and therefore without mysteries. The difference is that Wittgenstein, in attempting to bring the mind back to earth, puts it, so to speak, in behavior, stating, for example, that when we write, we think with our hands, or that when we speak, we think with our mouths and the larynx (see WITTGENSTEIN, 1958, p. 6). Fodor takes thoughts back to their natural place, namely, the mind, and proposes to remove the air of mystery that surrounds intentional mental states by treating them as involving symbols physically instantiated in the brain. Fodor observes (1975, p. 4) that behaviorists like Ryle seem to assume that the acceptance of mental states implies the acceptance of dualism, and therefore the acceptance of a conception of the mind as something private and mysterious. The only acceptable alternative would be behaviorism. But this is indeed a false dilemma, since RTM and CTM can be conceived as mentalist but non-dualistic theories (about substance). There is no reason to suppose that we have no way of knowing other people's mental states – a skeptic would certainly pressure Fodor at this point, but there is no reason to suppose that philosophy of mind should be guided by epistemological precepts (as it was commonly assumed).

5. Conclusion

We have seen that the representational and the computational theories of mind conceive mental states such as beliefs and desires as relations that individuals have to mental representations, and mental processes as algorithmic transformations of mental representations. In so doing, they preserve the essential properties that propositional attitudes have for common sense psychology, namely, semantic evaluability (and content) and causal efficacy. Behaviorism, on the other hand, opposes reference to mental processes and representations, conceiving mental states as behavioral dispositions. In so doing, it is not able to preserve the idea that intentional mental states have semantic content and causal efficacy. I illustrated some of the difficulties that behaviorism faces by considering Wittgenstein's approach to linguistic understanding. Against Wittgenstein, linguistic understanding is not just a matter of using language in a certain way, or of being disposed to behave in certain ways. It is instead part of what

causes or explains language use and behavior. In dispensing with mental processes to account for linguistic understanding, Wittgenstein is unable to preserve the causal efficacy that we ordinarily attribute to understanding. He moves a great deal away from common sense psychology.

As we've seen, part of what motivates Wittgenstein to avoid reference to mental symbols and processes is that they would introduce certain difficulties, such as, supposedly, an infinite regress. But, as Fodor says, "there are various things that you can usefully do when your car gets a ping in its cylinders; but declining to quantify over the engine is not among them. You need a story about the engine to explain how the car behaves." (1987, p. xi). Likewise, the semantic aspect of mental representations can be especially problematic and require an explanation, but eliminating or ignoring mental representations is not the best answer. With them we eliminate a relevant explanatory level, thus getting an incomplete understanding of ourselves and reality.

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86

It is worth adding that the practice of cognitive psychology has proven to be contrary to behaviorist precepts, since mental representations are assumed by most contemporary psychological theories.[16] As Fodor observes,

> The behaviorist has rejected the appeal to mental representation because it runs counter to his view of the explanatory mechanisms that can figure in psychological theories. Nevertheless, the science of mental representation is now flourishing. The history of science reveals that when a successful theory comes into conflict with a methodological scruple, it is generally the scruple that gives way. Accordingly the functionalist has relaxed the behaviorist constraints on psychological explanations. There is probably no better way to decide what is methodologically permissible in science than by investigating what successful science requires. (FODOR, 1981b, p. 123).

Thus, although behaviorist ideas and methodology have played a big role in psychological theories in the 20th century, it is now incompatible with the vocabulary of most current cognitive and social psychology and with the kinds of explanation they offer. Behaviorism, by denying causal powers to representational mental states, forces us to reconceive not only common sense psychological explanations but also widely held explanations in cognitive science. Representational and computational theories, on the other hand, are compatible with both common sense and cognitive science. This suggests that representational and computational theories of the mind are to be preferred when compared to behaviorism.

References

AYDEDE, M. The Language of Thought Hypothesis, *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), Disponível em: https://plato.stanford.edu/archives/spr2019/entries/language-thought/, 2010.

BOONE, W.; PICCININI, G. The cognitive neuroscience revolution. *Synthese* 193:1509–1534, 2016.

CAIN, M. J. *Fodor: Language, Mind and Philosophy*. Cambridge: Polity Press, 2002.

CRANE, T. *The mechanical mind: a philosophical introduction to minds, machines and mental representation* (2nd edition). Taylor & Francis e-Library, 2003.

DEVITT, M. *Ignorance of language*. Oxford: Oxford University Press, 2006.

DAWES, A. J. et al A cognitive profile of multi-sensory imagery, memory and dreaming in aphantasia. *Scientific Reports* 10, 10022, 2020.

FODOR, J. A. *The Language of Thought*. Cambridge, MA: Harvard University Press, 1975.

FODOR, J. A. Methodological solipsism considered as a research strategy in cognitive psychology In: *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press Bradford Books, 1981a.

FODOR, J. A. The Mind-Body Problem. In: *Scientific American*. 244:114-25, 1981b.
FODOR, J. A. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press Bradford Books, 1987.
FODOR, J. A. *Theory of content and other essays*. Cambridge, MA: MIT Press Bradford Books, 1990.
FODOR, J. A. *In critical condition*. Cambridge, MA: MIT Press Bradford Books, 1998.
FODOR, J. A. More peanuts. In: *London Review of Books*, Vol. 25, No. 19, pp. 16-17. 2003
HUTTO, D. D.; MYIN, E. *Radicalizing enactivism*. Cambridge, MA: MIT Press Bradford Books, 2013.
PLACE, U. T. Is consciousness a brain process? In: *British journal of psychology*, 47:1, pp. 44-50, 1956.
PUTNAM, H. Brains and behavior, in: Heil, J. (ed.) *Philosophy of Mind: a guide and anthology.* Oxford: Oxford University Press, 2004 (1963).
PUTNAM, H. Psychological Predicates, in: Heil, J. (ed.) *Philosophy of Mind: a guide and anthology.* Oxford: Oxford University Press, 2004 (1967).
QUILTY-DUNN, J; MANDELBAUM, E. Against dispositionalism: belief in cognitive science. *Philosophical Studies*, 175:2353–2372, 2018.
RYLE, G. *The concept of mind*. London: Penguin Books, 1990 (1949).
SEARLE, J. R. Minds, brains and programs. In: *Behavioral and Brain Sciences* 3 (3): 417-457, 1980.
SEARLE, J. R. *Intentionality.* Cambridge: Cambridge University Press, 1999 (1983).
STRAWSON, G. Consciousness isn't a mystery. It's matter. In: *The New York Times,* 2016.
WITTGENSTEIN, L. *Philosophical investigations*. Chichester: Wiley-Blackwell, 2009 (1953).
WITTGENSTEIN, L. *The blue book*. New York: Harper Torchbooks, 1965 (1958).
ZEMAN, A.; DEWAR, M.; DELLA SALA, S. Lives without imagery - Congenital aphantasia. *Cortex* 73, 378-380, 2015.

NOTAS

1 I would like to thank João Vergílio Gallerani Cuter, Evan Keeling, Plínio Smith and the anonymous reviewers for comments that helped to improve the paper. This research was supported by grant 2018/12683-9, São Paulo Research Foundation (FAPESP).

2 Mental states such as beliefs, desires, intentions, guesses, regrets, etc. are called propositional attitudes because it is generally assumed that they are different attitudes one can have to propositions. It is also common to call them intentional states, in the sense that they have intentionality, or are about something. I will here use "propositional attitudes" and "intentional states" interchangeably, and these are the states I have in mind when I say simply "mental states". Propositional attitudes are sometimes contrasted with mental states such as emotions and sensations, which appear to be essentially qualitative states.

3 Or, as Searle would say, intentional states such as beliefs and desires have conditions of satisfaction. For Searle, however, not all states with semantic content – which are characterized by being directed at or by being about objects or states of affairs in the world – are states that have conditions of satisfaction. Me being glad because my friend was awarded a prize, or upset because I insulted someone, as Searle notes, do not seem to be states that are satisfied or not by some relation to the world, in the way that a belief is satisfied if it is true, or not satisfied if it is false (cf. SEARLE, 1983, p. 8). What is important to note is that even if not all propositional attitudes can be said to be semantically evaluable, they all have semantic content, in the sense of being about something.

4 According to Aydede, "thinking is not proceeding from thoughts to thoughts in arbitrary fashion: thoughts that are causally connected are in some fashion semantically (rationally, epistemically) connected too. If this were not so, there would be little point in thinking—thinking couldn't serve any useful purpose. Call this general phenomenon, then, the *semantic coherence* of causally connected thought processes. LOTH [the language of thought hypothesis] is offered as a solution to this puzzle: how is thinking, conceived this way, physically possible?" (AYDEDE, 2010). Strictly speaking, though, it

is CTM that is offered to explain this (and not LOTH). While CTM presupposes a language of thought, the language of thought is not in itself sufficient to explain the semantic coherence of reasoning.

5 Galen Strawson describes, in my view aptly, the adherents of eliminativism by saying that they, by being "passionately committed to the idea that everything is physical, make the most extraordinary move that has ever been made in the history of human thought. They deny the existence of consciousness." (STRAWSON, 2016).

6 Perhaps both behaviorism and the identity theory can be considered realist theories of mental states, in the sense of not being eliminativists (of not simply denying the existence of mental states). According to these theories, mental states exist, they just are not exactly what we think they are; they are reduced to other types of entities: behavioral dispositions, for the behaviorist, and brain states, for the identity theorist.

7 This might not be an insurmountable problem for the identity theorist, since talk of neural representation is now common in cognitive neuroscience (see Boone and Piccinini, 2016). Insofar as brain states are conceived as being themselves representational, an identity theorist could argue for the view that brain states are the real bearers of semantic/representational properties of the propositional attitudes to which they are identical.

8 Place (1956) presents a similar argument to say that there is nothing conceptually problematic about the hypothesis that consciousness is a brain process. He makes it clear that, with this hypothesis, one does not want to give a definition of what "consciousness" means. That is, he is not saying that sentences about sensations are reducible or analyzable in terms of sentences about brain states. The "is" in the sentence "consciousness is a brain process" is what he calls an "is" of composition, analogous to the "is" in "lightning is a motion of electric charges."

9 Fodor says that in the context of an objection to Kim (FODOR, 1998, p. 13).

10 See Crane (2003) and Quilty-Dunn & Mandelbaum (2018) for further criticism of the idea that beliefs can be reduced to behavioral dispositions.

11 This view is famously developed in the *Philosophical Investigations* (1953) but since my purpose here is not to go into exegesis, I will concentrate on some ideas that can be extracted from *The Blue Book*.

12 As Fodor remarks, "if *anything* is clear it is that understanding a word (predicate, sentence, language) isn't a matter of how one behaves or how one is disposed to behave. Behavior, and behavioral disposition, are determined by the interactions of a variety of psychological variables (what one believes, what one wants, what one remembers, what one is attending to, etc.). Hence, in general, any behavior whatever is compatible with understanding, or failing to understand, any predicate whatever. Pay me enough and I will stand on my head iff you say 'chair'. But I know what 'is a chair' means all the same." (FODOR, 1975, p. 63).

13 It is possible that Wittgenstein would argue against the view that linguistic understanding involves conscious mental images based on introspection, claiming that we don't usually visualize a red flower when we hear and understand the request to bring a red flower. That would be fine, but using introspection to deny the more general claim that there are mental processes involved in understanding a sentence is a hasty move, since cognitive science works under the assumption that there are lots of unconscious mental processes that explain behavior.

14 Fodor does in fact give a similar answer when considering an objection against one of his arguments for the language of thought: "My view is that you can't learn a language unless you already *know* one. It isn't that you can't learn a language unless you already *learned* one." (FODOR, 1975, p. 65).

15 I do not intend to explore Wittgenstein's private language argument here. Suffice to say that Wittgenstein's concerns, as with most philosophers of the early twentieth century, are mainly epistemological. The private language that Wittgenstein finds impossible is a language whose words refer "to what only the speaker can know – to his immediate private sensations. So another person cannot understand the language" (1953, § 243). It is

at a minimum questionable that his observations apply to the language of thought, conceived as a construct accepted by cognitive psychologists. Someone who adopts the idea that the vehicle of thought is a language need not also adopt the very controversial thesis that its symbols refer to private experiences only knowable to the person who has them.

16 There is currently a movement in philosophy of mind called enactivism, represented by philosophers such as Daniel Hutto and Erik Myin (2013), that holds that mental representations are unnecessary to explain most behavior. They propose what they call a "radical view" about cognition, and assume that (basic) mentality should be understood by means of the "dynamic" interactions of an organism with its environment. On the face of it at least, this seems to be behaviorism in a new guise and, as such, subject to similar criticisms to those that Fodor and Putnam formulated decades ago. As we've seen, since behaviors are generally the effects of the interaction of intentional mental states, it is doubtful that behavior could be explained only in terms of the interactions of the organism with the environment, without reference to representational states.

AUFKLÄRUNG, João Pessoa, v.8, n.esp., Jul., 2021, p.71-86