The Journal of Mind and Behavior
Summer and Autumn 2020, Volume 41, Numbers 3 and 4
Pages 277–292
ISSN 0271–0137

Why Behaviorism and Anti-Representationalism Are Untenable

Markus E. Schlosser

University College Dublin

It is widely thought that philosophical behaviorism is an untenable and outdated theory of mind. It is generally agreed, in particular, that the view generates a vicious circularity problem. There is a standard solution to this problem for functionalism, which utilizes the formulation of Ramsey sentences. I will show that this solution is also available for behaviorism if we allow quantification over the causal bases of behavioral dispositions. Then I will suggest that behaviorism differs from functionalism mainly in its commitment to anti-representationalism, and I will offer two new objections to anti-representationalism. The first will be based on considerations concerning the contents of desires and intentions. The second objection concerns inner speech and mental imagery. We will see that the objections are of relevance to contemporary debates, as they apply with equal force to the currently popular anti-representationalist versions of embodied and enactive cognition.

Keywords: behaviorism, Ramsey sentences, anti-representationalism

The philosophy of mind literature distinguishes between two main types of traditional behaviorism: philosophical and psychological behaviorism (also known as analytical and methodological behaviorism).[1] Both views are usually presented as relics of the past, and the implicit message is often that they are still being taught only so that we can learn from their failure (Graham, 2019; Heil, 2012; Kim, 2010). It is certainly true that psychological behaviorism is widely considered to be hopeless as a general theory of mind, but there are some remnants of this view to be found in many areas of empirical psychology. Most importantly, it is still often held or implicitly assumed that phenomena that can be explained in behavioristic terms should be explained in this way — in particular, they should be explained, if possible, without the ascription of representational mental states.

I would like to thank Dr. Russ (Editor) and an anonymous referee for their very helpful comments on an earlier draft. Correspondence regarding this article should be sent to Dr. Markus Schlosser, School of Philosophy, University College Dublin, Newman Building, Belfield, Dublin 4, Ireland. Email: markus.schlosser@ucd.ie

[1] Philosophical behaviorism is most closely associated with the work of Gilbert Ryle (especially Ryle, 1949; see also Hempel, 1949). Sympathizers included Rudolf Carnap, Ludwig Wittgenstein, and W. V. O. Quine. The most important psychological behaviorists were Ivan Pavlov, John Watson, and B. F. Skinner.

Likewise, it is widely agreed that traditional philosophical behaviorism is an untenable theory of mind. But the more recent past has seen the emergence of various views that are in line with philosophical behaviorism in the sense that they propose accounts of embodied and enactive cognition that do not require the ascription of mental representations (Chemero, 2009; Hutto and Myin, 2013; Thompson, 2007; van Gelder, 1995; Varela, Thompson, and Rosch, 1991). My focus here is primarily philosophical behaviorism and its anti-representationalist dimension. I will return to psychological behaviorism briefly in the conclusion.

Philosophical behaviorism faces many objections and challenges. One of these objections is commonly singled out as the most devastating problem for the view. This is the problem of circularity. I will argue that this problem admits of a relatively straightforward solution if we make use of some of the insights that have emerged from the more recent debate on the nature of dispositions. If we refer to the causal bases of behavioral dispositions, we can solve, as I will argue, the circularity problem in the same way as functionalism does. This will force us to get clearer about the difference between behaviorism and functionalism, and it will lead us to the second part, in which I present two new objections to philosophical behaviorism. The first objection will be based on considerations concerning the contents of desires and intentions. In the second objection I argue that behaviorism cannot explain inner speech and mental imagery. It will become clear that those objections apply with equal force to the currently popular anti-representationalist versions of embodied and enactive cognition, and I will suggest that they point to clear limits for neo-behavioristic aspirations in psychology and cognitive science.

Before we turn to philosophical behaviorism, let me add a note on the overall dialectic, which may seem rather odd, at first. In the first part I defend behaviorism against the circularity objection, and in the second part I reject the view. As indicated, there is reason in this apparent incoherence. The reply to the circularity objection will bring into focus the key difference between behaviorism and functionalism. This will take us to my objections, which are objections that apply to all forms of anti-representationalism.

Philosophical Behaviorism and the Circularity Problem

Let me begin with an outline of traditional philosophical (or analytical) behaviorism (henceforth *behaviorism*, for short). This outline is based on the received view, as it is presented in standard textbooks (Graham, 2019; Heil, 2012; Kim, 2010). There are two main components of the view. The first is an analytical component that concerns the meaning of mental terms. It says, roughly, that statements about mental phenomena can be analyzed in terms of (or reduced to) statements about behavior and behavioral dispositions. The second component is ontological. It concerns the nature of mental states. It says, again roughly, that having certain mental states consists in the possession of behavioral dispositions and their manifestations in overt behavior. The analytical component is supposed

to yield the ontological component: when mental terms are analyzed in behavioristic terms we can see that the possession of mental states can be reduced to the possession and manifestation of behavioral dispositions.

To illustrate, let us consider two stock examples. It seems, for instance, that being in pain is typically accompanied by certain types of expressions, such as wincing, moaning, and by actions that lead to a cessation or alleviation of the pain (avoidance behavior). Behaviorism suggests that "being in pain" can be analyzed entirely in terms of such behaviors and behavioral dispositions, and that we have therefore no reason to think that the mental state of being in pain involves anything else than that. Or, it seems plausible to assume that what people believe is strongly correlated with what they are disposed to assert. Behaviorism suggests that the mental state type "belief that p" can be analyzed in terms of assertions and dispositions to assert, and that mental states of belief are nothing over and above such behaviors and behavioral dispositions.

At the time when behaviorism was developed and defended, it was commonly thought that dispositional properties, including behavioral dispositions, could be analyzed in terms of counterfactual conditionals. Behaviorism is still usually presented in this way, even though the conditional analysis is now widely rejected (more on this below). Assuming this analysis for now, a first approximation to an analysis of belief may take the following form:

S believes that p if and only if S would assert that p if S were asked about p.

There are, of course, obvious problems with that. What if the agent does not want to say or is afraid to say what is believed? What if the agent intends to express the belief without being asked? What if there is some misunderstanding or miscommunication? And so on.

Note, first, that one cannot avoid those issues by rejecting the conditional analysis. Nowadays, the conditional analysis of dispositions is widely rejected, mainly due to various counterexamples. Consider, for instance, a sorcerer who would change the intrinsic properties of a fragile glass if it were to be struck, so that it would not break when struck. Or suppose that the sorcerer would cover the fragile glass with protective coating if it were to be struck, so that it would not break when struck. The former is an example of a "finkish" disposition (Lewis, 1997), the latter is an example of a "masked disposition" (Johnston, 1992). Such examples show that the truth of the relevant counterfactual is not necessary. Reverse cases, sometimes called "mimics," show the opposite. Assume, for instance, that the sorcerer would turn a paper cup into a fragile cup whenever it was about to be struck. This appears to be a counterexample to the claim that the truth of the relevant counterfactual is sufficient (for more on this see Fara, 2005; Manley and Wasserman, 2008).

Various lessons can be drawn from this. One may suggest that this shows that dispositions cannot be explained reductively, and that it simply has to be accepted as a brute fact that dispositions may be finkish or masked in various ways. In light of

this, one may suggest that behaviorism can appeal to dispositions without analyzing them as conditionals, and that there is, therefore, no need to specify all the possible defeating conditions. Given this, the proposal may be rendered as follows:

S believes that p if and only if S is disposed to a sert that p when S is asked about p.

On this suggestion, the mentioned complications are not counterexamples. No claims are being made about what the agent *would* do, but only about what the agent is *disposed* to do. Suppose that, under certain circumstances, the agent does not and would not say what is believed. It may neverthewless be true that the agent is disposed to say what is believed (even in those circumstances).

This strategy may help to avoid apparent counterexamples, but it does not get to the bottom of the issue. The mentioned complications are not merely potential counterexamples to behavioral analyses. Rather, they highlight the fact that the nature of belief is more complex than the proposed analyses suggest. Consider the case in which the agent does not want to say what is believed. This points not merely to a potential counterexample, but to a connection between belief and desire that is essential to the nature of both belief and desire. The same holds for other examples. Certain desires may not merely prevent the manifestation of a pain, for instance. But the connections between desire and pain are essential to the nature of both desire and pain. This means that a rejection of the conditional analysis and endorsement of non-reductionism about dispositions does not help here.

In general, the way in which a mental state is connected to other mental states does not only generate potential counterexamples to behavioral analyses, but it is an essential part of that mental state's identity. This takes us to the main problem for philosophical behaviorism: the circularity problem (Chisholm, 1957; Putnam, 1965; see also Heil, 2012, pp. 65–67; Kim, 2010, pp. 71–78). A satisfactory analysis of pain would have to capture the (actual or counterfactual) connection with the belief that pain should be suppressed, the fear that the available avoidance behavior would lead to even greater pain, and so on. A satisfactory analysis of belief would have to capture the connection with the desire not to say what is believed, the intention to say what is believed without being asked, and so on. Any analysis that aims to capture the nature of those mental states has to capture those connections. And so any such analysis cannot be a purely *behavioral* analysis, because it would have to make reference to other *mental* terms — in other words, any such analysis would be circular, and quite obviously so.[2]

[2] Kim (2010, p. 170) points out that this problem can be construed either as a circularity or as a regress problem. There is an indefinite number of external and internal conditions that may prevent the manifestation of any mental state. Given this, behaviorism faces a *regress* problem: any attempt to give a behavioral definition of a mental state leads to an open-ended list of more and more qualifications that specify all the relevant external and internal conditions. As these qualifications would inevitably include reference to other *mental* states, this is also a *circularity* problem. The formulation in terms of circularity highlights the more fundamental problem, because it shows that the characterization of the nature of any mental state requires reference to other mental states.

As indicated, it is widely agreed that this circularity problem is the most devastating problem for philosophical behaviorism. One reason for this is that it is an *internal* problem that seems to show that behaviorism fails in its own terms — it seems to show that behavioral analyses are impossible. In the following sections, I will propose a solution to this problem for which I need two main ingredients: the solution to the circularity problem for functionalism and some more of the insights from the recent debate on the nature of dispositions.

The Solution to the Circularity Problem for Functionalism

According to standard functionalism, mental state types can be analyzed in terms of their functional roles, which are the causal roles that are occupied or realized by the relevant agent-internal states (Heil, 2012; Kim, 2010; Levin, 2018). It is usually assumed that those functional roles include causal connections between sensory input, internal states, and behavioral output. To use again the example of pain, one may begin with the suggestion that the typical causal input of a pain state is damage to the body, that the typical outputs are behaviors such as wincing and moaning, and that the typical connections with other internal states include a desire to alleviate the pain and a belief about how to do so. This functionalist approach seems to face the circularity problem as well, as it makes explicit reference to other mental states. Like behaviorism, functionalism claims to have an explanation of the nature of mental states. It allows reference to internal states, and so it must explain why certain internal states are *mental* states. If the characterization of a mental state always requires reference to other mental states, functionalism presupposes what it seeks to explain (see Heil, 2012, pp. 100–101; Kim, 2010, p. 170). Unlike behaviorism, functionalism has a standard solution to this circularity problem that is based on the formulation of Ramsey sentences (Lewis, 1972; see also Kim, 2010, pp. 170–172; Levin, 2018).[3] Here is how this would work for the example of pain. Quantifying over three internal states, x_1, x_2, and x_3 (for the pain state, the desire, and the belief), we first formulate the functional role of pain (FRP) as follows:

> $\exists x_1 \exists x_2 \exists x_3$ (x_1 tends to be caused by damage to the body, x_1 tends to cause wincing and moaning, x_1 tends to cause x_2 and x_3).[4]

[3] A Ramsey sentence renders a theoretical proposition free of terms that imply ontological commitments by replacing them with variables that are bound by existential quantifiers. For instance, the claim that electrons have the properties P and Q has the following Ramsey sentence: there is some x such that x has P and Q; in standard logical notation: $\exists x$ (Px & Qx). This formal method was used by logical positivists in their verificationist project to avoid reference to unobservable entities and to separate scientific theories from metaphysics. The details concerning their philosophy of science are not relevant here. Lewis (1972) was the first to use this method in order to show that mental states can be defined in terms of their functional roles without circularity. For this reason, Kim (2010, pp. 170–172) calls this procedure the Ramsey–Lewis method.

[4] For those unfamiliar with logical notation: there are three internal states, x_1, x_2, and x_3, such that x_1 tends to be caused by damage to the body, x_1 tends to cause wincing and moaning, x_1 tends to cause x_2, and x_3.

Or, for short:

$\exists x_1 \exists x_2 \exists x_3$ FRP (x_1, x_2, x_3).

The internal state x_1 is supposed to occupy the functional role of pain, and we want an analysis of what it is for someone, some agent, to be in a pain state. We get this from the following sentence for "being in pain," where now x_1, x_2, and x_3 refer to internal states of the agent S:

S is in pain if and only if $\exists x_1 \exists x_2 \exists x_3$ (FRP(x_1, x_2, x_3) & S is in x_1).

Ramsey sentences provide implicit definitions of mental states that make no explicit mention of other mental states. They quantify over *internal* states that are internal *mental* states in virtue of playing or occupying the right functional roles, which are the causal roles specified by the Ramsey sentence. This solves the circularity problem for functionalism, and it is generally agreed that this solution is not available to behaviorism. Behaviorism rejects reference to internal mental states, and so it seems clear that it cannot admit reference to internal states that are said to occupy the causal roles of internal mental states.

Dispositions

As pointed out, the conditional analysis of dispositions is now widely rejected, mainly due to various counterexamples. In response to this, some have proposed revised conditional analyses, others have pursued non-reductive and robustly realist approaches to the nature of dispositions. For our purposes, the important point here is only that it has become very common in this debate to make reference to a disposition's *causal basis*. Generally speaking, a disposition's causal basis is an intrinsic property of the object or agent that would interact with the stimulus condition so as to bring about the manifestation of the dispositional property. In the case of a fragile glass, for instance, this would be the particular molecular structure in virtue of which the glass is fragile. In other words, we can say that the causal basis is an intrinsic property that grounds the ascription of the disposition. In what follows, I will assume that behavioral dispositions have causal bases: intrinsic properties of the agent that ground the ascriptions of the behavioral dispositions in question.

Before we can turn to the solution to the circularity problem, we need to consider a few more clarifications. We assume now that behavioral dispositions have causal bases. Often, it is assumed that the causal basis of a disposition is an intrinsic property of the object or agent as a whole. Note, though, that it may also be an intrinsic property of one of the object's or agent's parts. Consider, for instance, a wine glass that has a thick and sturdy base. We say that the glass is fragile, as a whole, even if we think that its base is not fragile. Or consider someone's

disposition to fall asleep when tired. This is a dispositional property of the person, as a whole, but it seems clear that the causal basis of this disposition consists of the intrinsic properties of various parts. The physical composition of the person's toes, for instance, does not explain why the person has the disposition, and yet we attribute the disposition to the person as a whole. Further, it seems clear that an agent's internal state just is a property of one of the agent's parts. When we say, for instance, that someone is in a certain neural state we mean that a part of the agent's brain has a certain property (currently or for the time being). We can assume, then, that ascriptions of behavioral dispositions are grounded either in intrinsic properties of the agent (as a whole) or in intrinsic properties of the agent's internal states (construed as properties of parts).

Note here that we may attribute a behavioral disposition to the person as a whole even if we think that it is grounded in intrinsic properties of some of the agent's internal states. The agent as a whole has the disposition to fall asleep when tired, yet it is clear that the causal basis of this disposition consists of intrinsic properties of some of the agent's internal states (construed as properties of parts).

Finally, it is important to note that dispositions may interact or interfere with each other due to the interaction of their causal bases. Consider, for instance, someone's disposition to be hyper-vigilant and nervous when too much caffeine is consumed. It seems clear that the causal basis of this disposition may interact with the causal basis of that person's disposition to fall asleep when tried, so that the former may interfere with (or mask) the manifestation of the latter — he may not fall asleep even though he is very tired, because he drank too much coffee. Again, we make attributions to the person as a whole, but we assume that those attributions are grounded in interactions between internal states (properties of parts).

The Solution to the Circularity Problem for Behaviorism

With those components and clarifications in place, we can now see that the circularity problem for behaviorism has a straightforward solution. Consider once more the Ramsey sentence for "being in pain":

S is in pain if and only if $\exists x_1 \exists x_2 \exists x_3$ (FRP(x_1, x_2, x_3) & S is in x_1).

Now note that x_1, x_2, and x_3 may refer to intrinsic properties or internal states that are the *causal bases of behavioral dispositions.* As noted in the previous section, behavioral dispositions of the agent may be grounded in internal states, construed as properties of parts. Given this, we can see that the standard Ramsey sentence solution to the circularity problem can be deployed in the service of philosophical behaviorism.

This may seem too good to be true. At least one would expect an explanation of why this has not been noted before. I can suggest the following. An application of the Ramsey sentence method requires reference to internal states that can causally

interact and interfere with each other. The conditional analysis of dispositions does not enable this. It attributes dispositions to the agent as a whole, and it does not include reference to the disposition's causal basis. It does not enable us to see that an agent's behavioral dispositions may be grounded in internal states, and it does not enable us to see how an agent's behavioral dispositions may interact and interfere with each other. Discussions of behaviorism have either assumed the conditional analysis or they have ignored questions concerning the nature of dispositions. Once we reject the conditional analysis and make reference to causal bases that are grounded in internal states, we can see that there is nothing that prevents the application of the solution to philosophical behaviorism.

Behaviorism and Functionalism

Functionalism is often presented by means of the Ramsey-sentence method. I have just argued that the very same Ramsey sentences can be deployed by behaviorism. Does this mean that behaviorism collapses into functionalism? Would this not be a *reductio* of the proposed solution? For, clearly, behaviorism and functionalism *are* distinct theories.

The application of the Ramsey-sentence method to behaviorism does not collapse the view into functionalism. Rather, it helps us to bring the difference between them into sharper focus. As we have seen, behaviorism can make reference to internal states, but only if those states are the causal bases of *behavioral* dispositions. Functionalism is not limited by this. According to functionalism, the relevant internal states may be, and typically are assumed to be, internal mental states, construed as internal states that are the carriers or vehicles of representational content. Behaviorism categorically rejects reference to such internal representational states. So, the key difference does not concern the reference to *internal* states. As I have argued, reference to internal states does not mark a difference at all. Rather, the difference and the disagreement concerns the reference to internal *mental representations*: functionalism allows and endorses it, behaviorism rejects it. According to behaviorism, mental states and contents are to be ascribed to the whole person and in virtue of behavioral dispositions of the whole person. It can make reference to internal states only insofar as the ascription of behavioral dispositions to the whole person may be grounded in internal states. Functionalism, in contrast, ascribes mental states and contents to internal states, construed as the internal vehicles of mental representations.

Two Objections to Behaviorism and Anti-Representationalism

According to the proposed diagnosis, the key difference between the two views concerns mental representation: functionalism is a form of wholehearted representationalism and behaviorism takes a decidedly anti-representationalist position. This takes us to the two new objections to behaviorism that I will

put forward in this section. It will become clear that they apply also to current anti-representationalist versions of embodied and enactive cognition, as they object to the anti-representationalist dimension of those views — they object to anti-representationalism in general. The first objection concerns the contents of desires and intentions, the second concerns inner speech and mental imagery.

The Content of Desires and Intentions

Approaching this objection, let us first consider the contents of beliefs. At first, it seemed relatively straightforward to capture the contents of beliefs in terms of behavioral dispositions to assert. The main problem with this suggestion was that we needed to make reference to other mental states. Functionalism uses the formulation of Ramey sentences to solve this problem. If my argument so far is correct, then behaviorism can use the same solution, and so it seems that behaviorism can give an account of belief that is as good as the functionalist account.

Can this move be applied to the analysis of desire? Functionalism provides a standard analysis of desires in terms of dispositions to pursue goals and beliefs about how to achieve those goals. Ramsey sentences can be used to show that one can make reference to beliefs in the analysis of desire without circularity, and so desires and beliefs are said to be inter-definable on this analysis. Given that behaviorism can use the Ramsey-sentence method, it may seem that this standard analysis should now be available for behaviorism as well. But here we encounter a serious problem for the view.

Reference to beliefs is often required because we often desire a certain *outcome* or goal rather than the performance of a particular *action*. In other words, often we do not perform actions for their own sake, but in order to bring something about or pursue a certain goal. Beliefs provide the necessary information about how to bring about the desired outcome or goal. The crucial point is that such desires are directed at outcomes or goals, not at actions (or act-types). In such cases, the action is specified by the content of the belief as a means to an end, and the desire is not directed at a particular action or type of behavior, but at a future outcome or goal. This means that the content of such desires cannot be captured in terms of behavioral dispositions. It is, at least, very difficult to see how this could be done. Such desires are not directed at particular actions or types of behavior. So how could their content be captured in terms of dispositions to perform particular actions or types of behavior? Functionalism does not face this problem, because functionalism allows reference to mental representations. If we assume that desires are representational mental states, we can construe goal-directed desires simply as mental states that represent the outcome or goal in question. This points to a serious limitation of behaviorism, in comparison with functionalism. Some desires are directed, directly, at actions — we perform some actions for their own sake. But most of the things we do, we do in order to bring

something else about. You press certain keys on your computer in order to write a sentence or browse the internet. You say certain things in order to communicate a certain message. And so on. It seems quite clear that most of our desires are directed at abstract outcomes or goals, and it seems that this requires reference to the mental representation of outcomes or goals.

Note that this goes beyond the circularity objection. Now that we can make use of Ramsey sentences, the problem is not any more that the analysis of desires requires reference to beliefs. Rather, the problem is that the content of most desires is simply too abstract for an analysis in terms of behavioral dispositions — it abstracts from the pursuit of particular actions to the pursuit of outcomes and goals. This problem becomes even more obvious when we turn to the contents of intentions.

Like some desires, some intentions have concrete or specific contents in the sense that they are directed at particular actions (or act-types). Consider, for instance, the intention to drink some water (right now), to go out for lunch with a friend, to walk home instead of taking the train, and so on. But the contents of many intentions are more abstract and often also rather vague. The most obvious examples are provided by intentions to pursue long-term plans and projects. Consider, for instance, the intention to learn to play the piano, to move to another town, to do a Master's degree in law, to go on vacation some time in June, and so on. Such intentions are not directed at particular actions or types of behavior. At least in the early stages of intending and planning, the contents of such intentions are still vague and abstract and they are not yet directed at particular actions or types of behavior. Further decision-making is required before the agent arrives at a plan that specifies which actions to take in the pursuit of the goal.

The problem here is not just that the agent has not yet decided which particular actions to take. The problem is that there are no particular types of behavior that could ground the ascription of the content — neither now nor later in the planning and decision-making process. There is no such thing as the behavioral disposition to move to another town, or to do a Master's degree in law, because there are no particular types of behavior that correspond to those contents. In more technical terms, moving to another town, or doing a Master's degree in law, are not *basic* actions. One needs to figure out what to do before one can begin to take action and before one can be disposed to take action.

One might think that clusters of behavioral dispositions can provide the basis for the ascription of such contents. A cluster of behavioral dispositions would correspond to a cluster of particular actions that one intends to take in pursuit of the goal. The problem with this suggestion is that, in the early stages, the agent may not have yet any plan or idea about how to pursue the goal. At first, the intention may simply be to move to another town, or to do a Master's degree in law. The intention may be formulated at this abstract level only, and it seems clear that contents at this abstract level cannot be captured by reference to behavioral

dispositions. This may seem less clear for cases such as learning to play the piano. But even in this case, if one sits down at a piano only with the intention to learn to play, nothing will happen. In order to get started, the content of the intention must first be specified in more detail — one must develop some plan that specifies concrete actions.

Note, further, that Ramsey sentences are of no help here. The use of the Ramsey-sentence method facilitates reference to other mental states such as beliefs. But, in the early stages of the planning process, the agent may not yet have acquired any beliefs about what actions to take in order to pursue and bring about the intended end. Further deliberation and planning may be required before any particular types of behavior come into view, and before the agent can form concrete beliefs about how to pursue the goal. Until particular actions come into view, there is simply no way in which the contents of such intentions can be explained in terms of behavioral dispositions alone.

We can conclude that desires and intention with abstract contents cannot be analyzed in terms of behavioral dispositions. This problem arises for behaviorism, because behaviorism rejects reference to representational mental states. The view's anti-representationalism is the source of the problem, and so it is clear that this is a problem for anti-representationalist views in general, including the anti-representationalist versions of embodied and enactive cognition. Functionalism, in contrast, does not face this problem, because it makes reference to representational mental states.

Inner Speech and Mental Imagery

Inner speech and mental imagery are very familiar subjective experiences. Suppose I ask you to think about what you did yesterday (without saying anything out loud). Your thinking about yesterday will largely consist of inner speech and mental imagery. You will probably think of some descriptions of what you did, which are likely to be experienced as speech acts that you hear in your head, as it were. And you will probably think of some of the things that you did by recalling some of the situations you encountered, most likely by generating visual mental imagery. Examples can easily be multiplied. The average person spends a considerable amount of time in daydreams or with a mind wandering; and daydreaming and mind wandering consist largely of inner speech and mental imagery. Deliberation about future actions and plans also seems to consist largely of inner speech and mental imagery.

In general terms, inner speech can be defined, roughly, as the subjective experience of language in the absence of overt and audible articulation (Alderson–Day and Fernyhough, 2015). Mental imagery can be characterized as quasi-perceptual experience: subjective experience that resembles perceptual experience but occurs in the absence of the typical external stimuli (Thomas, 2014).

I take it that the reality of inner speech and mental imagery is undeniable. They are clearly part of our mental lives, and a general theory of mind must provide the basic resources to explain them. It need not give a detailed account, and if it is a philosophical theory it need not give an empirically informed account. But the ontology of a general theory of mind must be equipped to accommodate the reality of inner speech and mental imagery. I will argue that behaviorism and anti-representationalism are to be rejected because they do not provide the basic resources to accommodate inner speech and mental imagery.

Let us begin with inner speech. How might behaviorism account for it? Philosophical behaviorism holds that mental phenomena can be explained in terms of behavioral dispositions, and it says that those dispositions need not be manifested in overt behavior. For example, consider the proposal that the belief that p can be analyzed in terms of a disposition to assert that p in speech acts (in conjunction with further conditions that specify defeating conditions). Having the belief consists primarily in the possession of this disposition, and so having the belief does not require that it is manifested in overt speech. Can we construe inner speech as the manifestation of this disposition in *inner* speech? The basic idea would be that the disposition to assert that p may be manifested either in overt or in inner speech.

It is difficult to see how we can make sense of this within a behaviorist framework. Overt speech is a manifestation in behavior. What could the manifestation of a belief in *inner* speech possibly be? Within the behaviorist ontology, *where* would the belief manifest when it manifests in inner speech? Behaviorism admits only two basic categories: dispositions and their manifestations. An inner speech act is not itself a disposition. It is a manifest phenomenon: it is event-like in that it occurs. Further, it seems only plausible to construe inner speech acts as manifestations, as it seems that they are the manifestations or tokenings of beliefs that may also manifest in overt speech. But what does the belief manifest *as*, when it manifests in inner speech? What is the ontological category of the manifestation? Obviously, it does not manifest as overt behavior. What else could it be? As far as I can tell, behaviorism cannot provide an answer. Behaviorism lacks, fundamentally, the ontological resources to accommodate the reality of inner speech. Functionalism, in contrast, provides a straightforward answer: inner speech acts are tokenings of mental representations. Representationalism enables the answer. Behaviorism has no answer due to its commitment to anti-representationalism.

The point generalizes to other forms of anti-representationalism. Contemporary versions of anti-representationalism, as prominent among the proponents of embodied and enactive cognition, often appeal to the notion of re-enactment in their explanations of cognition and agency. One may suggest that inner speech acts are the re-enactment (and perhaps reconfiguration) of overt speech acts. But this reframing in terms of re-enactment is of no help at all. It generates the same obvious problem. What are inner speech acts re-enacted *as*? What is the ontological category of the inner speech act? It is not an overt speech act. What else

could it be within an anti-representational framework? It is, of course, tempting to say that an inner speech act is a re-enactment *in the mind*. But what does that mean? What that means is precisely what is in need of explanation here. According to representationalism, the re-enactment can be construed as the tokening of an internal mental representation. This answer is not available to any version of anti-representationalism. The problem is essentially the same as before. The proposal does not provide the resources to accommodate the reality of inner speech, and the root of the problem is the commitment to anti-representationalism.

Mental imagery generates the same problem. To make this as clear as possible, let me consider another move that is available to the proponents of anti-representationalism. It has been suggested that mental imagery can be construed as the re-enactment (or reconfiguration) of direct perception (Degenaar and Myin, 2014). I will not try here to provide an account of direct perception — we can leave that to the proponents of embodied and enactive cognition. Suffice it to say that direct perception involves the embodied interaction with the object of perception rather than the formation of an internal mental representation.

The proposal is puzzling. How can direct perception be re-enacted at all? On the face of it, it seems that direct perception requires the presence of what is perceived, and so it seems that it cannot be *re*-enacted as *direct* perception. And if it is not re-enacted as direct perception, we need to know what it is re-enacted as. This takes us straight back to the problem identified above. It is clear that mental imagery is a manifest phenomenon. Within the behaviorist ontology, it would have to be construed as the manifestation of a disposition. It is clearly not manifested as overt behavior. What else could it be? Behaviorism has no answer to this question. It lacks the ontological resources to accommodate the reality of mental imagery.

As before, the root of the problem is the commitment to anti-representationalism and so the problem generalizes. Anti-representationalism cannot accommodate mental imagery because it fails, fundamentally, to provide an appropriate ontological category. Concerning the appeal to the re-enactment of direct perception, it remains a mystery what the re-enactment of direct perception is supposed to be re-enacted as. It cannot be overt behavior. It cannot itself be direct perception (as the perceptual stimulus is not present). And it cannot be the tokening of a mental representation. What type of entity is the re-enactment supposed to be? As far as I can tell, anti-representationalism does not and cannot provide an answer.

To sum this up, inner speech acts and mental imagery are manifest phenomena. Within the metaphysical framework of behaviorism, they would have to be construed as manifestations of dispositions. They are not manifestations in behavior, and behaviorism has no alternative answer as to what those manifestations might be. In particular, behaviorism cannot make reference to tokenings of mental representations, which would be an obvious candidate answer to the metaphysical question. Essentially the same challenge arises for the more recent anti-representationalist proposals. Concerning the suggestion that inner speech and mental

imagery are to be construed in terms of re-enactment, we need to know the metaphysical category of the manifestation. Like behaviorism, recent versions of anti-representationalism lack the resources to answer the metaphysical question.

Note that there are two dimensions or components to this objection. First, it is suggested that behaviorism and recent variants of anti-representationalism are to be rejected because and insofar as they do not provide answers to obvious and fundamental questions. Second, it is suggested that an explanation of inner speech and mental imagery requires the ascription of mental representations. In other words, inner speech and mental imagery raise a general challenge to anti-representationalism because they seem to be "representation-hungry" phenomena (Clark and Toribio, 1994).

Conclusion

I have argued that, contrary to common opinion, traditional philosophical behaviorism does not have a circularity problem. If we borrow the Ramsey-sentence solution from functionalism, and if we allow reference to the causal bases of behavioral dispositions, we can see that the problem has a surprisingly straightforward solution. This move brought into focus the question of what the difference is between behaviorism and functionalism. I suggested that the main difference consists not in the reference to internal states, but in the reference to internal mental representations: internal states that are assumed to be the vehicles of representational content. In the second part, I have argued that this commitment to anti-representationalism generates fundamental problems. I considered the nature of desire, intention, inner speech, and mental imagery to make this point. We have seen, moreover, that the same problems arise for anti-representationalism in general, including the forms of anti-representationalism that are currently popular in the research on embodied and enactive cognition.

Let me close with a few remarks on psychological behaviorism. I mentioned that one behavioristic imperative is still widely accepted in psychology and cognitive science. That is, roughly, the imperative to explain as much as possible without the ascription of mental representations. Note that this does not have to be construed a behavioristic imperative, because it can also be motivated and justified on the ground of ontological parsimony. Further, it can be said in favor of this imperative that it encourages a pluralistic approach to explanation and to unexplored theoretical and experimental avenues. This seems to be a good thing — the more the merrier! However, the presented objections to anti-representationalism should put this pluralistic enthusiasm into perspective. It seems that a very great deal of human agency is in some way influenced and guided by desires and intentions and by other mental states such as inner speech and mental imagery. Indeed, it seems difficult to find clear instances of human agency that are altogether isolated from such influence and guidance. I granted that some desires

and intentions may admit an analysis in behavioristic terms. But I have argued that the contents of most desires and intentions will be too abstract or too vague to allow a reduction to specific behavioral dispositions. Moreover, even simple and specific acts are often constrained by longer-term desires and intentions that are operative in the background. Consider, for instance, going for a walk or whistling a tune just because one feels like doing so. It may seem that such actions can be explained in terms of simple and specific desires. But we usually do not go for a walk or whistle whenever we feel like doing so. The execution of such actions tends to be constrained by beliefs and by longer-term intentions (such as intentions about how to structure the day or standing beliefs about appropriate behavior). The heuristic to pursue non-representationalist explanations has delivered some interesting proposals and opened up new possibilities. But it remains to be seen how many, if any, types of human agency and cognition can be explained without any reference to mental representations (see Schlosser, 2018). I have argued that desires and intentions with abstract contents require the ascription of mental representations, and I have argued that the explanation of inner speech and mental imagery does so as well. Given that most of our agency is in some way influenced, guided, or constrained by desires and intentions with abstract contents, and given that a great deal of cognition involves inner speech and mental imagery, the prospects for general anti-representationalist accounts of the mind seem rather bleak, and the prospects for anti-representationalist explanations of particular capacities appear to be rather limited.

References

Alderson–Day, B., and Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*, 931–965.

Chemero, T. (2009). *Radical embodied cognitive science.* Cambridge, Massachusetts: MIT Press.

Chisholm, R. M. (1957). *Perceiving.* Ithaca, New York: Cornell University Press.

Clark, A., and Toribio, J. (1994). Doing without representing? *Synthese, 101,* 401–431.

Degenaar, J., and Myin, E. (2014). Representation-hunger reconsidered. *Synthese, 191*, 3639–3648.

Fara, M. (2005). Dispositions and habituals. *Noûs, 39,* 43–82.

Graham, G. (2019). Behaviorism. *The Stanford encyclopedia of philosophy.* Retrieved from https://plato.stanford.edu/entries/behaviorism/

Heil, J. (2012). *Philosophy of mind: A contemporary introduction* (third edition). New York: Routledge.

Hempel, C. (1949). The logical analysis of psychology. In H. Feigl and W. Sellars (Eds.), *Readings in philosophical analysis* (pp. 373–84). New York: Appleton–Century–Crofts.

Hutto, D., and Myin, E. (2013). *Radicalizing enactivism: Basic minds without content.* Cambridge, Massachusetts: MIT Press.

Johnston, M. (1992). How to speak of the colors. *Philosophical Studies, 68,* 221–263.

Kim, J. (2010). *Philosophy of mind* (third edition). Boulder: Westview Press.

Levin, J. (2018). Functionalism. *The Stanford encyclopedia of philosophy.* Retrieved from https://plato.stanford.edu/entries/functionalism/

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50,* 249–258.

Lewis, D. (1997). Finkish dispositions. *Philosophical Quarterly, 47,* 143–158.

Manley, D., and Wasserman, R. (2008). On linking dispositions and conditionals. *Mind, 117*, 59–84.

Putnam, H. (1965). Brains and behavior. In R. Butler (Ed.), *Analytical philosophy: Second series* (pp. 1–19). Oxford: Blackwell.
Ryle, G. (1949). *The concept of mind.* London: Hutchinson.
Schlosser, M. E. (2018). Embodied cognition and temporally extended agency. *Synthese, 195,* 2089–2112.
Thomas, N. J. T. (2014). Mental imagery. *The Stanford encyclopedia of philosophy.* Retrieved from https://plato.stanford.edu/entries/mental-imagery/
Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
van Gelder, T. (1995). What might cognition be if not computation? *Journal of Philosophy, 92, 345–381.*
Varela, F., Thompson, E., and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, Massachusetts: MIT Press.