



# Entity Realism About Mental Representations

Bence Nanay<sup>1,2</sup> 

Received: 27 May 2019 / Accepted: 3 October 2019 / Published online: 26 October 2019  
© The Author(s) 2019

## Abstract

The concept of mental representation has long been considered to be central concept of philosophy of mind and cognitive science. But not everyone agrees. Neo-behaviorists aim to explain the mind (or some subset thereof) without positing any representations. My aim here is not to assess the merits and demerits of neo-behaviorism, but to take their challenge seriously and ask the question: What justifies the attribution of representations to an agent? Both representationalists and neo-behaviorists tend to take it for granted that the real question about representations is whether we should be realist about the theory of representationalism. This paper is an attempt to shift the emphasis from the debate concerning realism about theories to the one concerning realism about entities. My claim is that regardless of whether we are realist about representational theories of the mind, we have compelling reasons to endorse entity realism about mental representations.

## 1 Introduction: Mental Representations

The concept of mental representation is a central concept of cognitive science and of philosophy of mind. When we describe any of our actions, or actions of other people, it is difficult to do so without talking about representations: I opened the window because I *thought* it was too hot and I did not *want* to turn on the air conditioner. This reference to thoughts and wants, to beliefs and desires has become the overarching framework for the philosophy of mind of the last 30 years at least: much of recent philosophy of mind is about the nature of these propositional attitudes, their relations to one another and to the brain.

Some philosophers of mind and most cognitive scientists use a much less restrictive concept of representation, which includes much more than just beliefs, desires

---

✉ Bence Nanay  
bn206@cam.ac.uk  
<http://www.bencenanay.com>

<sup>1</sup> Centre for Philosophical Psychology, University of Antwerp, D 413, 15 Prinsstraat, 2000 Antwerp, Belgium

<sup>2</sup> Peterhouse, University of Cambridge, Cambridge CB2 1RD, UK

and other propositional attitudes. Importantly, cognitive scientists (and some philosophers of mind) talk about representations in the perceptual system and representations that allow us to perform actions successfully. In fact, the very origins of cognitive science could be traced to the opposition to behaviorism in positing some kind of representations in the mind (again, not necessarily beliefs or desires).

This emphasis on representations provided a common ground for philosophy of mind and cognitive science that facilitated interaction between the two disciplines. But this emphasis on representations has been more recently questioned by what I will call – with some malice – ‘neo-behaviorist’ views of the mind.

The neo-behaviorists attempt to explain much of what goes on in our mind without any reference to representations. One motivation for the neo-behaviorist approach comes from worries about over-intellectualizing the mind (Hurley 2001). If we describe our mental life as the combination of beliefs, desires and other propositional attitudes, this makes it problematic to talk about animal minds and the minds of young children, but also to talk about many of our everyday actions without over-intellectualization. The neo-behaviorists, rather than trying to remedy this issue by being more precise about what kind of representations we use when describing the mind, reject any concept of representations at least when describing simple actions (of children, animals and of adult humans).

Neo-behaviorism comes in different flavors, some more radical than others. Most proponents of this view would accept that some complex, maybe linguistic human behavior could not be described without talking about representations, but insist that the vast majority of our actions and also our perception can be fully explained in non-representationalist terms (Chemero 2009; Hutto and Myin 2014; see also Ramsey 2007 for a nuanced summary of this approach).

Many neo-behaviorists focus on perception and argue that perception is not a representational process: there are no perceptual representations. As Dana Ballard put it, “the world is the repository of the information needed to act. With respect to the observer, it is stored ‘out there’, and by implication not represented internally in some mental state that exists separately from the stimulus” (Ballard 1996, p. 111; see also Brooks 1991). There are two major alternatives neo-behaviorists give to the representationalist framework. Some of them think of perception as an active process, some form of dynamic interaction with the world, which does not require representations (Noë 2004; Hurley 2001). Some others take perception to be a non-representational relation to the perceived object (Campbell 2002; Martin 2004; Travis 2004; Brewer 2011).

My aim here is not to assess the merits and demerits of these views or of neo-behaviorism in general. I am a representationalist and I will argue for representationalism in this paper. However, I will do so not by criticizing the neo-behaviorist views, but by taking their challenge seriously and asking the question: What justifies the attribution of representations to an agent?

This question is really a question about scientific methodology. Cognitive science posits an unobservable entity: representation (see Bechtel 2016 for a similar way of framing the problem, but see also Thomson and Piccinini 2018 who question this framing). Just as in physics and philosophy of physics, there are debates about when we are justified to posit an unobservable subatomic entity, there should

also be a debate in the philosophy of cognitive science and the philosophy of psychology about when we are justified to posit an unobservable<sup>1</sup> mental entity: mental representation.

## 2 Realism About Theories Versus Realism About Entities

My aim is to argue that if we apply some of the standard apparatus of the scientific realism debate in the context of the philosophy of cognitive science and philosophy of psychology, we can make real progress on the old questions about whether and when we are entitled to posit mental representations. As we have seen, the entities that are the most important ingredients of explanations in psychology and cognitive science are unobservable entities, just like electrons or subatomic particles. The question then is: how do the considerations about realism concerning unobservable entities we are familiar with from the scientific realism literature apply to them?

As we know from the classic scientific realism literature, there are different realism versus antirealism debates (see Newton-Smith 1978; Hacking 1983; Chakravartty 2007; Nanay 2013b for taxonomies). One can be realist or antirealist about theories. Scientific realism about theories is the view that “scientific theories are either true or false independent of what we know: science at least aims at the truth and the truth is how the world is” (Hacking 1983, p. 27). In other words, science aims to give us literally true claims about the world and there is a fact of the matter, independent of us, about how the world is. This definition has two conjuncts: one about what science aims to do and the other about the relation between scientific theories and the world. Scientific antirealists can deny either of these two conjuncts.

But this debate about theories is very different from, and logically independent of, the debate about theoretical entities. Realism about entities is the view that theoretical (unobservable) entities that scientific theories postulate really do exist. And, at least on the face of it, we can be realist about a theoretical entity a theory postulates without being realist about the theory that postulates it (see more on how this may work below).

On the rare occasions when philosophers or cognitive scientists defend representationalism, they assume that the view they need to defend is realism about representationalism as a theory. This can take the form of arguing that a representationalist theory of the mind is superior to a non-representationalist theory, in terms of explanatory scope, simplicity or empirical adequacy. And neo-behaviorists also assume this framework in their fight against the concept of representation: they aim to give a theory that covers all the available empirical facts and has the same explanatory scope, but that has an advantage over representationalist theories in terms of simplicity because it can do all that without positing representations (see esp. Chemero 2009; Hutto and Myin 2014 and see also Ramsey 2007 for a good summary).

---

<sup>1</sup> There is a debate in philosophy of perception whether we can represent some of the mental states of other people perceptually. If we can, then in some sense these mental states are observable. Note, however, that this is not the sense of ‘observable’ that is at stake in the scientific realism debate.

In short, both representationalists and neo-behaviorists tend to take it for granted that the real question about representations is whether we should be realists about the theory of representationalism. This paper is an attempt to shift the emphasis from the debate concerning realism about theories to the one concerning realism about entities. My claim is that regardless of whether we are realist about representational theories of the mind, we have compelling reasons to endorse entity realism about mental representations. And this is more than enough justification for talking about representations when talking about the mind.

### 3 Entity Realism

Entity realism is the view that, to quote Ian Hacking, “a good many theoretical<sup>2</sup> entities do really exist” (Hacking 1983 p. 27; see also Cartwright 1983, p. 89; see also Nanay 2019). Not a particularly controversial view these days. What is somewhat more controversial is what methodology entity realism should use. But the real controversy about entity realism comes from Hacking’s (and Cartwright’s) insistence that one can be realist about entities and antirealist about theories.

First, what I take to be the core commitment of entity realism is that the more evidence we have about the causal powers of *x*, the more reason we have to be realist about *x*. Taking causal powers as evidence for existence is hardly a very controversial move—the same connection does a lot of work in various metaphysical debates, for example, the one about the causal role of properties (e.g., Shoemaker 1979, Crane 2009). Jaegwon Kim even came up with a catchy label for it: “This we might call ‘Alexander’s Dictum’: to be real is to have causal powers” (Kim 1993, p. 202; see also Cargile 2003).

What is more controversial is how we can find out about the causal powers of an unobservable entity—the methodology of entity realism. And here we get some variations within the entity realist camp—as how one should proceed in this question clearly depends on one’s conception of causation in general. According to Nancy Cartwright, we have reason to endorse realism about entity *x* if *x* figures essentially in causal explanations of observable phenomena (Cartwright 1983, 1999). Ian Hacking’s view is slightly different: if we can manipulate *x* in such a way that this has direct influence on observable phenomena, we have reason to endorse realism about entity *x*. As his famous one-liner goes, “so far as I’m concerned, if you can spray them, they are real” (Hacking 1983, p. 23). I prefer another, less famous, but somewhat more informative, one-liner: “When we use entities as tools, as instruments of inquiry, we are entitled to regard them as real” (Hacking 1989, p. 578). Or, even more informatively:

We are completely convinced of the reality of electrons when we regularly set out to build and often enough succeed in building new kinds of device that use

---

<sup>2</sup> I will use the less controversial term ‘unobservable entities’ instead of Hacking’s original ‘theoretical entities’ in what follows.

various well-understood causal properties of electrons to interfere in the more hypothetical parts of nature (Hacking 1983, p. 265).

Hacking and Cartwright give two different ways of cashing out the very same idea: namely, that discovering the causal properties of an entity is what justifies realism about this entity. But they offer very different methodologies for establishing these claims. I will focus on Hacking's methodology mainly because Cartwright's relies on a fairly specific stance on the relation between causal explanation and causation that is easier to criticize (see, e.g., Psillos 2008).

When we are assessing entity realism, we should use the following methodology then: if we can manipulate 'well-understood causal properties' of a certain kind of entity in a way that would have direct observable effects, we can be convinced of the reality of this kind of entity. This, of course, leaves open the question about how we can manipulate unobservable entities (see Giere 1988, pp. 111–140), a question that will play an important role in Section IV.

These core commitments of entity realism have been criticized both for being too restrictive and for being too inclusive (see Shapere 1993 and Gelfert 2003, respectively). But the real controversy comes from the proposal that entity realism does not presuppose realism about theories. This is the reason why entity realism was a major shift in the realism versus antirealism debate: it pushed the debate about the truth of theories (or their aiming at the truth) in the background. Again, here is Hacking on the relation between entity realism and the realism versus antirealism debate about theories:

One can believe in some entities without believing in any particular theory in which they are embedded. One can even hold that no general deep theory about the entities could possibly be true for there is no such truth (Hacking 1983, p. 29).

In short, one can be realist about observable entities and be noncommittal about the realism versus antirealism debate about theories. Entity realism may even be consistent with antirealism about theories. As Erman McMullin sums up, according to entity realism, we may "know that the electron [exists], even though there is no similar assurance as to what it is" (McMullin 1984, p. 63; see also Clarke 2001). Further, Hacking argues that this seems to fit the actual scientific practice very well—scientists do things with unobservable entities without (or before) having any firm theories about them. So an experimental scientist may or may not be realist about theories, but as long as she conducts experiments with the help of unobservable entities, she must be realist about entities.

This tenet of entity realism—that it is consistent with antirealism about theories—has been very controversial (Morrison 1990; Resnik 1994; Massimi 2004; Chakravartty 2007; Psillos 1999, 2008; Hardin and Rosenberg 1982; Laudan 1984; Musgrave 1996), so I do not want to rely on it in this paper. But if it is true that entity realism does not imply realism about theories, then one can accept the main claim of this paper, that we should be entity realist about mental representations, while being noncommittal about whether realism about representationalist theories of mind is correct.

However, my main aim is to use the methodology of entity realism to find out more about the ontological commitments of psychology and cognitive science. Those who are unconvinced that entity realism is consistent with antirealism about theories should have no problem going along with the arguments presented in the rest of the paper.

## 4 Entity Realism About Representations

Mental representations are about some distal entity in the world. They attribute properties to a distal entity: they represent this entity as having certain properties. This can go wrong in two different ways. The represented entity may not exist. In this case, the mental representation misrepresents. Or the represented entity, although it exists, may not have the represented properties. This is also a case of misrepresentation. If the represented entity exists and has the represented properties, then the mental representation is correct.

Mental representations are not necessarily propositional attitudes like beliefs and desires. They are not necessarily syntactically structured. Further, mental representations are not necessarily conscious, and we do not necessarily have particularly reliable introspective access to them.

Representations attribute properties to distal entities in a way that can misrepresent. This allows for talking about representations in the visual system, even in the early visual system, like the primary visual cortex. But the photosensors of the retina will not count as representations—they carry information about the scene in front of our eyes, but they can't misrepresent. The primary visual cortex can. The concept of representation covers representations of very different kinds, from the ones in the primary visual cortex to beliefs and desires. My claim is that we should be entity realist about at least some of these. And the reason for this is that we can manipulate these representations.

What would constitute a good reason for attributing a mental representation to a subject? Before I address this question, I want to first set aside a very widespread bad reason for attributing mental representations: introspective report. If a subject introspects and says that she has such and such mental representation, this is not a very good reason to attribute this mental representation to her. Why? Because introspective access is very unreliable (see Nisbett and Wilson 1977 for the *locus classicus* and Schwitzgebel 2008; Spener and Bayne 2010; Spener 2011 for philosophical summaries). This is a vast literature and I can't do justice to all the wrinkles here. But a large number of studies from very different disciplines and using very different methodology have suggested that introspection is an unreliable guide to what goes on in our mind (see, e.g., Carruthers 2011 for a summary).

What would then constitute a good reason for attributing a mental representation to a subject? This is where entity realism comes in. Hacking's criterion in this context would be that we should be realist about mental representations if we can manipulate mental representations in such a way that it would have direct observable influence—which, in the present context would mean that it would influence our observable behavior directly. I will argue that we have strong reason

to endorse entity realism about mental representations if we follow this methodology. But before I do so, it is important to see that entity realism about mental representations is not an obviously correct view—it has its opponents. In fact, it has many opponents.

Behaviorists deny that there are mental representations. There is sensory input and there is motor output, but there is nothing representational between them (Watson 1930). It is important that at least some versions of behaviorism are consistent with the claim that a lot is going on between the sensory input and the motor output and we can learn about some of these processes. In other words, behaviorism is not committed to the caricature idea that the mind is a black box. But whatever mediates between the sensory input and the motor output is not something representational. Skinner, for example, explicitly allows for a neuroscientific description of such mediation:

The organism is, of course, not empty, and it cannot be adequately treated simply as a black box, but we must carefully distinguish between what is known about what is inside and what is merely inferred [...] The physiologist of the future will tell us all that can be known about what is happening inside the behaving organism. His account will be an important advance over a behavior analysis, because the latter is necessarily “historical”—that is to say, it is confined to functional relations showing temporal gaps. Something is done today which affects the behavior of the organism tomorrow. No matter how clearly that fact can be established, a step is missing, and we must wait for the physiologist to supply it. He will be able to show how an organism is changed when exposed to contingencies of reinforcement and why the changed organism then behaves in a different way, possibly at a much later date. What he discovers cannot invalidate the laws of a science of behavior, but it will make the picture of human action more nearly complete (Skinner 1974, pp. 233–237).

But the behaviorist stance towards neuroscience is ambivalent to say the least (see Catania and Harnad 1988). Skinner himself is not always as concessive as in the quote above. In the following quote he talks about the relation between sensory input (the ‘first link’), the neural processes (the ‘second link’) and the motor output (the ‘third link’):

Unless there is a weak spot in our causal chain so that the second [neural process] link is not lawfully determined by the first [sensory input], or the third [motor output] by the second, the first and third links must be lawfully related. [...] Valid information about the second link may throw light on this relationship but can in no way alter it. (Skinner 1953, p. 35)

The main point is that regardless of how much role behaviorists envisaged for the neural processes that mediate between sensory input and motor output, these processes are to be described in nonrepresentational terms. Behaviorism would, presumably, need to posit some neural processes (in order to account for various forms of learning, for example), but, it does not need to posit representations.

Behaviorism is not particularly popular today, but, as we have seen, neo-behaviorism is. They claim that sensory input and motor output are so closely intertwined in a dynamic process that we do not need to posit any representations that would mediate between them. The neo-behaviorist, like the old-fashioned behaviorist, would deny entity realism about mental representations.

My aim is to argue that entity realism, when applied to the question about whether mental representations exist gives a fairly straightforward positive answer.<sup>3</sup> We have seen that according to Ian Hacking's criterion, if we can manipulate mental representations in a way that would have direct influence on behavior, we would have a strong case for entity realism about mental representations. The problem is that it is far from clear what it would mean to manipulate mental representations. More generally, it is far from clear how we can manipulate unobservable entities. What would be the equivalent of the spraying of electrons in the domain of mental representations?

The general proposal is that if changes in what a mental representation represents directly influence our behavior, this would constitute a reason to accept entity realism about this kind of mental representation. The problem is that it is not clear how we could be certain that we change what this mental representation represents (and not some nonrepresentational processes leading to the behavior).

To put it differently, in order to establish entity realism about mental representations, we need to argue for two claims: First, that changes in a certain kind of mental state directly influence our observable behavior. Second, that these changes in our observable behavior cannot be explained in terms of nonrepresentational processing of the sensory input. I will label these as:

**Criterion A:** Changes in mental state *M* directly influence our observable behavior.

**Criterion B:** The changes in our observable behavior cannot be straightforwardly explained in terms of nonrepresentational processing of the sensory input.<sup>4</sup>

If both criteria are satisfied, we have good reasons for positing representation *M*.

I have said a lot about Criterion A in the last section. The importance of Criterion B can be made more explicit if we consider a version of Morgan's Canon. Here is Morgan's Canon in its original phrasing:

---

<sup>3</sup> Somewhat surprisingly, some advocates of neo-behaviorism do appeal to entity realism when they defend the reality of some entities in their nonrepresentational apparatus (see Chemero 2009, chapter 9, esp. pp. 192–194, where he defends entity realism about 'affordances'). I find this surprising, because it seems to me that there are vastly stronger empirical reasons to be entity realist about representations (which neo-behaviorists would want to avoid) than about affordances (which some neo-behaviorists endorse).

<sup>4</sup> It is worth noting that Criterion B could be replaced by a much weaker one and we could still get a strong reason for positing representations: that the representationalist explanation is more powerful/simpler/in some sense better than the non-representationalist one. As none of these ways of comparing explanations is unproblematic, if we manage to find evidence for Criterion B, this would be a much stronger reason for positing representations.



**Morgan's Canon:** "In no case may we interpret an action as the outcome of the exercise of a higher psychological faculty if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale." (Morgan 1894/1903, p. 53)

Morgan's Canon has been proposed as the correct methodology for attributing mental states to organisms. There may be reasons to doubt the evolutionary arguments in favor of Morgan's Canon and it is not clear how 'higher' and 'lower' psychological faculties are supposed to be distinguished (see Sober 1998, 2005, Karin-D'Arcy 2005, Buckner 2013). But for the purposes of this paper, I want to accept a much weaker version of Morgan's Canon as applied to representations (rephrasing Morgan's original formulation):

**Representational Morgan's Canon:** in no case may we interpret an action as the outcome of a representational process if it can be interpreted as the outcome of a nonrepresentational process.

A behaviorist (or neo-behaviorist) move against a realist stance towards mental representations would be to rely on the Representational Morgan's Canon (which, for the purposes of this paper I take to be uncontroversial) and insist that whatever causes the differences in our behavior is not a representation but some nonrepresentational processing of the sensory input.

The analogy often used in this context is with the heliotropism of some plants (see, e.g., Dennett 1991, p. 191). Some plants move their leaves or flowers in a way that tracks the location of the Sun. Examples include the snow buttercup (*Ranunculus adoneus*) and the bud (but not the mature plant) of sunflower (*Helianthus annuus*). If we follow the Representational Morgan's Canon, we should resist the temptation of attributing the representation of the spatial location of the Sun to these plants because there is a well-understood mechanism that explains heliotropism without postulating any representations (Sherry and Galen 1998; Galen and Stanton 2003; Vanderbrink et al. 2014). Because of the Representational Morgan's Canon, we should not be entity realist about the representation of the location of the Sun in plants (but see Morgan 2014 for some worries about this claim). My aim is to show that even if we accept the Representational Morgan's Canon,<sup>5</sup> we have strong reasons to be realist about some mental representations.

The mental representation I want to focus on is one that is directly involved in the successful performance of actions. I will label it 'motor representation' here (I called it 'pragmatic representation' earlier, see Nanay 2013a). Motor representations represent those parameters of the situations that are necessary for the successful performance of action. Just what these parameters are is debated: they may include the properties of the objects one acts upon, the properties of one's own body, one's bodily movement that is needed to complete the action or maybe the properties of the goal state the action is aimed at (see Jeannerod 1997;

---

<sup>5</sup> Note that not accepting the Representational Morgan's Canon would make the job of arguing for entity realism about mental representations much easier.

Nanay 2013a; Poincaré 1905/1958; Bach 1978; Brand 1984; Pacherie 2011; Milikan 2004; Butterfill and Sinigaglia 2014 for very different proposals about this).

For simplicity, I will assume that motor representations represent simple shape, size and spatial location of the distal objects the action is directed at. Arguably, simple representations of this kind are involved in all the accounts of motor representations I mentioned in the previous paragraph. These properties need to be represented in order for the agent to be able to perform the action at all. Suppose that the action is to pick up a cup. If I didn't represent the size of the cup, I would have no idea what grip size I should approach it with. If I didn't represent its spatial location, I would have no idea which direction I should reach out towards. And so on. Motor representations are genuine representations: they can misrepresent. If I represent the shape property of the cup correctly, then I will be more likely to approach it with the appropriate grip size, which makes it more likely that my action will be successful. And if I represent the spatial location of the cup correctly, I will be more likely to reach out in the right direction, which, again, makes it more likely that my action succeeds.

These motor representations do not need to be (and arguably they normally are not) conscious. But then how do we know what property (say, shape property or spatial location property) they attribute to the cup? Clearly not by introspecting. We can infer what shape property this motor representation attributes to the cup from the grip size the agent approaches the cup with. And we can infer what spatial location property it attributes to the cup from the direction of my reaching. In other words, if the shape property the motor representation attributes to the cup changes, this affects my behavior, that is, the grip size of my approaching hand, directly. And if the spatial location property the motor representation attributes changes, this also affects my behavior—the direction I reach out towards (see Jeannerod 1997 for a number of case studies of how intervention on the motor representation leads to observable changes in our behavior). In one famous experiment, in the middle to the performance of the reaching movement the target was changed—either its spatial location or its size. And this influenced the action execution—the reaching movement changed direction in the course of the execution of this action. The subjects are almost always unaware that anything has changed (Paulignan et al. 1991; Pelisson et al. 1986; Goodale et al. 1986).

Thus, Criterion A for entity realism about mental representations is satisfied: we can manipulate the motor representations to bring about direct changes in our observable behavior. But I said nothing about the Criterion B: that it is really the representation that got manipulated. I need to also show that these changes in behavior could not be explained in a purely nonrepresentational way (that is, in terms of the nonrepresentational processing of the sensory input).

And here we need to turn to the cognitive neuroscience of action. There are lots of empirical reasons to think that the grip size of grasping movements is determined by a representation (that is, motor representation) and not some nonrepresentational processing of the sensory input. What I take to be the most impressive piece of evidence is the following (this is by no means an isolated example, see Jeannerod 1997; Nanay 2013a for a more systematic treatment).

Two very widely used brands of matches in the UK are “Swan Vestas” and “Scottish Bluebell.” The box of Swan Vestas is 25 percent larger than that of Scottish Bluebell. It was tested whether the brand of the matchboxes influences our grip size when grasping them, and it was found that it does (McIntosh and Lashley 2008). When the subjects were grasping the 1.25-scale replica of the Scottish Bluebell box, their grip size was smaller than it was when grasping the normal Swan Vestas of the same size. And when they were grasping the 0.8-scale replica of the Swan Vestas box, their grip size was larger than it was when grasping the normal Scottish Bluebell box. Hence, the recognition of the brand of the matchboxes influences the grip size we approach it with. In a follow-up study, it was also pointed out that this influence is not due to local or short-term learning, but to long-term familiarity with these matchbox brands (Borchers et al. 2011; Borchers and Himmelbach 2012; Roche et al. 2015).

This is going to be difficult to explain in nonrepresentationalist terms. What happens in motor control in this case is an integration of the sensory input with high-level information (about the matchbox’s brand or maybe about some reliably co-occurring features of these matchboxes). This integration is crucial for yielding the grip-size we approach these objects with. And this integration is not the mere nonrepresentational processing of the sensory input—it is the combination of the sensory input with some fairly complex stored information—information about previously experienced matchboxes (Borchers et al. 2011; Borchers and Himmelbach 2012; Roche et al. 2015).<sup>6</sup>

Some hard-core neo-behaviorists may push back at this point and argue that as long as we allow for a non-representational way of explaining learning (including perceptual learning), we may be able to give a non-representational analysis of these match-box experiments. They could argue that while there is information in the system (as a result of past exposure to different kinds of match-boxes, our movement, and our grip size can be explained fully in terms of the processing of the input (of certain color combinations on the box). I am not entirely sure how such an explanation would work, but as I don’t want to exclude the possibility that it might, I want to modify the argument in a way that would be even more difficult for the neo-behaviorist to counter.

We have seen that manipulating motor representations leads to observable effects: if my motor representation attributed a different size property to the object, I would approach it with a different grip size. In other words, motor representations satisfy Criterion A for positing mental representations—the potential problems were with Criterion B. But there is another kind of mental state that satisfies Criterion A: a mental state I call, following Nanay (2013a), ‘pragmatic mental imagery’. I hope to show that pragmatic mental imagery fares (even) better than motor representations when it comes to Criterion B.

---

<sup>6</sup> It should also be noted that the matchbox study is not an isolated example. There are many results that show that motor representations are sensitive to various top-down factors: the subject’s attention (Marrett et al. 2011), her language skills and lexical recognition (Deng et al. 2012; Pulvermuller and Hauk 2005), and her expectations or knowledge (Roche et al. 2015).

Suppose that there is a cup in front of me. I can pick it up while looking at it: in this case, the visual feedback helps me to do so. I can adjust my movements in the light of my visual experience of how my action succeeds: if my initial reach was too forceful, I can adjust its course in response to the visual feedback (some of this happens unconsciously, see Paulignan et al. 1991; Pelisson et al. 1986; Goodale et al. 1986; see also Brogaard 2011).

But I can also perform this action, and do so fairly successfully, without looking. I am looking at the cup, close my eyes, count to ten and then reach out to grab it. In this case, it is my mental imagery that guides my action. It is a special kind of mental imagery inasmuch as it attributes very similar properties as motor representations do: egocentric spatial location properties (that allow me to reach out in the appropriate direction), egocentric size properties (that allow me to approach the cup with the appropriate grip size) and so on. And it is also, like motor representation, a genuine representation as it can misrepresent.

Manipulating pragmatic mental imagery leads to observable behavioral changes in the same way as manipulating motor representations leads to observable behavioral changes: if my pragmatic mental imagery attributed a different size property to the cup, I would approach it with a different grip size. So Criterion A is satisfied. How about Criterion B, which proved to be more problematic in the case of motor representations?

Consider the following slightly modified version of the previous example: there is a cup in front of you and you close your eyes, count to ten and try to pick it up, but before you do so, your friend tells you that she moved the cup to the left by 10 cm. You can still perform this action fairly successfully.

If we attribute representations to the subject, we have no problem explaining how this happens: your mental imagery attributes a spatial location property to where the cup was originally and the information you received from your friend modifies this spatial location property leading to the pragmatic mental imagery that represents the correct spatial location property. And this pragmatic mental imagery can guide you to perform the action successfully.

But what could the non-representationalist explanation be? The direction of my reach is directly influenced by a mental state that is the combination of two stimulus-independent informational states: my mental imagery (of where the cup used to be) and the verbally coded information about how my friend moved it. This is the combination of two information-carrying states. Even if neo-behaviorists can somehow explain how an information-carrying state influences behavior (and even if this information is influenced by previous exposure as in the case of motor representations), I see no way in which they could explain how the information from two information-carrying states would be combined without positing representations, especially given that one of these two information-carrying states is linguistically coded, which would be considered to be representational even by the most ardent proponents of neo-behaviorists (see, e.g., Hutto and Myin 2014).

It may help to contrast this case with the simpler example of just looking at the cup, closing my eyes, counting to ten and then picking it up with my eyes closed. In this case, the neo-behaviorist could say that I am acting on the sensory input but with a bit of delay—no representation needs to be posited.

But the case I gave above requires a much more complex mental process: the mental imagery of the cup I have when I close my eyes carries information about the cup that is coded in an egocentric frame of reference. The '10 cm to the left' information, in contrast, is not coded in an egocentric frame of reference. In order to revise my mental imagery and reach out in the appropriate direction, I need to put together these two very differently coded pieces of information.

Integrating two pieces of information that are coded in comparable reference frames (which is what happens in multimodal sensory integration, for example) could be argued to be explicable in non-representationalist terms (but see Nanay 2014 for an argument about how multisensory integration may give us reasons for positing perceptual representations). But I do not see how one can explain the integration between two pieces of information that involves translating linguistically coded information about (allocentric) distances into egocentric information one can act on.<sup>7</sup>

One may worry that the talk about mental imagery in this argument is ontologically suspicious. Didn't Ryle—himself very much in the behaviorist camp—refute any appeal to mental imagery, after all? In response, it is important to emphasize that the concept of mental imagery at play here is very different from the 'little pictures in the head' conception Ryle was making fun of. Mental imagery, in the meantime, has become a scientifically respectable concept that neuroscience has a lot to say about. And the way neuroscientists and psychologists talk about mental imagery has nothing to do with little pictures in the head. Neuroscience considers mental imagery to be perceptual processing that is not triggered by corresponding sensory stimulation in the relevant sense modality (Kosslyn et al. 1995; Pearson et al. 2015; Nanay 2015, 2018). Mental imagery in this sense is as scientifically and ontologically unproblematic as perception. And, a fortiori, so is pragmatic mental imagery.<sup>8</sup>

Let's go back to the explanations of the movement of heliotropic plants. This movement can be fully explained in terms of the nonrepresentational processing of the sensory input (Sherry and Galen 1998; Galen and Stanton 2003; Vanderbrink et al. 2014). The only information that is relied on when determining the movement of the plant is carried by the sensory input. So we need to conclude, via the Representational Morgan's Canon, that there is no need to postulate any representations here. Criterion B is not satisfied.

But in the case of pragmatic mental imagery, the fine-grained movements of action execution are determined by two informational states, which code spatial location information very differently: the mental imagery of where the cup used to be and verbal information from my friend. Given that we need to explain how these two very differently coded pieces of information are combined and used to guide our movement, this rules out any explanation by means of purely nonrepresentational

---

<sup>7</sup> It should be emphasized that the kind of representations posited for the reasons outlined above would be consistent with the kind of representations that Ramsey 2007 would allow for, so my view is not inconsistent with his.

<sup>8</sup> Thanks for an anonymous referee for pushing me to clarify what I mean by mental imagery.

processing of the sensory input. So the Representational Morgan's Canon does not apply in this case. Criterion B is satisfied.

We can now put together the argument for entity realism about motor representations: changes in the properties the pragmatic mental imagery attributes directly influence our motor behavior (Criterion A). And these changes in motor behavior cannot be explained in terms of nonrepresentational processing of sensory input (Criterion B). In other words, we have good reasons to be entity realist about at least some kinds of representations.

Note that these considerations give us reason to endorse entity realism about mental representations. If it is true that entity realism in general does not entail realism about theories, then they do not give us any reason to endorse representational theories of the mind. As Hacking pointed out, scientists do things with unobservable entities without (or before) having any firm theories about them and this is definitely true of neuroscientists (see Thomson and Piccinini for a related argument). More generally, entity realism about representations only requires that we can causally manipulate representations. This requires attributing some causal power to these representations, but all the other properties of these representations could be left unspecified. And this picture fits the argument above: the argument left it unspecified just how and what motor representations represent (and there are many radically different theories about this, see Butterfill and Sinigaglia 2014; Jeannerod 1997; Nanay 2013a; Poincaré 1905/1958; Bach 1978; Brand 1984; Pacherie 2011; Milikan 2004). These theories differ in their account of what motor representations represent (the object's actual properties vs. a goal state vs. the agent's potential action) and they also differ in how they connect up with the rest of our mind. But they are all entity realist about motor representations.

Stephen Stich said in 1984 that:

we now have an enormous collection of experimental data which, it would seem, simply cannot be made sense of unless we postulate something like [representations] (Stich 1984, p. 649).

I am not sure that this was in fact true in 1984—to the extent that would have satisfied an ardent proponent of the Representational Morgan's Canon. But with the advances of the cognitive neuroscience of action, Stich's claim is definitely true now.

## 5 Conclusion: What Kinds of Representation?

The scope of the argument I presented in the last section was very limited: it was about motor representations and pragmatic mental imagery. This argument gives us strong reasons to endorse entity realism about motor representations and even

stronger reasons to endorse entity realism about pragmatic mental imagery. Can we generalize this argument to other representations—maybe ones more familiar to us from our folk psychology, like beliefs and desires?

In this brief conclusion, I want to warn against any such generalization. The argument in the last section was based on very specific empirical findings that we could only explain if we posit motor representations or pragmatic mental imagery. It provides no justification for drawing any conclusions about any other kinds of representations (although it could be thought of as providing an example of what kind of evidence someone who wants to argue for entity realism about other kinds of representations should be looking for).

In other words, the conclusion of this paper would be consistent with an all-out entity realist stance about various forms of representations (including beliefs and desires). But it would also be consistent with the classic 1980s eliminativist stance against beliefs and desires (Stich 1983; Churchland 1981, 1986, 1988).

**Acknowledgements** This work was supported by the ERC consolidator Grant [726251], the FWO Odysseus Grant [G.0020.12N] and the FWO research Grant [G0C7416N]. Special thanks for comments by Patrick Butlin, Anna Ichino, Alex Geddes, Lu Teng, Manolo Martinez and three anonymous referees.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bach, K. (1978). A representational theory of action. *Philosophical Studies*, 34, 361–379.
- Ballard, D. H. (1996). On the function of visual representation. In K. Akins (Ed.), *Perception* (pp. 111–131). New York: Oxford University Press.
- Bechtel, W. (2016). Investigating neural representations: The tale of place cells. *Synthese*, 193, 1287–1321.
- Borchers, S., & Himmelbach, M. (2012). The recognition of everyday objects changes grasp scaling. *Vision Research*, 67, 8–13.
- Borchers, S. A., Christensen, L., Ziegler, & Himmelbach, M. (2011). Visual action control does not rely on strangers—Effects of pictorial cues under monocular and binocular vision. *Neuropsychologia*, 49, 556–563.
- Brand, M. (1984). *Intending and acting*. Cambridge: The MIT Press.
- Brewer, B. (2011). *Perception and its objects*. Oxford: Oxford University Press.
- Brogaard, B. (2011). Are there unconscious perceptual processes? *Consciousness and Cognition*, 20, 449–463.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Buckner, C. (2013). Morgan's Canon, meet Hume's Dictum: Avoiding Anthropofabulation in cross-species comparisons. *Biology and Philosophy*, 28, 853–871.
- Butterfill, S., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, 88, 119–145.
- Campbell, J. (2002). *Reference and consciousness*. Oxford: Oxford University Press.
- Cargile, J. (2003). On 'Alexander's' dictum. *Topoi*, 22, 143–149.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Cartwright, N. (1983). *How the laws of physics lie*. New York: Oxford University Press.

- Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge: Cambridge University Press.
- Catania, C., & Harnad, S. 1988 (eds). *The selection of behavior: The operant behaviorism of B. F. Skinner: Comments and consequences*. Cambridge: Cambridge University Press.
- Chakravartty, A. (2007). *A metaphysics for scientific realism: Knowing the unobservable*. Cambridge: Cambridge University Press.
- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: MIT Press.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78, 67–90.
- Churchland, P. (1986). *Neurophilosophy*. Cambridge: MIT Press.
- Churchland, P. M. (1988). Folk psychology and the explanation of behaviour. *Proceedings of the Aristotelian Society*, 62, 209–221.
- Clarke, S. (2001). Defensible territory for entity realism. *British Journal for the Philosophy of Science*, 52, 701–722.
- Crane, T. (2009). Causation and determinable properties: on the efficacy of colour, shape and size. In J. Kallstrup & J. Howhy (Eds.), *Being reduced*. Oxford: Oxford University Press.
- Deng, Y., Guo, R., Ding, G., & Peng, D. (2012). Top-down modulations from dorsal stream in lexical recognition: And effective connectivity fMRI study. *PLoS ONE*, 7, e33337.
- Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.
- Galen, C., & Stanton, M. L. (2003). Sunny-side up: flower heliotropism as a source of parental environmental effects of pollen quality and performance in the snow buttercup, *Ranunculus adoneus*. *American Journal of Botany*, 90, 724–729.
- Gelfert, A. (2003). Manipulative success and the unreal. *International Studies in the Philosophy of Science*, 17, 245–263.
- Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.
- Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, 320, 748–750.
- Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.
- Hacking, I. (1989). Extragalactic reality: the case of gravitational lensing. *Philosophy of Science*, 56, 555–581.
- Hardin, C. L., & Rosenberg, A. (1982). In defence of convergent realism. *Philosophy of Science*, 49, 604–615.
- Hurley, S. L. (2001). Perception and action: Alternative views. *Synthese*, 129, 3–40.
- Hutto, D. D., & Myin, E. (2014). *Radicalizing enactivism*. Cambridge, MA: MIT Press.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford: Blackwell.
- Karin-D'Arcy, M. R. (2005). The modern role of Morgan's canon in comparative psychology. *International Journal of Comparative Psychology*, 18, 179–201.
- Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.
- Kosslyn, S. M., Behrmann, M., & Jeannerod, M. (1995). The cognitive neuroscience of mental imagery. *Neuropsychologia*, 33, 1335–1344.
- Laudan, L. (1984). Discussion: Realism without the real. *Philosophy of Science*, 51, 156–162.
- Marrett, N. E., de-Wit, L. H., Roser, M., Kentridge, R. W., Milner, A. D., & Lambert, A. J. (2011). Testing the dorsal stream attention hypothesis: electrophysiological correlates and the effects of ventral stream damage. *Visual cognition*, 19, 1089–1121.
- Martin, M. G. F. (2004). The limits of self-awareness. *Philosophical Studies*, 120, 37–89.
- Massimi, M. (2004). Non-defensible middle ground for experimental realism: Why we are justified to believe in colored quarks. *Philosophy of Science*, 71, 36–60.
- McIntosh, R. D., & Lashley, G. (2008). Matching boxes: Familiar size influences action programming. *Neuropsychologia*, 46, 2441–2444.
- McMullin, E. (1984). A case for scientific realism. In J. Leplin (Ed.), *scientific realism*. Berkeley: University of California Press.
- Millikan, R. G. (2004). *Varieties of meaning*. Cambridge: The MIT Press.
- Morgan C. L. (1894/1903). *An introduction to comparative psychology* (2 edn). London: Walter Scott.
- Morgan, A. (2014). Representations gone mental. *Synthese*, 191, 213–244.
- Morrison, M. (1990). Theory, intervention and realism. *Synthese*, 82, 1–22.
- Musgrave, A. (1996). 'Realism, truth, and objectivity'. In R. S. Cohen, R. Hilpinen, & Q. Renzong (Eds.), *Realism and anti-realism in the philosophy of science* (pp. 19–44). Dordrecht: Kluwer.
- Nanay, B. (2013a). *Between perception and action*. Oxford: Oxford University Press.



- Nanay, B. (2013b). Singularist semirealism. *British Journal for the Philosophy of Science*, *64*, 371–394.
- Nanay, B. (2014). Empirical problems with anti-representationalism. In B. Brogaard (Ed.), *Does perception have content?* (pp. 39–50). New York: Oxford University Press.
- Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies*, *172*, 1723–1736.
- Nanay, B. (2018). Multimodal mental imagery. *Cortex*, *105*, 125–134.
- Nanay, B. (2019). Entity realism and singularist semirealism. *Synthese*, *196*, 499–517.
- Newton-Smith, W. (1978). The underdetermination of theory by data. *Proceedings of the Aristotelian Society, supplementary*, *52*, 71–91.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Noë, A. (2004). *Action in perception*. Cambridge, MA: The MIT Press.
- Pacherie, E. (2011). Nonconceptual representations for action and the limits of intentional control. *Social Psychology*, *42*, 67–73.
- Paulignan, Y., MacKenzie, C. L., Marteniuk, R. G., & Jeannerod, M. (1991). Selective perturbation of visual input during prehension movements: I. The effect of changing object position. *Experimental Brain Research*, *83*, 502–512.
- Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, *19*, 590–602.
- Pelisson, D., Prablanc, C., Goodale, M. A., & Jeannerod, M. (1986). Visual control of reaching movements without vision of the limb: II. Evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus. *Experimental Brain Research*, *62*, 303–311.
- Poincaré, H. (1905). *The value of science*. New York: Dover.
- Psillos, S. (1999). *Scientific realism: How science tracks truth*. London: Routledge.
- Psillos, S. (2008). Cartwright's realist toil. In S. Hartmann, C. Hofer, & L. Bovens (Eds.), *Nancy cartwright's philosophy of science* (pp. 167–194). London: Routledge.
- Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.
- Resnik, D. (1994). Hacking's experimental realism. *Canadian Journal of Philosophy*, *24*, 395–412.
- Roche, K., Verheij, R., Voudouris, D., Chainay, H., & Smeets, J. B. J. (2015) Grasping an object comfortably: Orientation information is held in memory. *Experimental Brain Research*, in print.
- Schwitzgebel, E. (2008). *The Unreliability of Naive Introspection Philosophical Review*, *117*, 245–273.
- Shapere, D. (1993). 'Astronomy and Anti-Realism'. *Philosophy of Science*, *60*, 134–150.
- Sherry, R. A., & Galen, C. (1998). The mechanism of floral heliotropism in the snow buttercup, *Ranunculus adoneus*. *Plant, Cell and Environment*, *21*, 983–993.
- Shoemaker, S. (1979). 'Causality and Properties', in *Identity, Cause, and Mind* Cambridge: Cambridge University Press: 206–233.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B. F. (1974). *About behaviorism*. New York: Vintage.
- Sober, E. (1998). Morgan's canon. In C. Allen & D. Cummins (Eds.), *The evolution of mind* (pp. 224–242). Oxford: Oxford University Press.
- Sober, E. (2005). Comparative Psychology Meets Evolutionary Biology: Morgan's Canon and Cladistic Parsimony. In L. Daston & G. Mitman (Eds.), *Thinking with animals: New perspectives on anthropomorphism* (pp. 85–99). New York: Columbia University Press.
- Spener, M. (2011). Using first person data about consciousness. *Journal of Consciousness Studies*, *18*, 165–179.
- Spener, M., & Bayne, T. (2010). Introspective humility. *philosophical Issues*, *20*, 1–22.
- Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge MA: MIT Press.
- Stich, S. (1984). Is Behaviorism Vacuous? *Behavioral and Brain Sciences*, *7*, 647–649.
- Thomson, E., & Piccinini, G. (2018). Neural representation observed. *Minds and Machines*, *28*(1), 191–235.
- Travis, C. (2004). The silence of the senses. *Mind*, *113*, 57–94.
- Vanderbrink, J. P., Brown, E. A., Harmer, S. L., & Blackman, B. K. (2014). Turning heads: The biology of solar tracking in sunflower. *Plant Science*, *224*, 20–26.
- Watson, J. B. (1930). *Behaviorism*. New York: W.W. Norton & Company Inc.